OECD Education Working Papers No. 83

# Assessment for Qualification and Certification in Upper Secondary Education: A Review of Country Practices and Research Evidence

**Stefanie Dufaux**

https://dx.doi.org/10.1787/5k92zp1cshvb-en

OECD

**Unclassified** **EDU/WKP(2012)20**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development **04-Sep-2012**

**English - Or. English**

**DIRECTORATE FOR EDUCATION**

EDU/WKP(2012)20
**Unclassified**

# ASSESSMENT FOR QUALIFICATION AND CERTIFICATION IN UPPER SECONDARY EDUCATION: A REVIEW OF COUNTRY PRACTICES AND RESEARCH EVIDENCE

**OECD Education Working Paper No. 83**

**by Stefanie Dufaux**

*This paper was prepared by Stefanie Dufaux, who joined the OECD Directorate for Education in September 2010 as a Carlo Schmid Fellow, and worked as a Consultant in the Education and Training Policy Division until July 2011. This paper is part of the work undertaken by the OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes.*

*The OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes is designed to respond to the strong interest in evaluation and assessment issues evident at national and international levels. The overall purpose is to explore how systems of evaluation and assessment can be used to improve the quality, equity and efficiency of school education. The Review looks at the various components of assessment and evaluation frameworks that countries use with the objective of improving student outcomes. These include student assessment, teacher appraisal, school assessment and system evaluation. More information is available at: www.oecd.org/edu/evaluationpolicy.*

Contacts: Ms. Deborah Nusche [Tel: +33(0) 1 45 24 78 01; e-mail: deborah.nusche@oecd.org]
and Mr. Paulo Santiago [Tel: +33(0) 1 45 24 84 19; e-mail: paulo.santiago@oecd.org].

**JT03325796**

**Complete document available on OLIS in its original format**
***This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.***

**English - Or. English**

**OECD EDUCATION WORKING PAPERS SERIES**

This series is designed to make available to a wider readership selected studies drawing on the work of the OECD Directorate for Education. Authorship is usually collective, but principal writers are named. The papers are generally available only in their original language (English or French) with a short summary available in the other.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The opinions expressed in these papers are the sole responsibility of the author(s) and do not necessarily reflect those of the OECD or of the governments of its member countries.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgement of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org.

Comment on the series is welcome, and should be sent to edu.contact@oecd.org.

---

www.oecd.org/edu/workingpapers

---

Copyright OECD 2012

## ABSTRACT

Within the policy field of student assessment, the assessment of students for qualification and certification in upper secondary education has special importance since key decisions for the progression of students may be taken on the basis of assessment results. Students in most OECD countries face increased specialisation in upper secondary education and high stakes are associated to their performance when assessment results are used as a criterion for selection, both for access to higher education and other educational programmes and for access to the labour market. On the basis of research findings and country practices, this paper describes key features of assessment for qualification and certification in upper secondary education and discusses issues regarding its design and implementation.



## RÉSUMÉ

Dans le domaine de l'évaluation des élèves, les évaluations pour qualification et l'obtention d'un diplôme au deuxième cycle de l'enseignement secondaire ont une importance particulière puisque des décisions clés sur la progression des élèves peuvent être prises sur la base des résultats d'évaluation. Les élèves dans la plupart des pays de l'OCDE font face à une spécialisation accrue au deuxième cycle de l'enseignement secondaire et à des enjeux élevés associés à leurs performances, lorsque les résultats d'évaluation peuvent être utilisés comme critère de sélection à la fois pour l'accès à l'enseignement supérieur et autres programmes d'éducation, et au marché du travail. Dans ce rapport, les caractéristiques principales des évaluations pour qualification et l'obtention d'un diplôme au deuxième cycle de l'enseignement secondaire, ainsi que les thématiques associées à leurs conceptions et mises en œuvre sont présentées à la lumière des résultats de la recherche et des pratiques nationales.

## TABLE OF CONTENTS

1. INTRODUCTION ..... 5

2. TERMS AND DEFINITIONS ..... 7

3. KEY FEATURES OF UPPER SECONDARY EDUCATION INFLUENCING STUDENT ASSESSMENT AT THIS LEVEL ..... 8

3.1 Pathways for certification ..... 8
3.2 Individualisation of certification ..... 11

4. DESIGN AND GOVERNANCE OF ASSESSMENT FOR QUALIFICATION AND CERTIFICATION ..... 13

4.1 Purposes of assessment for qualification and certification ..... 13
4.2 Sharing of responsibilities for assessment for qualification and certification ..... 15

5. PROCEDURES FOR ASSESSMENT FOR QUALIFICATION AND CERTIFICATION ..... 18

5.1 The role of standards in qualification frameworks in upper secondary education ..... 18
5.2 Continuous versus end-point assessment ..... 20
5.3 Formats of assessment for qualification and certification ..... 27
5.4 Grading issues at upper secondary level ..... 28

6. USE OF RESULTS ..... 35

6.1 Entrance to higher education ..... 35
6.2 Entrance to the labour market ..... 35

7. CONCLUSION ..... 36

ANNEX I – CONCEPTUAL FRAMEWORK FOR ASSESSMENT FOR QUALIFICATION AND CERTIFICATION IN UPPER SECONDARY EDUCATION ..... 37

ANNEX II – EXAMPLES OF COUNTRY APPROACHES ..... 38

REFERENCES ..... 44

## 1. INTRODUCTION

1. Upper secondary education is characterised by increased subject specialisation and the preparation of students for their future professional or educational career. The purpose of summative assessment at this level of education is therefore a very specific one: it records students' achievement and certifies abilities to the labour market and further education institutions. As a result, certification has a crucial impact on students' future career opportunities.

2. Assessment for qualification and certification refers to the testing of students for the official recognition of their competencies. A qualification relies on defined standards of achievements, set by a responsible institution (OECD, 2007). These defined standards are usually embedded into a qualification framework, meaning a rank order of several qualification levels (OECD, 2010a). Awarding a qualification is thus an official recognition of a student's achievements by a competent body. In a further step, certification assures the documentation of students' skills which allows them to communicate their competencies to others (OECD, 2007). As the end of upper secondary education represents an important point of decision for a student's career progression in most OECD countries, assessment for qualification and certification at this stage of education carries high stakes, which leads to several particularities for its design and implementation.

3. OECD countries show a great variety of approaches towards assessment for qualification and certification in upper secondary education. Qualification frameworks may lead students into different pathways, such as the progression into higher education or entry into the labour market and may offer possibilities of subject choice for individualised specialisation. Centralised versus decentralised approaches to processing assessment for qualification and certification may have implications for grading practices and reliability. The choice of assessment tools depends on their capacity to produce a reliable picture of what students have learned. There are a number of policy options to be considered by policy makers and practitioners to design assessment for qualifications in a way as to provide the right incentives to teachers and students and ensure transparency in communicating students' skill levels to further educational institutions or the labour market.

4. This paper intends to address challenges in assessment for qualification and certification in upper secondary education by reviewing the research literature in this field. It will review key points of debate in the literature and present studies pointing to both positive and adverse effects of given approaches. For example, several scholars find that curriculum-based external exit exams entail positive motivational and performance effects for students (Woessmann *et al.*, 2007; Bishop 1997a, 1999) while others argue that they may reduce the scope of teaching and learning (Lee and Lee, 2001; Volante, 2004).

5. This review will synthesise research findings in the area and use them as a basis for the reflection on country practices. The focus of the paper will be on ISCED level 3, meaning the final stage of secondary education, typically at an entrance age of 15 or 16 years, and will not consider other forms of assessment which lie outside of upper secondary education, such as higher education entrance exams, or assessment of vocational training outside of general education.

6. Comparative literature on assessment practices in upper secondary education is scarce (Köditz, 2009; Le Métais, 2002). Most English-language quantitative studies in the field build their evidence on

data from the United States. Nevertheless, findings from these studies can provide some useful insights for other countries as well. In addition, the review considers information from a range of other education systems, such as New Zealand and Scotland (Raffe, 2003; Strachan, 2002). As the organisation of assessment for qualification and certification in upper secondary education is a very fast changing policy field, this paper only refers to recent studies for the comparison of policy practices.

7. In this paper, the following issues will be addressed:

- What are the options for the design of assessment for qualification and certification in upper secondary education, and what are their potential effects?

- Which approaches are chosen by countries and what are their respective advantages and challenges?

8. After this introduction and a short section on terms and definitions, this paper addresses these questions in four main sections. Section 3 presents key features of upper secondary education that influence student assessment practices at this level. Section 4 then looks at the design and governance of assessment for qualification and certification in different contexts. It considers the purposes of assessment for qualification and certification as well as the different options for sharing responsibilities for their development and implementation. Section 5 reviews the procedures used for assessment for qualification and certification. This includes a discussion on the role of standards in qualification frameworks, the issues surrounding internal versus external assessment, the different formats used for assessment and the approaches used for correcting and grading assessments. Section 6 discusses different uses of assessment results, namely for entrance to higher education or to the labour market. The paper then concludes on research findings and policy practices among OECD countries. Annex I presents the conceptual framework used for this paper.

## 2. TERMS AND DEFINITIONS

9. Assessment for qualification and certification is a specific form of summative and high-stakes assessment. For this paper, it is important to clarify the meaning of each of the key terms and concepts used throughout the paper. The main concepts used in the following discussion are summative assessment, qualification, certification and high stakes.

10. Summative assessment measures learning that has taken place with the purpose of presenting a summary of students' achievements over a defined period, in this case upper secondary education. Hence, summative assessment reports on students' learning achievements for judgement at a specific point in time. It is often described as "assessment *of* learning" and contrasted with "assessment *for* learning". For a comprehensive literature review on summative assessment, see Nusche (forthcoming).

11. Assessment for qualification and certification carries high stakes for the student, meaning that real consequences result for the learner and possibly for the teacher, the school or the community (Au, 2007). In upper secondary education, these consequences usually relate to decisions on the student's future, such as entrance into higher education or the labour market. Results from assessment for qualification and certification may be used by a wide range of stakeholders, namely the student, parents, teachers, higher education institutions and employers.

12. Qualification is a formal recognition of learning. It relies on a predefined standard of competencies that is achieved by a learner when a competent body concludes so. This competent body, which can consist of teachers, schools, or agencies external to schools, bases its qualification confirmation on a validation process, usually an assessment or in some cases the completion of a specific learning unit which the learner attended (OECD, 2007). The awarding of qualifications grants an official recognition of a student's competencies.

13. Certification evolves subsequently to the assessment for qualification. A certificate is an official document which records qualifications and thus serves as a communication tool to all stakeholders (OECD, 2007). In the case of assessment for qualification and certification in upper secondary education, certificates convey information in particular to higher education institutions and the labour market. As such, certification fulfils an important social function (Remesal, 2011).

14. As a product of all these components, assessment for qualification and certification in upper secondary education serves a very specific validation and communication purpose which implies an important valuation of students' competencies, and consequently has a great importance for their career.

## 3. KEY FEATURES OF UPPER SECONDARY EDUCATION INFLUENCING STUDENT ASSESSMENT AT THIS LEVEL

15. Assessment for qualification and certification needs to be seen in the context of upper secondary education structures. This section provides background information on how OECD countries design upper secondary education systems and explains some key concepts such as pathway engineering and individualisation of certification, which influence assessment practices at this level of education.

### 3.1 Pathways for certification

16. Students' pathways often become differentiated at the stage of upper secondary education, sometimes on the basis of assessment. Pathway choice is highly dependent on the options and the degree of flexibility provided by the education system.

17. In many OECD countries, students follow the same education track until the end of compulsory education, when they are aged 15 or 16 years old. In fact, more and more countries take the approach that tracking and streaming of students should be applied as late as possible in the education process (Mueller and Wolbers, 2003).

18. In post-compulsory education, a certain degree of flexibility through the organisation of intersecting pathways provides opportunities and incentives for students to make transitions (Bouder *et al.*, 2008). While preparing students for their future career path and offering alternative decisions, pathway engineering in post-compulsory education appears to respond to the needs of rapidly changing labour markets and post-secondary education systems. While Raffe (2001) emphasises the importance of pathway diversity, Mueller and Wolbers (2003) argue that less pathway differentiation but greater diversity and possibilities for individualisation within one general education programme enables students to change and adapt their study plans at any given time.

*The concept of pathways*

19. The concept of pathways is a metaphor which describes and contrasts educational options in post-compulsory education (Raffe, 2001). Even though pathways are expected to lead to specific career destinations according to Lund (2008), they may interconnect and thus allow students to change direction. This implies that students are in need for guidance in order to find their personal pathway. In upper secondary education, pathways define the transition between compulsory education and post-secondary educational options. The main pathway differentiation made in many countries is between tracks leading towards higher education and tracks leading directly to the labour market.

20. Pathway decisions are usually made based on assessment results at different stages of progression. Assessment thus marks points of successful educational completion at which the student may have the option to choose between different pathways. Conversely, failing tests at these points of intersection may result in a limitation of options for the student or even the arrival to a dead end. Thus, assessment determines ways of proceeding at pathway intersections.

*Designing pathways*

21. Many countries are engineering their qualification systems to enhance flexibility of pathways. In this context, pathways are being brought closer together or are being unified within one system of common curricula, governance and quality assurance through assessment (Raffe, 2001). Modularisation and credit systems, as well as qualification frameworks with several qualification levels, enable students to transfer their achievements into other pathways. The existence of multiple decision points over time can also lead to broader options for students. In New Zealand, for example, an elaborated qualification framework has been put in place to enable students to progress easily on the basis of recognition of their prior learning (Bolstad, 2006). Flexibility of pathways thus creates opportunities and incentives for students.

22. Systems in which both types of subjects, academic and vocational, are combined within the same school, are referred to as integrated systems (Le Métais, 2002). This is for example the case in Sweden and Norway. Trends in policy making in many countries are trying to bring these two main tracks closer together so that students can keep their options open while taking courses in different fields (Raffe, 2001). Nevertheless, integrated school systems may also include specialised schools besides integrated institutions.

23. Introducing Vocational Education and Training (VET) into the general framework of upper secondary education not only provides a real choice for students who do not aim to continue to higher education but also an alternative for students at risk of drop-out (Fullarton, 2001).

24. Mueller and Wolbers (2003) argue that pathway flexibility may decrease students' adequate preparation for various needs on the labour market. According to these authors, an early track differentiation could prepare students better in an environment which corresponds to their level of abilities, capacities, performance and aspirations. Especially in preparation for the labour market, some countries have very specific expectations towards the mix of skills and knowledge a student should have.

25. Raffe (2001) explains that flexible systems seem to promote individualisation which allows students to gain more control over their personal itineraries regardless of social structures such as social and gender inequalities. However, he also points out that a higher differentiation of itineraries not necessarily means more control over them and that inequalities persist in a system which promotes students' personal responsibility for transition.

*Country practices in designing pathways*

26. According to Mueller and Wolbers (2003), Sweden and Finland are among the countries with the lowest differentiation of study tracks. But also Canada, Ireland, England, New Zealand, Spain and the USA have integrated systems. In Ireland, courses taken lead to three different types of certification, namely the Leaving Certificate (Established), the Leaving Certificate (Vocational) and the Leaving Certificate (Applied) (Le Métais, 2002). In New Zealand, the National Qualifications Framework allows students to combine courses in schools, colleges and private institutions.

27. In Switzerland, only students attending the Gymnasia receive access to higher education (INCA Switzerland, 2009). But students may alternatively choose a qualification pathway, combining academic and vocational courses which provide access to both apprenticeships and tertiary non-university education (OECD, 2000). The example of Switzerland indicates that successful pathway engineering highly depends on further education options, especially those which combine tertiary and further vocational education.

*Second chance institutions for drop-outs*

28. In OECD countries, the average enrolment rate in upper secondary education was 82% in 2008 and has the tendency to grow further (OECD, 2010b). However, this does not mean that all students pass the final exam and achieve certification. To a large extent, attainment rates also rely on the system's ability to provide second chance pathways for drop-out students. These chances can be given through simple retest opportunities, but also though alternative certification by providing different pathways and tests.

29. Retaining students in upper secondary education and lifelong learning is becoming a declared policy priority in many countries, like Denmark or the Flemish Community of Belgium (Danish Ministry of Education, 2011; Vlor, 2010; OECD, 2000). Most education systems try to keep low performing students at upper secondary level in the system by providing retesting opportunities and second chance options. But numerous factors, such as the psychological effects of failing the test or students not being informed properly about retaking options may prevent from students retaking the exam (Ou, 2010).

30. In some countries, education safety nets offer alternative pathways leading to equivalent certificates for students. In the United States, for example, assessment for the General Education Development certificate (GED) provides the possibility for students to achieve equivalent certification after drop-out. Tyler and colleagues (2000) emphasise the benefits of the GED, pointing out that students may not only signal competences which they hold, but also enhance their skills while studying for the exam and create an opportunity to proceed in education by formalising their achievements.

31. In the French Community of Belgium, students who could not qualify for entry into higher education during upper secondary education may follow a special one-year programme in continuation to upper secondary school (*7e année de l'enseignement professionnel secondaire*) (OECD, 2005).

32. Ireland similarly runs the so called Learning Certificate Applied (LCA) programme, motivated by the desire to retain at-risk students within the full-time education system. Students who drop out may change into the LCA and achieve certification. This certificate, however, is not equivalent to the Established Leaving Certificate but allows the student to proceed into several post-graduation courses and thus to move on in education (Banks *et al.*, 2010).

33. Many education systems moreover deal with the risk of drop-out in a preventive manner. In Finland for example, students who do not meet the criteria for entering upper secondary education after grade 9, may attend a one-year specific course of study at the end of compulsory education to receive additional preparation for the continuation of their studies (Grubb *et al.*, 2005). If they can achieve the required test score, they may progress into upper secondary general education or choose among other educational programmes (Field *et al.*, 2007). Especially in the Scandinavian countries, emphasis is also given to the availability of guidance and counselling (OECD, 2000).

34. Tuck (2007) suggests that increased modularisation of learning and assessment increases the chances to retain students in the system. Such modularisation may increase access and options to progress individually within the qualifications framework. Financial costs and time requirements can be reduced for returners into the education system (Hart and Howieson, 2004). Furthermore, a greater variety of institutions may contribute. In New Zealand, the national qualification framework provides multiple opportunities for students to progress within consecutive modules. Moreover, both England's and Scotland's reforms, namely the *Curriculum 2000* and the *Higher Still* encompass clear trends towards increased modularisation to enhance pathway flexibility. However, as each unit requires assessment, the increased amount of assessment connected with this reform created challenges in both countries (Hodgson *et al.*, 2004).

### 3.2 Individualisation of certification

35. Another important element influencing assessment for qualification and certification is the degree to which students may choose their personal learning content. By choosing between different modules and subjects, students can influence the content of what they learn and will be assessed on. This may have major consequences for certification, communication of skills and progression.

36. In many OECD countries, national qualification systems do not only offer alternative educational tracks but also individual selection and composition of learning units within one track. Students are offered a subject choice for the individual composition of learning according to their personal interests. Achievements are measured in credits and add together for final certification. However, the degree of choice for students can vary significantly (Le Métais, 2002). In most OECD countries students are required to study the national language, mathematics as well as one science and humanist subject. One art subject and/or physical education are also often included, but may not be subject to assessment. Thus, students have a number of core subjects to study which they may complement with elective courses.

37. According to Smyth and Hannan (2007), the possibility of subject choice is an important incentive for students. It allows students to focus their learning on topics of their personal interest, which is found to have positive effects on their probability to progress into higher education and to participate in education in the long term.

38. Not only the content but also the prescribed amount of subjects to be studied may vary. In England students may chose to graduate in only one subject, while Korea expects students to study nine subjects for graduation (Le Métais, 2002). In Germany, students are usually assessed in a minimum of four subjects. They may choose two advanced subjects and a third and fourth with a stronger weight on their final grade. The selection of these four subjects, however, is required to cover all main subject fields, such as linguistics, social sciences, mathematics and natural sciences (INCA Germany, 2005).

39. Students make their subject choice on the basis of several factors. Elective courses either allow to broaden or to deepen the student's field of study. Mostly, however, students tend to choose subjects from one field of study. Students are highly dependent on the grades they receive at graduation for their future career. Therefore, subject choices are often based on grading experience from lower secondary education (Lund, 2008). Moreover, less rigorous perceived subjects tend to be preferred by students over others in the expectation to have better chances for a higher grade (Bishop, 2006). A study by Smyth and Hannan (2007) conducted in Ireland furthermore shows gender difference in the choice of subjects, related to previous performance in lower secondary education. As girls often have better results during compulsory education, they are more likely to choose courses of higher levels than boys. Students' choice of subjects may also be influenced by assessment practices or requirements in the subject. According to a study by Fullarton and colleagues (2003), Australian students changed their course choices in response to changing assessment practices.

40. As a consequence of a large choice of subjects, certification at the end of upper secondary education is highly individualised. For students this means that they may freely select courses according to their personal interests and profile. In regard to continuation, the individualisation of certificates, however, may cause a number of difficulties. First, certain subject combinations may retain the student from continuing on pathways of another focus. In this case, the decision point for continuation lies already within upper secondary education. Second, comparability may be reduced. The challenge for education policy is to design flexible assessment systems within the boundaries of standardisation (Smyth and Hannan, 2007).

41. In Victoria, Australia, the Victorian Certificate of Education, issued for the completion of upper secondary education, relies on the student's subject choice from 43 study areas. For certification, a student has to accomplish a minimum of 16 units within two years. Subjects usually include four units. In Queensland, certification is similarly processed with 20 units to be accomplished. This procedure leaves students with a choice to tailor their studies according to their own interests (INCA Australia, 2008).

42. Ireland and England are among the countries which have the lowest prescription of subjects (Le Métais, 2002). Students may freely choose their subjects for graduation. However, some literacy and numeracy qualifications may be required. In Ireland, students have to take a minimum of five subjects but usually take seven. One of the subjects chosen must be Irish language (see NCCA website[1]). The certification of qualification at the end of upper secondary education is thus highly individualised.

43. Other countries like France, Ireland, Italy, the Netherlands, Sweden and Switzerland offer a division into specialisation tracks such as languages, social sciences or fine arts and music (Le Métais, 2002). All students have common core subjects but specific compulsory subjects according to their track of specialisation.

44. In the UK and Ireland, students may not only choose between different subjects, but also different levels of requirements within the subjects (Mueller and Wolbers 2003; NCCA). Students in Ireland may select ordinary or high level. This choice, however, can be a further precondition for certain career paths after graduation.

45. These choices may also determine students' opportunities for progression. While the German *Abitur* does not constrain the student's course choice at higher education, students in France or England may be constrained by their upper secondary study field when applying to higher education (INCA England, 2010: INCA France, 2008; INCA Germany, 2005).

46. However, individualisation of pathways also implies an enhanced challenge for certification of students' achievements. While pathways are expected to lead students to a specific educational or professional destination for which a distinct qualification is required, pathway changes make it difficult to track and certify competencies and skills acquired. The student no longer develops a clear profile for a particular occupation (Raffe, 2001).

[1] National Council for Curriculum and Assessment: The Leaving Certificate (Established), *www.ncca.ie/en/Curriculum_and_Assessment/Post-Primary_Education/Senior_Cycle/Leaving_Certificate_Established_/*

# 4. DESIGN AND GOVERNANCE OF ASSESSMENT FOR QUALIFICATION AND CERTIFICATION

47. This section describes a range of aspects related to the design and governance of assessment frameworks for qualification and certification. First, it describes the key purposes of assessment and qualification and certification. These include communicating student results to a range of stakeholders and thereby reducing information asymmetries as well as motivating students for high performance. Second, the section looks into the responsibilities of different actors within the assessment framework for qualification and certification. In particular, it describes how responsibilities are distributed between the school level and the education system level.

## 4.1 Purposes of assessment for qualification and certification

***Reducing information asymmetries***

48. The key purpose of assessment for qualification and certification in upper secondary education is its communication component. As certificates contain information for a student's progression into higher education or the labour market, they are a key factor for students' educational and career development opportunities.

49. Certification plays a central role in the reduction of asymmetric information between the student and stakeholders. For decisions of recruitment, employers and further education institutions have a great interest in finding out whether graduates have proficiency in the skills and knowledge required for their purposes. Thus, certification of upper secondary education programmes may have a key signalling function to reduce the costs of screening students' abilities for stakeholders. Students' credentials are often central in the communication of achievements.

***Signalling of competencies***

50. Certificates themselves are only a piece of paper, but they signal the presence of competencies which employers associate with increased productivity and which further education institutions perceive as evidence for the presence of skills and knowledge required to proceed in education (Martorell and Clark, 2010; Betts and Costrell, 2000). This means that an economic value is connected to certificates and therefore they play an important role for the future of students (Papay *et al.*, 2008).

*Signalling value varies across systems*

51. The signalling value and recognition of certificates varies strongly among countries. Even though certificates provide information on a student's achievements in school, stakeholders do not necessarily consider certificates as appropriate communication tools of knowledge and skills (Cedefop, 2010). As a result, further education institutions and employers often complement information provided through certificates by information collected through their own assessment procedures. To this end, educational institutions may create entrance exams and employers may measure the skills of applicants in assessment centres. This, of course, reduces the influence of certificates and increases screening costs for educational institutions and employers.

52. Depending on the country, certificates are often not the only signalling factor that may convey information on students' achievements. The extent to which certificates and individual grades are valued by stakeholders varies, and in some cases, the reputation of the attended school has a significant signalling value. Schools may be especially concerned about their signalling ability if the national education system holds schools accountable on the basis of their students' performance. Schools may spend additional attention to the signalling function of certificates if they are judged on the basis of students' progression (Woessmann, 2003). In countries where school results are published, rankings may add as a significant signalling factor.

53. The signalling value of certificates may be reduced if there is a system-wide lack of trust into certification as an institution (Cedefop, 2010). Those who use certificates to obtain information about graduates may perceive varying levels of competencies connected to one qualification standard and thus lose confidence in its reliability. Defective or complex description of competencies can augment the lack of confidence into certificates and induce ineffective signalling (Cedefop, 2010).

***Signalling for labour market employability***

54. Even though stakeholders have similar interests towards the signalling function of certificates, certain aspects need differentiation for employers and further education institutions. On the labour market, certificates may bring about earning advantages for students who successfully passed upper secondary graduation tests, as their diplomas confirm that students master a specific set of skills. Conversely, assessment for qualification and certification might negatively affect employment conditions for students who fail the examination, as the absence of a certificate might signal absence of skills to employers. Standards for grading or passing, which will be discussed in depth below, may thus have important implications for a student's signalling ability and earning chances on the labour market (Betts and Costrell, 2000).

55. Martorell and Clark (2010) analysed the signalling value of high school diplomas by comparing earnings of passing and failing students in the United States. But no strong effects were found, which could be a sign that other sources of information or mistrust in the value of diplomas may have an impact on later earnings. Moreover, employers do not seem to pay much attention to whether students received their diploma at the first try or only after several attempts. According to the study, companies seem to ask only rarely for high school transcripts and pay little attention to grades.

56. These findings are in line with a study by Bishop (1999), who found that Canadian employers do not seem to pay a lot of attention to grades in the selection of employees. Checking a small sample of job applicants, Bishop found that employers typically do not ask for grade point averages (GPA) or grades. The degree of emphasis given to grades can be a cultural component which may strongly influence the signalling value of certificates.

***Signalling for admission to higher education institutions***

57. For admission to higher education, the successful completion of upper secondary education is a minimum requirement in most OECD countries. In most systems, the actual grades obtained only have an importance for the selection into specific programmes. The level of standards determines the passing threshold which decides on students' progression to higher education. Higher standards in assessment for qualification and certification can form an obstacle for low-performing students and can contribute to reducing overall higher education admission rates.

58. With the development of mass higher education, many countries face the problem of having less higher education places available than students applying. In France, for example, the overall grade of the

*Baccalauréat* is increasingly important for access to higher education, especially to be accepted into a high-quality institution. In this context, students' choice of subjects in upper secondary education also has a certain signalling value. The choice of mathematics and physical sciences is considered as most prestigious and is thus highly valued in the admission process. Students sometimes opt to repeat a year to improve achievements and increase their chances for admission into certain higher institutions (INCA France, 2008).

***Motivating students for high performance***

59. Assessment for qualification and certification is often also intended to improve students' achievement by providing extrinsic incentives for high performance. Students may perceive assessment for qualification and certification as a concrete challenge to take. Usually, assessment for qualification and certification is connected to the permission to proceed with a certain career path, such as to higher education. This opportunity might motivate students for higher achievements than without such assessment. As assessment for qualification and certification is expected to increase the overall level of achievement, the proportion of students proceeding into higher education may be positively affected. But findings on incentive effects are diverse in international and cross-regional comparative and national studies (more on this below).

**4.2 Sharing of responsibilities for assessment for qualification and certification**

60. The division of responsibilities for assessment in upper secondary education tends to vary considerably among countries, but in most cases responsibilities are shared between teachers and a central agency. Assessment responsibilities encompass the design of tests, processing of assessment, grading and issuing of certificates. The main choice is whether the main responsibility rests with schools or central agencies.

61. Across OECD countries, there is a growing trend towards including centrally processed components in assessment for qualification and certification. This is the case for example in Australia, Austria, Denmark, England, France, Ireland, Italy, Norway and Slovenia (INCA Australia, 2008; Eurydice, 2009). In these countries, central agencies supervise or process assessment with the intention of increasing fairness and reliability.

62. In highly centralised systems, tests are designed, processed and graded by a national agency for all schools. In France, for example, all these tasks are state controlled. The development of test questions and the overall monitoring of assessment is the responsibility of the Ministry of Education. The *recteur*, who is the local representative of the Ministry of Education, holds the main responsibility for the processing the assessment. Despite this centralisation, the French system strongly relies on the principle of participation and thus teachers are involved in all processes (INCA France, 2008).

63. By contrast, teachers in decentralised systems execute all tasks of assessment by themselves, like it was traditionally the case in Germany (INCA Germany, 2005). Even though differences existed between the assessment procedures across the German *Länder*, the only tool for verification was the collection of some sample papers for central checking. By now, however, most *Länder* have introduced central examinations to some extent. In some *Länder*, assessment of several subjects may still be entirely processed by teachers, but the trend clearly moves away from decentralised testing.

64. In Japan, assessment for qualification and certification in upper secondary education is also largely the responsibility of teachers and there is no external moderation. School principals are the ones who issue the certificates. However, after completion of upper secondary education, students who want to proceed with education are requested to pass the *Juken,* which is the higher education entrance exam. The

Japanese qualification system thus relies rather on entrance than leaving exams to determine the pathways of students (INCA Japan, 2007).

65. Most education systems have divided responsibilities for assessment in upper secondary education between central agencies and schools. In Sweden, national tests in the core subjects are centrally designed by the Swedish National Agency for Education (Skolverket), but they are processed and graded by the students' own teachers (Nusche *et al.*, 2011a). The National Agency for Education provides extensive support to teachers and also makes sure that teachers are involved in the design process to make sure tests match the competencies of students. Moreover, teachers have access to a database, containing assessment material to support them in the preparation of students.

66. In Australia, all states rely on partly externally and partly internally processed systems of assessment in upper secondary education. However, through work programmes teachers receive strong guidance for the conduct of tests, especially in Queensland, where internal assessment has the highest priority (INCA Australia, 2008). In this State, the work of a sample of students, which is considered typical under specific criteria, is sent to the Queensland Studies Authority (QSA) at the end of Year 11 for feedback concerning the attainment of standards. Towards the end of Year 12, subject specific portfolios of students' work are sent to the QSA in combination with a judgement on quality. In case these do not match the standards, the panel invites the school for consultation. In a post-hoc analysis, the QSA collects random work samples for examination by a non-home district. This elaborated procedure is meant to leave some liberty of decision with schools and to simultaneously assure a high level of quality. The external moderators in Queensland play an important role in providing feedback to support teachers in the assessment process and ensure fair grading.

67. According to Mercurio (2006), a significant differentiation between assessors and moderators is made in Australia. "Assessors" are defined as teachers who assess achievement standards of their own students and "moderators" are teachers who asses achievement standards of other teachers' students. Each assessor in Australian upper secondary education has access to moderation support. In this relationship Mercurio (2006) sees the chance for intensive, professional, collegiate and supportive exchange, which in his regard encourages fairness. Mercurio (2008) considers a mixture of internal and external assessment as a way that strongly promotes fairness. However, in most states in Australia there is a stronger focus on external assessment (INCA Australia, 2008; QSA, 2010).

68. The distribution of responsibilities for assessment in upper secondary education also tends to depend on the subject. In many systems, including Austria and several German *Länder*, central institutions hold responsibility for designing assessments in core subjects such as mathematics, the language of instruction and foreign languages, while teachers hold the sole responsibility for other subjects. There is, however, a general trend for central agencies to take on assessment responsibilities in an increasingly broader range of subjects.

***Transparency and comparability aspects***

69. In terms of transparency, it is argued that externally processed assessment for qualification and certification allows a clearer communication of students' achievements (Cedefop, 2010; OECD, 2010a). Stakeholders are more likely to understand which skills to associate with a graduation certificate and to compare students' achievements. This transparency for the communication of skills also promotes mobility, as students' certificates will easily be recognised throughout the region (OECD, 2010a). Centralised assessment and standards may thus promote readability. In Germany great differences exist between assessment practices of different regions, probably reducing readability across the country. As a means to overcome this diversity within the country and to enhance common understanding of standards,

the Standing Conference of the Ministers of Education and Cultural Affairs of the *Länder* agreed on standard requirements for 33 subjects (INCA Germany, 2005).

70. On the basis of data from the Longitudinal Study of American Youth (LSAY), Betts (1995) argues that decentralisation, on the other hand, may lead to a lowering of standards. The reason is that employers and higher education institutions cannot adequately distinguish between the varieties of all school districts. As a result, districts have incentives to lower standards and to benefit from being pooled with other districts of higher standards. The signalling value of districts with lower standards is consequently above their actual level of performance and a free-riding issue emerges. As districts of higher standards have signalling power below their actual level of performance, they are likely to adjust standards to the lower. For employers and higher education institutions, these effects reduce trust into certification and make a comparison of certificates between regions unreliable.

71. Centralisation of assessment for qualification and certification inhibits free riding and a race to the bottom. It raises standards in low-achieving districts but lowers standards in high-achieving districts. Centralisation may thus imply negative reputational and motivational effects for some provinces. Betts (1995) therefore recommends the setting of minimum centralised standards combined with the option for provinces to set further higher standards. This way a race to the bottom is prohibited by a predetermined threshold while incentives to enhance standards are still given.

## 5. PROCEDURES FOR ASSESSMENT FOR QUALIFICATION AND CERTIFICATION

72. Establishing effective procedures for assessment leading to qualification and certification requires consideration of a range of possible options. First, effective assessment requires the establishment of reference standards and definitions of expected levels of achievement. Second, choices need to be made between different assessment instruments, such as continuous versus end-point assessment, and written versus oral assessment. There are also a number of issues connected to grading process, which influence the effectiveness of assessment in upper secondary education.

### 5.1 The role of standards in qualification frameworks in upper secondary education

73. Qualifications are defined levels of competency and are often regarded as the only means to recognise learning outcomes by providing a definition of what has been learned (OECD, 2010a). They are used as a proof of students' performance which is usually measured through some sort of formal assessment. Qualifications thus create standards in education, as they define levels of competency which can be assessed and certified for all students. Moreover, qualification standards enable comparability across regions or nations (Cedefop, 2010).

74. The creation of a qualification framework implies the ranking of these qualifications along with the establishment of an assessment system as a tool for comparability among students. The successful completion of one stage is regarded as evidence for readiness to proceed in education. Qualification systems thus function as a rank of standards, a framework for the structured progression of students (Young, 2003). Not only are routes of progression defined for the student, but levels of qualification provide strong extrinsic incentives for students (Cedefop, 2010). These standards define goals for personal achievement, as usually decisions for the future of the student are connected to the passing of these milestones.

75. In Ireland and England, for example, students are provided with a qualification framework which clearly defines levels of achievement to increase learners' flexibility in terms of orientation and pace (INCA Ireland, 2011; INCA England, 2010). In England, the learner is awarded credits, one counting for ten hours of study, for every qualification level he achieves. Nine levels, from entry to specialisation, are differentiated in England and ten in Ireland. The specific qualifications to be certified on each level are defined in a National Database of Accredited Qualifications in England (INCA England, 2010).

***Learning benefits of qualification standards***

76. Standards play a key role in qualification frameworks in upper secondary education. This is the period in which students' achievements significantly determine their future career and thus assessment for qualification and certification is an important milestone on the educational path of many students. Qualification frameworks provide a number of structural advantages for learning and teaching in upper secondary education. Through the ranking of qualifications on standardised levels of achievement, both students and teachers experience increased guidance in a study period characterised by high stakes and increased pathway differentiation.

77. For the structure and organisation of learning, detailed qualification frameworks may thus play a central role. Agreement on qualification standards throughout a region allows to set common learning

targets and to measure these at different stages of learning across an education system. This allows students to progress vertically to deepen knowledge and horizontally to broaden the scope of learning (Young, 2003). The measurement of achievements relative to an external standard and not within the class allows a fair comparison of students' achievements across schools and regions (Bishop *et al.*, 1997). It increases flexibility and mobility. But qualification systems also promote communication between all stakeholders as well as transparency. In this sense, qualification frameworks can also be perceived as a means for quality assurance (Cedefop, 2010).

78. On the basis of standardised degrees, graduates of upper secondary enjoy increased mobility. Employers and higher education institutions rely on highly readable and comparable certificates for the choice of graduates within a common qualification system. The presence of ranked qualification standards thus may supply students with increased guidance, flexibility and opportunities to signal their achievements.

79. Standards are assumed to have the potential of raising motivation and average achievements of graduating students as they raise expectations towards students' achievement (Bishop, 2006). On the basis of multiple studies by Wood and colleagues (1987) which focus on goal setting, Bishop (2006) states that the presence of central standards significantly raises students' performance in solving highly complex tasks. This study compares a cohort of students who were given clear achievement goals with other students who were only told to do their best. According to the study, the formulation of some goals raised achievement by 47%. All these studies are based on the assumption that students confronted with higher expectations learn more. Especially, external or national standards provide a means for comparison across a large group of students and allow students to set their own performance in relation to a larger group of peers. In the following the specific challenges and implications of externally processed assessment for qualification and certification will be discussed.

***Possible negative implications of qualification standards***

80. Standards may not only function as a tool for the improvement of learning. Negative motivational, structural or implementational constraints and effects may result.

81. The establishment of standards throughout a region requires consensus on the definition of those. Even though a growing region of common standards implies higher levels of transparency, mobility and recognition of certificates, it also implies the need of consensus and recognition among an increasing number of stakeholders (Swanson and Lee Stevenson, 2002).

82. According to Lee and Lee (2000), the creation of standards is often associated with additional workload for teachers and with very considerable cost implications. Given the impact on teachers' workload, opposition from teacher unions may arise (Lee and Lee, 2000). Further implementation challenges may arise in the creation of standards for performance assessment. Standards may lack appropriate assessment tools to be accurately applied (Hambleton *et al.*, 2000). The design of assessment tools which can catch the full range of skills a student is expected to learn and to make them comparable is one of the main challenges for standard-based systems.

83. Qualification frameworks may also be misused or shift priorities at the cost of other areas. As Raffe (2003) points out, the danger connected to the creation of standards is that societies may put too much emphasis on certification at the cost of learning. He also states that he sees a risk for an enhanced focus on vertical progression of students towards higher education, as the hierarchy of the qualification framework may suggest to the expense of vocational and community learning.

### 5.2 Continuous versus end-point assessment

84. Countries vary in whether students accumulate points or credits based on continuous assessment or whether they have high stakes end-point examinations, also referred to as exit examinations. Mostly, continuous assessment is internally processed, while end-point assessment tends to involve external processes. The reason for internal processing of continuous assessment might be the huge organisational and financial effort associated with externally processed assessment.

85. Some systems like Spain strongly rely on continuous assessment (INCA Spain, 2007). The successful completion of all courses results in receiving the *Bachillerato*. In Korea, the diploma of high school education is similarly based on students' achievements throughout their time in high school, but it has only very limited importance for the student's future career. For access to higher education, students in Korea have to pass the College Scholastic Ability Test (CSAT) which is neither directly connected to schools nor to higher education institutions (INCA Korea, 2008). This highly independent and standardised national test is an externally processed exam to replace completion assessment in upper secondary education by an entrance oriented approach.

86. Some systems which employ continuous assessment rely on credits for tracking the achievements of students over time and adding them up for graduation. Credit systems can be employed for different purposes. As a measure of achievement, credits may enable vertical or horizontal transfer within an education system. Furthermore, the accumulation of credits is a way of quantifying educational attainment and progression (Cedefop, 2010). Major benefits associated with credit systems are simplification, increased access, flexibility and quality assurance (Hart and Howieson, 2004). But also participation can be enhanced with new groups of learners being integrated into the system. Credit systems may allow students to study at their individual pace while respecting the needs of the learner (Tait, 2003).

87. Credits can provide a means to conduct continuous assessment over time, and to sum up a student's achievements for a final grade. Moreover, achievements can be weighted and thus educational priorities can be set. Credits can also be awarded for non-graded activities. In Sweden for example, students have to accomplish a project work, for which they receive credits for graduation (Swedish Ministry of Education and Research, 2010). Ungraded credits are especially valuable for the development of skills which are difficult to assess. Many countries rely on a mixed system in which the final grade is composed partly of continuous assessment over one or two years on the basis of credits and a final exam (Köditz, 2009).

88. In Queensland, Australia for example continuous, internal, subject-specific assessment, moderated by review panels constitutes one part of assessment for qualification and certification in upper secondary education, complemented by a cross-curriculum externally processed test (INCA Australia, 2008). In England students may collect credits throughout their learning career with one credit being equivalent to ten hours of study (INCA England, 2010). Also in New Zealand, great emphasis is given to the progression of students in a credit-based qualification framework. While receiving credits for the achievement of standards, students advance until being granted the National Certificate of Educational Achievement (NCEA) (INCA New Zealand, 2008; New Zealand Ministry of Education, 2010).

89. In most cases, as for example in the Netherlands, England, Scotland, Portugal or Greece, mixed systems of continuous and end-point assessment are used (Eurydice, 2009). In France not only a final exam, but also continuous assessment is centrally organised (INCA France, 2008). Despite the strong focus on centrality, a strong emphasis is given to participation, guaranteed through the involvement of all stakeholders and teachers from all over the country in the development of tests. In Hungary meanwhile, teachers experience high autonomy in continuous assessment (INCA Hungary, 2006).

90. It could be assumed that continuous teacher-developed assessment better matches the content taught in class and therefore provides stronger evidence on a student's level of achievement. Bishop (1999), however, argues that teacher developed tests focus on the assessment of lower order thinking skills to a much higher extent than tests developed by national or regional institutions.

***The role of exit exams in assessment for qualification and certification***

91. This section analyses the different types of exit exams and the impact that they may have on students' learning experience. The literature mostly understands exit exams as a one-time event at the end of upper secondary education. It describes a standardised, externally processed and centralised high-stakes exam to be taken by all students as part of their requirements for graduation. Many researchers discuss the role of exit exams at the end of upper secondary education focusing their analysis on different aspects (Dee and Jacob, 2006; Bishop, 1999). Among these, Holme and colleagues (2010) give a most comprehensive overview on issues connected to exit examinations.

92. But the term exit examination is not synonymous to assessment for qualification and certification in upper secondary education. Assessment for qualification and certification in upper secondary education is used to mark the successful completion of this educational period in general, while exit examination mostly only refers to a final exam taken at the end of upper secondary education. This section describes different types of exit examinations and considers some of their potential effects.

***Minimum competency exams (MCEs) and curriculum-based external exit exams (CBEEEs)***

93. In the literature from the United States, a differentiation is made between the certification of minimum competency exams (MCE) and curriculum-based external exit exams (CBEEE). While MCEs communicate only a single level of competencies to decide on success or failure, CBEEEs assess a large portion of what the student is expected to have learned. Certificates of CBEEEs convey a lot more detailed information and signal a much wider range of competencies (Bishop, 2006). In consequence, for certification on the basis of CBEEEs, the collection of assessment material is more complex in order to create evidence on the full range of levels of achievement. However, these supplementary benefits in terms of information provision largely contribute to the fact that CBEEEs gain increasing recognition in many countries.

94. Assessment for qualification and certification can either require the completion of a set of minimum skills or measure achievements on a scale. MCEs bear the risk that low standards discourage students to try any harder. If, however, standards are set very high, large drop-out rates risk to result. Many systems have therefore moved to assessment for qualification and certification which allows certification at different ranks and thus signal achievement at multiple levels (Bishop, 1999). As a result, literature predominantly discusses assessment on the basis of gradual grading and certification.

95. Meanwhile, CBEEEs assess a large portion of what the student is expected to have learned and thus a wider variety of a student's competencies may be signalled. This also implies that the design of assessment material for CBEEEs gains complexity. But most importantly, CBEEEs provide the necessary information to signal competencies on multiple levels while MCEs rely on a single level to decide on a student's success or failure (Bishop, 2006). These supplementary benefits in terms of information provision largely contribute to the fact that CBEEEs gain increasing recognition in many OECD countries.

*Implications for teaching in CBEEE systems*

96. High stakes are connected to CBEEEs. In consequence exit exams may have strong motivational implications for teachers, both positively as well as negatively. While the importance of quality teaching

may rise with the introduction of high stakes and teachers' role be redefined, narrowing of teaching content and fragmentation of knowledge could result as an unintended side effect.

97. Enhanced guidance and incentives through assessment for qualification and certification may have an impact on the role of teachers. Under CBEEEs, teachers experience different conditions for the conduct of their lessons. CBEEEs describe external assessment for the purpose of qualification and certification, which refers to multilevel and centrally set standards (Bishop, 1997b). In CBEEE systems, teachers may no longer form their expectations towards students according to their own ability of teaching (Bishop, 1999). Teachers may feel encouraged to raise the level of achievement and to meet the national level (Holme *et al.*, 2010). Benchmarks would no longer be set in the school or classroom itself but on a national level.

98. Nationally defined standards support teachers in providing students with clear information on what is expected from them, as well as in communicating needs for improvement (Rawlins *et al.*, 2005). Standards may thus offer an important guideline to teachers. Moreover, Bishop (1999) finds in his study on the basis of TIMSS data, that teachers in CBEEE systems are significantly more likely to make students do experiments in class. This positive effect was also reproduced in Canada in his study based on International Assessment of Education Process (IAEP) data (Bishop, 1999).

99. As teachers give up their function as an assessor, they take an entire different role in the presence of CBEEEs. Teaching input and learning outcomes are no longer divided, but teacher and student focus together on outcomes. According to Bishop's study on TIMSS data (1999), teachers in CBEEE systems spend significantly more time on student consultation and tutoring which might be a result of this changed role. The teacher's role changes from expert to coach who is fighting with his students on the same side (Bishop, 2006).

***Possible negative implications for teaching due to exit exams***

100. The creation of national standards to be assessed by an external exam bears the danger of teaching to the test, meaning the limitation of curriculum content in favour of the focused preparation of a student for examination. Limitations in the availability of assessment tools may further increase this effect, which is extensively discussed in different contexts by Looney (2011), Nusche (forthcoming), and Rosenkvist (2010). Teachers may neglect other study content which cannot be captured in tests (Lee and Lee, 2000; Volante, 2004). The consideration of "teaching to the test" effects becomes especially crucial and is found particularly strong at the upper secondary level, as testing is connected to high stakes for students (Au, 2007; Looney, 2009). Exit exams largely determine students' future career and teachers are sometimes even evaluated on the basis of their students' performance.

101. Under the pressure of high stakes for students and even teachers in exit exams, negative effects on teaching, especially in low-stake subjects are likely. Typically, mathematics, the country's language and foreign languages are among the high-stakes subjects, while subjects such as sciences, arts and history, classified as low-stakes subjects, risk to be neglected. Winters and colleagues (2010) describe this as an incentive problem for schools, which encourages teaching focus on certain subjects only. Teachers may feel encouraged or even obliged to shift resources from low stakes to high-stakes subjects (Jacob, 2005). This effect is described as a crowding out of learning outcomes in low stakes subjects.

102. In Florida, Winters and colleagues (2010) conducted a study which analyses the effects of high-stakes exams on such subjects which are not assessed for graduation. For students as well as for schools high stakes are connected to the performance on the test. Crowding out effects in science, which is a low stake subject were analysed in association with high-stake reading and mathematics assessment. However, Winters and colleagues (2010) did not find any significant crowding out effects in science. It could be even

assumed that some positive spill-over effects result. But data was collected in fifth grade which may mitigate conclusions to be drawn on the effects of exit exams.

103. As a further result of performance pressure related to exit exams, teachers are reported to disconnect logically assembled learning content and to fragmentise knowledge in favour of an increased focus on test content (Au, 2007). Because standards and therefore assessment too refer to single skills the acquisition of integrated knowledge risks to be split into disjointed learning units (Rawlins *et al.*, 2005). Thus, standardised tests lead to atomisation of learning. Standards are more suitable for the assessment of technical and practical than higher order thinking skills, as Rawlins and colleagues claim (2005). For teachers exit exams mean a much higher degree of stress, as their students' assessment is to a certain extent also an evaluation of their abilities. Young or low performing teachers may even tend to feel overloaded.

***Risks for decreasing educational attainment***

104. While exit exams are perceived as an important incentive for students to raise their efforts and believed to raise performance, low performing students may suffer a feeling of discouragement under the pressure of high expectations. As a result, numbers of drop-outs may increase. A careful weighting of incentives for increased efforts and disincentives for quitting may be needed to assess overall effects.

105. High standards combined with high stakes for students in upper secondary education assessment may raise the hurdle for low performing students to successfully complete upper secondary. There are three different scenarios how high standards may have negative effects on attainment. In a first scenario, some students might have to simply retake the exam and pass it at a later stage. This implies spending more time in education which means high costs for the students themselves and the school (Bishop, 2006). Others may experience repeated failure on the exit exam, despite making the effort of spending more time in education. After several tries these students may be forced to leave the school as the regulations require it or the student himself may decide to drop out as a result of discouragement. A third group of students may even drop out or switch immediately to less demanding programmes without even taking the exam in fear of failure (Papay *et al.*, 2008).

106. When students make the experience of failing, exit exams may no longer provide an additional incentive to try harder. The loss of motivation may lead to increased drop-outs. In particular, among low income, minority and low-performing students, numbers of drop-outs are high (Papay *et al.*, 2008; Ou, 2009). For the progression in education, such as into higher education, high standards may thus even have negative effects. Even though high standards for the assessment for qualification and certification may enhance overall students' performance by increasing motivation, higher numbers of drop-outs might mitigate these positive effects.

107. As a policy reaction some countries have special financial incentive systems in place, such as the New York City school system paying students for passing grades on the New York State Regents exams or the Baltimore City Public School District which pays students for their improvement of scores on state graduation exams (Angrist and Lavy, 2009). In a study in Israel, Angrist and Lavy (2009) analyse performance effects of financial incentives for low-achieving schools. They find that girls, especially those who perform just below the threat hold without financial incentives, spend significantly more time on test preparation and thus clearly perform better under the prospect of financial reward.

***Evidence from quantitative studies***

108. Reardon and colleagues (2009) analysed the issue of increasing drop-outs in reaction to exit exams which are highly standardised tests with high stakes for students. According to their study low performing students do slightly less persist in high school in the presence of exit exams, based on external

standards. Schools with such exit exams also have substantially lower graduation rates. In their study in California graduation rates decreased by 3.6% to 4.5% in response to the introduction of external exit exams. Moreover, Holme and colleagues (2010) argue that the introduction of exit exams provokes a reduction of student attainment and increases drop-out rates among at risk students in upper secondary education. The assumption that low performing students can be motivated to work harder in the presence of exit examinations can thus not be proven right.

109. According to Bishop (2006), exit exams do not have significant effects, neither on enrolment rates of 15-to-19-year-olds, nor on graduation rates from upper secondary or on the years spent in school. Warren and Edwards (2005) similarly analysed the introduction of exit exams and their consequences for graduation rates in the United States, but found no effects.

*International studies on the effects of exit exams*

110. Woessmann conducted several studies on the effects of external exit exams, relying on data from TIMSS and PISA (Woessmann *et al.*, 2007). On the basis of TIMSS and TIMSS-Repeat data, Woessmann (2003) finds that students from countries with curriculum-based external exit exams performed significantly higher in mathematics and science, controlling for students' performance level as well as family background. On the basis of data from PISA 2003 he finds further evidence that students in a country which rely on CBEEE systems perform higher than in other countries. Even though he controls for student characteristics, family background, the location and resources of the school, GDP per capita, expenditure per student, autonomy in staffing and budgeting as well as the proportion of private and governmental schools, students in systems with curriculum-based external exit exams perform higher than students elsewhere by what an average student learns within half a year (Woessmann *et al.*, 2007).

111. Bishop (1997a; 1999) supports the findings of Woessmann, referring to a number of studies analyzing the effects of CBEEEs on students' achievements. He also refers to TIMSS data and finds positive effects for students' performance in mathematics and science (Bishop *et al.*, 1997). In a test based on the data of the 1991 International Assessment of Educational Progress (IAEP) which contains data on the performance of 13-year-old students in mathematics and science for 15 countries, Bishop (1997a) finds large positive effects of CBEEEs on students' performance.

112. In mathematics, students achieve significantly higher results in systems with CBEEEs (Bishop, 1999). The variance of student achievement is not affected by the presence or absence of CBEEEs, according to his study. In science effects were insignificant. Relying on IAEP data, Bishop (1999) analysed in addition effects in geography, even though for this subject data for only nine countries is available. However, no significant effects were found here.

*Cross-regional studies on the effects of exit exams*

113. The IAEP provides also province specific data for Canada which allowed Bishop (1999) to conduct a cross-regional analysis on effects of CBEEEs within the country. He finds that those provinces with a CBEEE at the end of high school perform significantly higher both in mathematics and science. Even when endogenous school policies and inputs, the amount of homework assignment and the time of watching TV are held constant, effects on performance in mathematics are still significantly positive (Bishop, 1999). For science, however, results decrease to insignificance.

114. According to a study from Betts and Costrell (2000), graduates from systems with MCEs as analysed in Massachusetts and California, which also rely on centrally assessed standards, earn about nine percent more in the calendar year after graduation than graduates who do not take such a test, despite controlling for high school completion, characteristics of the local labour market and college attendance.

115. Other than probably expected, the size of classes is not smaller and time used for the preparation of classes is not longer in CBEEE systems. Bishop (1997a) comments that those provinces which have strong exam systems in place may simply give higher priority to education.

116. However, direct effects of CBEEEs in upper secondary on students' motivation and performance may be questioned. As TIMSS data focuses on 13-year-old students, a direct link between CBEEEs in upper secondary and performance changes of much younger students may not be established. Effects for younger students may rather result from overall priority changes in the education system, induced by the introduction of CBEEEs.

117. On the basis of data from 37 states of the United States collected in 1993 by Graham and Husted, Bishop and colleagues (1997) analyse the education system of New York. Again, positive effects of CBEEEs on students' performance are found.

118. Research in the United States on MCEs has shown some motivational effects on low performing students. Low performing students in systems with MCEs performed a lot higher than students who were not taking such a test (Holme *et al.*, 2010). But, most countries' testing culture has grown above the assessment of minimum competencies and therefore these results may not hold for other systems.

119. Reardon and colleagues (2009) analysed the effects of exit examinations in California. For this purpose they compared two different cohorts of students, one in 2003, the other one in 2004 or 2005, with observable similarities. Both cohorts undertook the California High School Exit Exam (CAHSEE) test in their tenth grade under the belief that it would contribute to their graduation. While this was the case for the 2004 and 2005 cohort, and their test results were part of the graduation requirements, this was not the case for the 2003 cohort. Thus, the two groups of comparison differ in their experience after tenth grade. However, from their results Reardon and colleagues (2009) could neither conclude on positive nor negative effects of exit exams on students' achievements at high school.

120. Similarly, the study of Jurges and Schneider (2010) in Germany did not find any effects of exit exams on students' performance in mathematics. Students in provinces with and without exit exams achieved at equal levels. However, empirical evidence from Germany shows that a certificate based on performance in a national assessment is a better predictor of actual productivity than a diploma obtained in a local assessment (OECD, 2010a).

*Factors of influence*

121. Bishop (1999) suggests the following possible reasons for his findings, based on data from the questionnaires. First of all, parents in provinces with CBEEEs talk more to their children about mathematics and have a strong desire for their children to perform well. Moreover, schools in provinces with CBEEEs schedule more hours for mathematics and science education, assign more homework, were equipped with better science labs and had more specialised teachers (Bishop, 1997a). Also, the time spent in front of the TV is significantly lower and the use of computers for schoolwork higher among students in CBEEE systems (Bishop *et al.*, 1997).

122. Moreover, Bishop (1999) tested performance effects in dependence of time spent on reading for pleasure and watching science programmes. On the basis of TIMSS data he found no effects among students who like mathematics and science. In Canada, however, based on IAEP data, students in CBEEE provinces were surprisingly found to spend more time reading for pleasure and watching science programmes while they overall spend less time watching TV.

123. Assessment for qualification and certification, especially if processed externally, has also a certain influence on spending priorities taken by schools and pedagogy (Bishop, 1999). Resources are

redirected and learning as well as teaching for subjects which are assessed gain considerably more attention. Despite expectations, however, countries with CBEEEs do not have significantly higher spending on primary and secondary education overall as the IAEP data shows (Bishop, 1999). Nevertheless, TIMSS focuses on student performance in mathematics and science only, just as the IAEP. Therefore, no evidence for effects in other subjects is found in these studies.

*Other effects of central exit examinations*

124. On the basis of international and cross-regional data not only motivational effects were analysed. Studies found a number of further positive as well as negative effects of external exit exams.

125. Students who are subject to external assessment may change their behaviour in many respects under central exit exams as a reaction to changing incentives. Bishop (1999) believes if students are assessed relative to outside standards, they no longer have an interest in slowing down teaching or detain other students from studying in the hope that the amount of content assessed will be reduced. Teachers and students cooperate more and students are no longer motivated to try to lower standards (Bishop, 2006).

126. An additional benefit of exit exams is a potential reduction of "nerd harassment", as all students want to pass, and standards are not set in the school but enhance external comparability (Bishop, 1999). High performing students thus may no longer stick out among their class mates, as their level of achievements is seen in a larger context.

127. Meanwhile, external exit exams may raise stereotype effects. While comparing high stakes and low stakes exit exams in California, Reardon and colleagues (2009) analysed the impact of stereotype threat. Stereotype threat describes the risk that a person internalises low expectations connected to his negative stereotype, performs accordingly on high stakes exams and thus confirms expectations (Holme *et al.*, 2010; Steele, 1997). This issue is usually related to ethnicity or gender. Reardon and colleagues (2009) confirm in their study on the basis of data from the California High School Exit Exam (CAHSEE) that Blacks, Hispanics and English learners do perform on a lower level on high stakes exams. Minority students performed disproportionately much worse while no performance effects were found for white students. Moreover, girls were found to underperform in mathematics exams. Reardon and colleagues (2009) conclude that the diploma does thus not produce a reliable signal of students' competencies to employers for minorities and girls and is not neutral.

128. In the attitude of students, Jurges and Schneider (2010) find that those who take central exit exams enjoy mathematics less and find it more difficult and boring. At the same time, these students do only achieve slightly higher in test scores while receiving more homework. These findings are based on TIMSS data, comparing students of centrally and locally processed exit exams.

129. Surprisingly, a study of Bishop (1999) reveals that students in CBEEE systems are more unlikely to belief that memorisation for learning mathematics and sciences was necessary which could be a strong argument to deny effects of teaching to the test (Bishop, 1999).

130. As a consequence of stronger signalling ability, students who are subject to external exit exams are believed to have higher earnings after graduation (Bishop, 2006). On the basis of data from 37 states of the United States, collected in 1993, Graham and colleagues tested this effect (Graham *et al.*, 1993 in Bishop *et al.*, 1997). As Bishop and colleagues (1997) state on the basis of this study, even while controlling for student achievement, CBEEEs seem to have positive effects on earnings of graduating seniors.

### 5.3 Formats of assessment for qualification and certification

131. The choice of assessment tools faces great pressure to guarantee reliability and fairness as assessment is connected to high stakes for students at upper secondary. Countries show a certain variety of how they handle this concern.

132. Formal examination and thus also examination for the purpose of qualification and certification in upper secondary is mostly written as for example in Greece, Portugal or Finland (Eurydice, 2009). Reliability and fairness seem to be highest under this form of assessment. But certain skills might not be assessed with this tool. Therefore, many countries like the Czech Republic, Denmark, Germany, France or the Netherlands conduct examination for certification of upper secondary education both written and oral (Eurydice, 2009; Köditz, 2009).

133. While external assessment mostly relies on written, time-bound assessment under standardised conditions, internal assessment allows more flexibility. In Queensland, Australia, a number of assessment tools such as short tests, projects, field studies, practical and aural tests, assessing understanding of listening, under various conditions, including supervision, open-book and handing-in are used to assess different aspects of learning (INCA Australia, 2008). This section will discuss the nature as well as advantages and disadvantages of used forms of assessment.

*Written examinations*

134. Race (2007) reveals several aspects to be considered in connection with written examinations. There are a number of advantages:

- Written tests are time and cost efficient. Only one set of tasks needs to be designed and all students can be assessed at the same time.

- They allow a clear definition of the field a student is expected to study, while in project work or essays the student only focuses his learning on a specific topic.

- As questions or tasks are the same for all students and preparation usually takes place commonly in class, a high level of equality of opportunity can be achieved.

- Teachers perceive written examination as the most common tool of assessment and therefore feel highly confident with it.

135. But written examinations also imply several disadvantages for which other forms of assessment are sometimes chosen to compensate.

- Written examinations produce a strong form of extrinsic motivation for students and do not necessarily promote the desire for learning. Students usually try to guess the topics being assessed and concentrate their learning on selective topics and methods to successfully pass the exam. Surface learning is encouraged with students clearing their mind after having passed the test to concentrate on the preparation of the next exam.

- Handwriting and neatness are often included in grading. Therefore, the content risks getting out of focus. Often exam papers are marked in a rush and without sufficient attention as it can be a very monotonous and tiring work. This further contributes to a loss of content focus.

- If exam sheets are corrected by different staff, reliability and comparability may decrease. Correctors do not always set the same criteria and students may receive different grading according to their corrector.

- A number of skills, which are increasingly included in curricula such as teamwork, leadership or lateral thinking, are difficult to assess in written examinations.

136. As students tend to highly focus their learning on expected tasks of assessment, successful learning outcomes on written assessment require adequate design of tasks. Otherwise, fragmental and surface learning may be encouraged. But some of the negative effects of written examinations can be attenuated. Preparation of questions in teams diminishes ambiguity of language and increases validity. Question design which ensures the integration of all learning units encourages comprehensive learning. The formulation of clear answering criteria can prevent disagreement in grading among different correctors.

*Oral examinations*

137. Some systems choose to enlarge their range of assessment methods, mostly with oral exams. But oral testing implies the risk of missing consistency in the assessment of different students (Miguel and Larson, 2006). In Germany, besides selecting their two advanced subjects, students select a third subject in which they are orally assessed in some cases and a fourth subject for oral assessment only (INCA Germany, 2005). In France students have to take seven examinations, of which each is possibly complemented by an oral part (INCA France, 2008). Also practical assessment tools are used in specific fields of study. Similarly, students in Italy have to pass one oral exam, dealing with various subjects. But also continuous assessment in Italy partly relies on oral assessment (INCA Italy, 2008).

### 5.4 Grading issues at upper secondary level

138. Grades in lower secondary education often have no larger importance than to inform student and parents on the level of achievement. In some countries, grades may also decide over year repetition, but usually no higher stakes are connected to grading.

139. In upper secondary education, however, grades do not only determine progression within school but also the student's career and play a major role in signalling a student's abilities to higher education or the labour market. In many educational systems students strongly depend on their grades for access to higher education or alternative career paths. Judgements on students' achievements determine their future career options and are therefore crucial to rely on fairness. This section will therefore deal with different aspects of grading, including fairness, students' perception and implications of grading standards.

***Reference points for grading student achievement***

140. Grading implies both the measurement of achievements and its comparison with the achievements of others. While the methods for measurement strongly rely on the tool of assessment chosen, the comparison part of grading can be achieved in three different forms, namely criterion-referenced, norm-referenced and individual-referenced grading (Nusche, forthcoming). These definitions differentiate whether the reference is made to a standard, to a group of students or the student's previous achievements (Dalberg *et al.*, 2007; Lee and Lee, 2000; Rosenkvist, 2010).

141. For the progression of students into higher education, norm-referenced grading may be used for the selection of students into limited higher education places. This is for example the case in Australia, where students are centrally certified and ranked for entry to tertiary education (INCA Australia, 2008). In

terms of fairness this procedure requires that all applying students are part of the same reference group. In Germany, quotas are set for certain courses, where the number of applicants is much higher than places available (INCA Germany, 2005). But waiting time and social criteria are also taken into account to guarantee fairness. In France, the Baccalauréat général theoretically provides access to higher education, but with increasing numbers of students and limited higher education places this is no longer guaranteed (INCA France, 2008).

142. Norm-referenced grading is also criticised for encouraging competitive behaviour among students (Guskey, 2000). Cooperation for learning becomes a threat for students and assessment divides students into a group of winners and losers. Moreover, Guskey criticises that a very high grade does not provide any information on the actual level of performance, especially in case of small reference groups. Consequently, a failing grade does not necessarily prove students to be low performing, but at the same time has major implications for their future.

143. In practice, assessment often relies on a mix of different references. Especially norm- and criterion-referenced methods tend to be mixed. In Portugal for example, the successful completion of upper secondary is a general criterion for access to higher education but a *numerus clausus*, depending on the number of applicants and the number of places available, decides on a student's admission. Danish students are continuously assessed in reference to national criteria but have to take end-point exams to complete their upper secondary education (Danish Ministry of Education, 2010).

144. Even though teachers rely on standards for assessment they tend to compare a student's performance with those of his peers. Sometimes, the class or the school is the only group of reference for teachers to define levels of performance. Moreover, students' individual improvement over time is likely to be rewarded by teachers with a good grade.

***Fairness in grading***

145. Fairness is a major concern in schools, especially if high stakes are in for students. Students expect to be treated with fairness by their teachers. Not only for certification fair grading is of central importance, but the experience of fairness also has major implications for a student's future trust in societal institutions and may even have effects on his behaviour (Dalberg *et al.*, 2007).

146. In many systems grading refers to students' level of achievement. Nevertheless, research shows that teachers tend to take into account other factors than performance (Bishop, 1999; Randall and Engelhard, 2010). Bishop (1999) assumes that teachers also take into consideration the effort students make and may neglect fairness while grading them. Remesal (2011) adds that good behaviour and extra-credit work are hardly being ignored by teachers in their grading decision. Especially, when decisions on passing or failing have to be taken, effort becomes an important decision criterion. A number of aspects such as homework completion, participation in class, preparedness for the class, notebooks or attentiveness serve as indicators for effort. The resulting problem is that it becomes difficult to know what a grade stands for. Moreover, behaviour is often taken into account for grading. The conduct and attitude of students may have both positive and negative effects on their grade.

147. CBEEEs, as Bishop (1999) states, may reduce visible student effort describing attempts undertaken by the student to impress teachers. Instead, students may increase invisible efforts for success on the test itself. In the absence of external assessment, students often try to present themselves to their teacher in a positive way in order to raise their grades. When CBEEEs become the only source of assessment for grading, this mechanism vanishes which might encourage a shift from visible to invisible efforts. However, in his cross-regional study in Canada, based on IAEP data, Bishop (1999) could find no

evidence for such an effect. But grading in Canadian provinces also relies on supplementary teacher-based grading so that there is no incentive for students to entirely abandon strategies for impressing teachers.

148. The assessment of other students' characteristics distinct from academic achievements, such as punctuality, can be highly desirable for certification. Especially employers have a specific interest in non-academic achievements of graduates. But, when such characteristics are included in general grading, confusion about the meaning of grades may emerge. For certification appropriate information for stakeholders may no longer be guaranteed. In Italy students are therefore additionally assessed on their attitude and behaviour for which they have to achieve a certain grade in order to graduate (INCA Italy, 2008). Similarly in Norway, students receive a distinct grade for order and conduct which is displayed on their leaving certificate (Norwegian Directorate for Education and Training, 2011). In a study conducted in the United States, Randall and Engelhard (2010) found that teachers largely take into account student behaviour in grading. On the basis of these findings they suggest that final grade assessment should receive more attention in teacher education.

***Student perceptions of fairness in grading***

149. Even though educators would wish that students are entirely intrinsically motivated, grades are an important tool for extrinsic motivation. Especially in post compulsory education, when grading determines students' pathways the pressure to achieve high grades increases. While grades may provide a strong incentive to students to enhance achievements, often they are perceived as a form of punishment, increasing stress and pressure on students.

150. Many students are driven by the fear of failure or of not meeting grading criteria. Sometimes, students who do not meet grading standards are given extra work to improve or they have to attend summer courses, which can be perceived as a punishment (Guskey, 2000; Jacob, 2005). For low performing students, pressure of grading standards may even discourage them and result in dropping out (Papay *et al.*, 2008).

151. According to a study of Fan and Chan (1999) among students from China and Hong-Kong, students aim to be evaluated equally, meaning according to norm or criteria-referenced grading. While the degree of equality in norm-referenced grading largely depends on the size of the group of reference this method might seem more transparent to students, as Dalberg and colleagues (2007) find in an analysis of students' perception of grading fairness in Germany.

152. The study of Dalberg and colleagues (2007), moreover, reveals that students consider criterion-referenced grading most just. While individual-referenced grading is also perceived as rather fair, norm-referenced grading is considered as unjust. However, the questionnaire presented to students is designed in a way that norm-referenced grading relies on the class as the reference group (Dalberg *et al.*, 2007, p. 425). This is the smallest possible entity for norm-referenced grading and thus provides the least possible comparability. If the questionnaire would refer to the whole administrative region as reference group, students might have a higher perception of justice in norm-referenced grading.

153. The study also shows that students with higher grades value criterion-referenced grading higher in terms of justice. No connection was found between the subject assessed and justice perception of grading. However, the study only takes into account core subjects, such as German language, mathematics and a foreign language. If core subjects would be compared with other subjects, students' judgement may change.

154. Further influence on students' perception of fairness in grading may derive from the individual cultural or educational background. Children, even with the same educational background have very

different perceptions of justice in education. In many cases the personal performance, students' self-esteem as well as their personal relation towards the teacher influences their sensitivity towards justice. Furthermore, the perception of justice is strongly culture-based and cannot be generalised for all countries (Sabbagh *et al.*, 2006). While some students emphasise that performance should be subject to rewards, others belief that grading should rather focus on efforts made (Dalberg *et al.*, 2007).

***Grading standards***

155. Grading, if it is not norm-referenced, relies on standards, defining clear criteria for the measurement of students' achievements (Guskey, 2000). But these criteria can vary strongly. High standards, as discussed earlier are associated with strong incentives and better performance but also higher drop-out rates. Lenient grading, however, is both argued to lower students' efforts as well as to positively reward and encourage students for continuation (Montmarquette and Mahseredjian, 1989). The assumption of this effect is particularly important for assessment for qualification and certification at upper secondary, as students may feel enhanced pressure to perform well for their desired progression into further training. Simultaneously, some students may also feel discouraged, reduce efforts and drop out.

156. Several studies analysed the effects of an increase in grading standards on students' performance. Two major effects are observed in the research undertaken. One is a sorting effect and the second is an incentive effect (Betts, 1995). Sorting describes changes in student pooling depending on the placing of the threshold line for passing and failing or for different levels of grades. Incentive effects describe students' choices in their adjustment of efforts. They may significantly influence the sorting effects of shifting grading standards. These effects will be explained in detail in the following.

*Raising grading standards*

157. In terms of sorting, shifting up the cut-off line means a decrease of those who pass the test and an increase of students failing the test (Betts, 1995). Enhancing grading standards has a similar distribution effect with fewer students receiving high grades and more students receiving low grades.

158. But once incentive effects are taken into account, a shift of the cut-off line means more than a simple relabeling or pooling of students. Queensland, Australia set up minimum standards with the introduction of the Queensland Certificate of Education in 2008, based on the attempt that these will increase students' motivation and achievements (INCA Australia, 2008).

159. When increasing grading standards high performing students will raise their efforts to maintain their grades or even not change their behaviour and still receive the highest grade (Betts and Costrell, 2000). These students show little or small changes in their effort and performance, but this group of passing students or highest performing students may shrink and thus their diploma gains more value. Belonging to this smaller group signals a higher average of achievement.

160. Low performing students with a level of effort close to zero, however, are likely not to change behaviour, but to keep efforts low as they may expect to stagnate in the group of failing or low performing students anyway. Their pool is likely to grow as the number of drop-outs will increase. Students at the bottom of the performance scale will now be pooled with more students performing better than those which may increase the signalling value of their group (Betts, 1995).

161. Students, who performed just above the cut-off line in the initial situation, may fail in consequence of a rise of standards. For these students costs to increase efforts may be too high to remain within the pool of passing students. Consequently, they might even feel discouraged and reduce their efforts (Betts, 1995).

162. Betts (1995) describes a further group of students, those consisting of students who, in terms of sorting, are just below the new cut-off line for passing. For these students costs to enhance efforts in order to maintain the initial grade are rather low and thus a rise of standards might incentivise these students for higher performance.

163. If students are indifferent between two grades initially, usually represented by a small number only, they will change their preference towards the lower grade (Betts and Costrell, 2000). The level of achievement will remain unchanged, but in terms of performance signalling, these students will be disadvantaged, being pooled together with students who perform below instead of above their own level. As discussed earlier it is assumed that grades have a significant impact on wages and chances for employment in regard to their signalling function. As grades of these students will drop, a rise of standards might also be connected to lower wages after graduation.

164. The likelihood of each of these scenarios depends on the distribution of students among these different groups and what their reaction will be towards an increase of grading standards. Betts and Costrell (2000) state, that the largest fraction of students is likely to increase efforts and performance. Students work harder and learn more.

165. On the basis of the Longitudinal Study of American Youth (LSAY), including data from around 6 000 students in secondary education, Betts and Costrell (2000) analyse grading standards effects on students' effort and performance and find positive motivational effects for most students. Grades are found to be more influential for the level of achievement than are the class size, the teacher's experience or his level of education.

166. Figlio and Lucas (2003) support these findings. They examined grading effects among primary students under teacher grading and find strong benefits for students in relation to high grading standards. They note, however, that unobserved teacher characteristics may influence the results and of course students at upper secondary may respond with slightly different behaviour. Similarly, Lillard and DeCicca (2001) assume that higher standards may lead to higher wages, employment probability and enhanced success in higher education. But the downside of increasing standards is that drop-outs significantly rise (Lillard and DeCicca, 2001). While higher requirements signal higher performance for some students, they also signal lower achievement for those who do not pass the threshold.

167. A study of Bonesrønning (1999) in upper secondary education further finds particularly strong effects for high achieving students. Fullarton and colleagues (2003) add that for the choice of subjects grading or the achievement level of students is an important factor of influence. Dryler (1999) distinguishes, that not the student's grade level but his relative standing in comparison to the class is a strong determinant for his motivation and choice of the subject. This means that grades can also be an important factor for students' orientation and continuation in education.

168. Meanwhile, the distribution of students may change in response to shifting standards. While the group of top students shrinks, the group of lowest performing or failing students grows. The effect is the formation of some sort of small elite with a higher average productivity and a stronger signalling value connected to their diploma (Betts, 1995). Simultaneously, the degree of stigmatisation of failing or low performing students relies on the average performance of this pool of students. As this group grows, the stigma may decrease (Betts, 1995). Students, who would not have graduated in the first place, may benefit from the enhanced signalling value of their group. Nevertheless, failing has no educational value and can imply strong effects of discouragement (Guskey, 2000). This category may thus be considered for elimination.

169. Especially in schools where parents show active interest in the education of their children, teachers feel pressured to keep standards up and to grade students on the basis of high expectations. Betts and Costrell (2000) also find that in the presence of higher standards, students spend more time on homework. This effect is particularly strong in mathematics. Results, however, have to be taken with caution, as parents are likely to have an imprecise observation on their children's time allocation to homework and as time allocation is an inexact measure of effort.

170. One reaction to increasing divides among students' levels of achievement in some states of the United States is to issue certificates of attendance for those students who attended high school but did not complete the degree (Betts and Costrell, 2000; Lillard and DeCicca, 2001). In most countries this is a group of students who has no means of signalling, even though these students may have attended upper secondary education for a number of years and developed skills during this period. With a locally issued certificate students could signal some non-cognitive skills, such as persistence, discipline or punctuality. Many of these non cognitive skills are highly valued by employers. Critics say that such locally issued certificates may raise confusion among stakeholders if they are not clearly distinguished from diplomas, issued on the basis of achievement standards (Betts and Costrell, 2000). A clear terminological differentiation between them, such as "diploma" and "certificate" may mitigate this problem.

*Lowering grading standards*

171. Conversely, lowering of grading standards is believed to have adverse effects. Grade inflation may decrease the effort and achievements of most students (Betts and Costrell, 2000). Especially among highest performing students, incentives to maintain their level of effort disappear.

172. Grade inflation is primarily a problem in competitive school systems which underlie teacher grading. Often the performance of schools is judged on the basis of students' grades. Teachers may thus be incentivised to give better grades in order to send positive signals to parents and society (Wikstrom and Wikstrom, 2005). Moreover, grades may be of high importance for the progression of students to higher education or on the labour market which may even increase teachers' motivation for giving high grades. According to Bishop and Mane (2001), 30% of American teachers feel pressured by parents or principles to give higher grades to students than what they consider to be appropriate. But for the selection of students who access higher education, grade inflation can lead to considerable mismatch (Wikstrom and Wikstrom, 2005). For assessment for qualification and certification at upper secondary level many countries therefore established central and standardised institutions to guarantee unbiased grading.

173. Nevertheless, many systems rely on mixed structures with one part of the final grade determined by teachers and another through an external institution (Köditz, 2009). Some other countries issue two certificates of which one only contains the grades provided internally by teachers and the other certificate communicates results from external assessment (Köditz, 2009). This is for example the case in the Czech Republic, Hungary, Poland, Slovakia and Finland. For the role of teachers on grading, grade inflation is thus a crucial risk to be taken into account. As a result, it could be important to introduce school internal mechanisms which prevent grade inflation.

174. Wikstrom and Wikstrom (2005) conducted an analysis on the effects of school competition on grade inflation in upper secondary schools in Sweden. For public schools they only found small effects of grade inflation while in independent schools, which are believed to be under higher pressure for the attraction of students, significant effects of grade inflation could be measured.

***Certificate formats and grading scales***

175. The certificate itself is a vehicle to convey information on a student's achievements. Grading scales and the design of certificates may therefore significantly influence the amount and the focus of information conveyed.

176. The main choice in determining the scale of grading lies between a binary system which only distinguishes between passing and failing students and a scaled system, supplying more information on the actual performance level of students. Moreover, such a scale can have different degrees of preciseness, depending on the number of levels available and the distance between them. For certification, the choice between different types of grading scales determines possibilities of signalling students' achievements to stakeholders.

177. Supporters of a binary system, as it is practised in Japanese upper secondary education, mainly argue that outcomes are more egalitarian (Betts, 1995; INCA Japan, 2007). In such a system, only two pools of students are differentiated and thus only two different categories of signals can be sent to employers or further education institutions. While binary systems clearly provide less information than scaled systems, it is questionable whether employers and higher education institutions actually use the enhanced information provided through scaled systems. Betts and Costrell (2000) find that additional information often remains unused. In binary systems, credentials other than the graduation certificate, such as artistic and musical achievements, may gain significance for the signalling of student competencies and impact on students' chances for admission to higher education or employment.

178. Scaled systems provide more fine-grained information. Different scales of accuracy are possible. In Norway, grades reach from 1 to 6, similar as in Queensland, Australia where grades reach from A to E (INCA Australia, 2008; Norwegian Directorate for Education and Training, 2011). In Tasmania, meanwhile, grades are only differentiated from A to D. In Victoria, a further differentiation is made with A+, A, B+, etc., which allows more precise grading. In France and Portugal, grades range from 0 to 20. In addition to the degree of differentiation, the naming of grades may be considered. While alphabetical letters are very neutral in their description, grades in Sweden for example are differentiated between "Pass with special distinction", "Pass with distinction", "Pass" and "Fail" (Swedish Ministry of Education and Research, 2010).

179. Scaled systems may help increase the efficiency in the matching of graduates and employers or further education institutions (Betts, 1995). The whole range of students from high to low qualification levels can signal accurate information on their level of competency, which supports efficient job-matching. The implications of scaled systems for motivation are similarly important. For students at the top end of the performance range, scaled systems provide an important incentive as their achievements would receive no recognition in binary systems. Students at the bottom of performance measurement similarly may only feel motivated to improve in scaled systems, where their efforts may give them access to a higher grade.

180. Certificates may also include course descriptions so as to further increase the level of information conveyed. In Tasmania, Australia, certificates include a 75-word description of the 12 most recently studied subjects to increase readability (INCA Australia, 2008). While a short text may add important information, overloaded certificates risk reducing readability.

# 6. USE OF RESULTS

181. The results of assessment for qualification and certification are of major importance for students' progression. On the basis of their performance, decisions such as entrance into higher education or chances on the labour market may be determined. This section will present the most important uses of assessment results.

## 6.1 Entrance to higher education

182. In almost all OECD countries, assessment for qualification and certification in upper secondary education serves to some extent a selection function for access to higher education. But different scenarios are possible.

183. In some countries, the successful graduation from upper secondary school is a minimum requirement for entrance into higher education as in France, Germany and the Netherlands (Le Métais, 2002). This does not mean that students are necessarily automatically admitted to higher education, but institutions of higher education and faculties may have their own supplementary entrance exams. In the Flemish Community of Belgium for example, medical and art students have to take entrance exams in addition to secondary leaving examination (Flemish Ministry of Education and Training, 2010). As Perna and Thomas (2009) report, some test takers in the United States do not see a link between exit exams and higher education entrance exams. William (2010), however, argues that high school grade point average seems to be the best prediction for a student's success in higher education. Nevertheless, he sees the Scholastic Aptitude Test (SAT), which is an admission test for higher education in the United States, as a useful supplement for the evaluation of students' qualification. Entrance exams may thus test for further competencies which are relevant for the specific subject to be studied at higher education.

184. Other countries as Ireland, England, Korea and Australia directly use assessment for qualification and certification in upper secondary for selection into higher education (INCA Korea, 2008; Le Métais, 2002). In a further group of countries, upper secondary certification may serve a selection purpose only for subjects which are on high demand.

## 6.2 Entrance to the labour market

185. Even though the labour market does not have any minimum entrance requirements, students who do not chose to continue in higher education or another educational programme, but to apply for a job may undergo a similar process of selection. If the certificate of upper secondary education is perceived as a trustworthy institution for the communication of students' skills, employers may strongly base their selection of employees on the information provided through the certificate. Thus, certificates may have a strong screening function for employers. Reducing information asymmetry, employers may use certificates as a reliable tool to decide whether the applicant matches the requirements.

186. The competencies which students develop during upper secondary education are important to be signalled to the labour market in a comprehensible and suitable manner. As a result, not only curriculum development is a major component of upper secondary education which is important to be adjusted to the needs of the labour market but also assessment and certification. The delivery of certificates which comprehensively communicate a student's skills are crucial for the student's opportunities of progression. At the basis of this stands an assessment, which captures a student's most important skills requested on the labour market (World Bank, 2005).

## 7. CONCLUSION

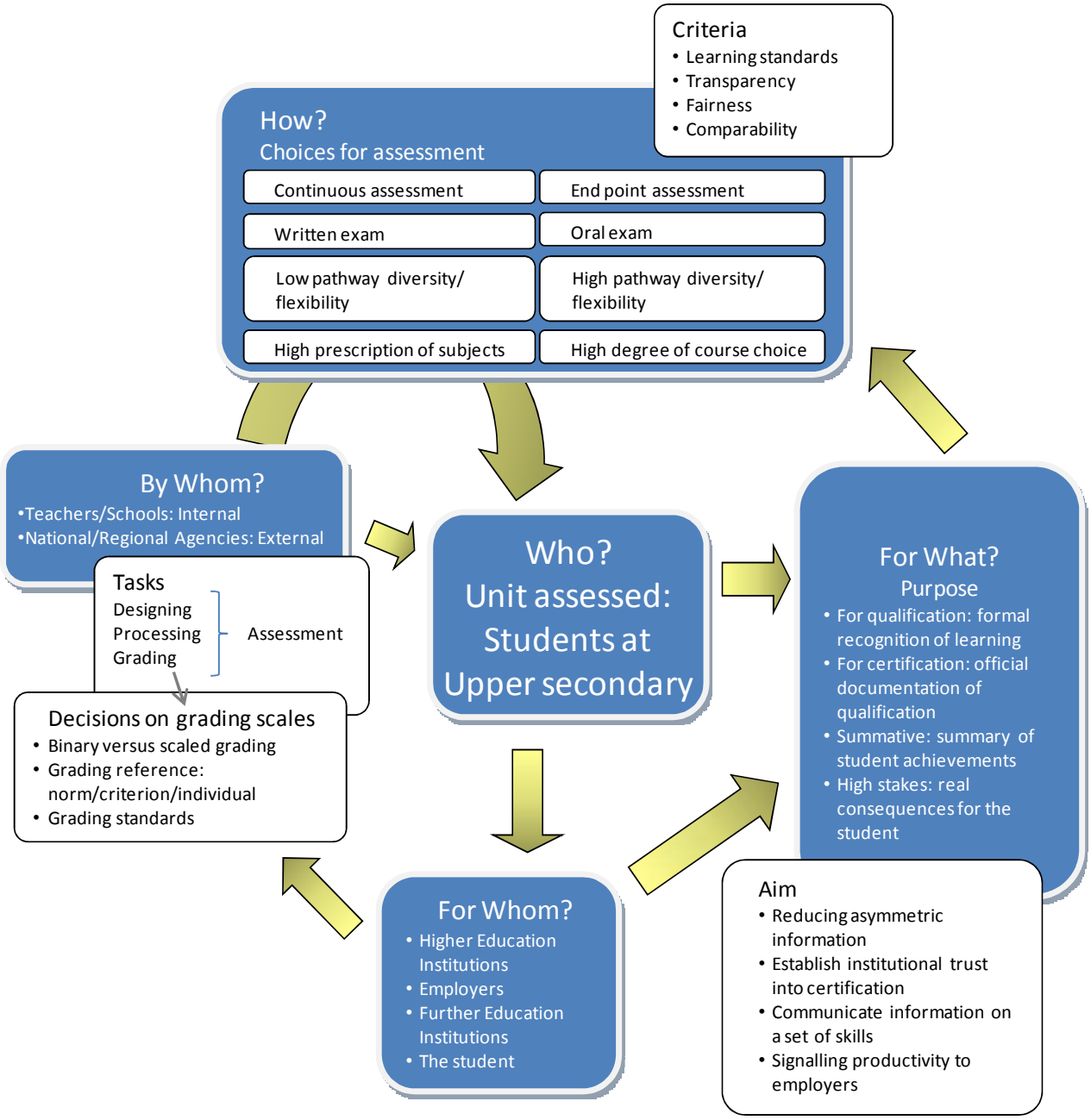
187. This paper discussed features as well as issues of assessment for qualification and certification in upper secondary education. Not only did it bring up significant evidence on the importance of structural options for the organisation of assessment, but also on its implications for students' learning.

188. As evidence from countries shows, responsibilities for assessment may be very differently distributed among systems which may strongly influence the role of teachers as well as the system's accountability. The design of qualification frameworks and pathways in upper secondary education also considerably determines students' flexibility and chances for progression and for continuous learning. But the creation of a qualification framework always relies on the establishment of standards. Several studies show how standards may influence students' performance to the better or to the worse, or how standards may contribute to decreasing attainment. Thus, standards are important for all sorts of testing. While MCEs only determine one standard level, CBEEEs rely on multileveled systems of performance which divides students into a higher number of achievement groups. The choice of written versus oral assessment can have further effects on accountability but implies also certain risks of bias in grading. Accountability is an issue of enhanced importance at the stage of upper secondary education as results are used by stakeholders such as higher education institutions and employers who may significantly influence the student's chances of progression. This leads to assessment for qualification and certification in upper secondary education being a key factor for students' career development.

189. While looking at issues of assessment for qualification and certification, this paper finds that external assessment is favoured by some countries for benefits of reliability and enhanced student performance while internal assessment provides teachers with opportunities for broader assessment of students' skills. Most OECD countries are found to rely on a mixed system with a combination of continuous teacher-based assessment and an externally processed end-point test. Even though among OECD countries there seems to be a preference for such a division of responsibilities for assessment, the focus of responsibilities and its structure vary largely among systems. There are indications in the research literature that grading issues are tightly connected to other aspects of a system's structure. Standards for grading may have positive or negative effects on students' performance. Furthermore, students' perception of fairness is a major factor for motivation. As grades may have a major effect on a student's future career, aspects of fairness are especially crucial in assessment for qualification and certification in upper secondary education. The consequences of this stage of education for students' careers, as shown by numerous studies, raises the importance of considerations of all features and issues of upper secondary assessment for qualification and certification presented in this paper to ensure students' attainment, excellent performance and best future career prospects.

# ANNEX I

## CONCEPTUAL FRAMEWORK FOR ASSESSMENT FOR QUALIFICATION AND CERTIFICATION IN UPPER SECONDARY EDUCATION



# ANNEX II
# EXAMPLES OF COUNTRY APPROACHES

## Australia

Australia has a federal school system with primary responsibility for school education granted to state and territory governments.

All states and territories have some form of senior secondary completion assessment covering both the government and non-government sectors. Years 11 and 12 subject-based exams and vocational education and training exams may be the most important assessments for students, as they are used for admission to tertiary education, work placement and employment. While these exams vary by state and territory, they provide the basis for the Australian Tertiary Admission Rank (ATAR), which combines students' relative performance with moderation procedures to place students on a common scale across all locales (except Queensland). While these exams vary by state and territory, there are moderation processes in place across Australia.

Students are given a wide choice of subjects, which essentially constitute the core of their secondary education. While each state and territory has its own system and own set of procedures for developing and approving courses (including specified learning goals, content, and exams), most combine student performance on external exams at the end of Year 12 with moderated, teacher judgments of coursework performance to arrive at scores for senior secondary certificates and high school completion. External exams are derived from a combination of multiple-choice, short answer and extended response tasks. Course work includes a variety of tasks, including extended performances. Examination systems in the Australian Capital Territory and Queensland are more school-determined and based, but achievement standards and scoring are externally moderated.

In Queensland, the moderation processes for the Senior Certificate (Year 12) involve subject-based panels of expert teachers providing advice to schools on the quality of their assessment programme and their judgments of quality of student performance based on sample portfolios. The system involves follow-up where panels identify difficulties. There is negotiation of the final results to be recorded on the Senior Certificate (Sebba and Maxwell, 2005).

*Source:* Santiago *et al.* (2011).

## Belgium (Flemish Community)

Secondary education in the Flemish Community of Belgium is organised in a uniform structure comprising three stages, four education programmes and several study areas. Students only make a final choice of study areas in the second stage so that they are first introduced to as many subjects as possible. The four programmes are: General secondary education (ASO), Technical secondary education (TSO), Secondary arts education (KSO) and Vocational secondary education (BSO).

The diploma of secondary education is generally awarded by schools (with exception of diplomas awarded by Examination Board of the Flemish Community). The diploma of secondary education means that students have reached the objectives set by the curriculum of the school. Attainment of these objectives implies that the attainment targets have also been sufficiently achieved.

The diploma of secondary education allows entrance to all forms of higher education at either Institutes for Higher Education or at universities. Possible restrictions by means of entrance exams (*e.g.* medicine or art studies) are generic and do not depend on the study area or educational programme (ASO-TSO-KSO-BSO) in which the diploma of secondary education was obtained.

However, high drop-out rates and low success rates in the first year of higher education have raised concerns about generic entrance. The current policy discussion paper for the reform of the structure of secondary education plans to limit entrance possibilities in higher education for students who graduated in secondary schooling preparing for the job market. It is planned that they will first have to attend a one year preparation course before being allowed in higher education.

*Source:* Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group (2010).

**Belgium (French Community)**

Secondary education in the French Community of Belgium comprises three stages and four education programmes (general, technical, artistic and professional).

The secondary education diploma is generally awarded by schools, and there is currently no common standardised examination required for certification. Certificates are delivered under the responsibility of the individual institution. There is, however, a procedure for external verification in case of appeal by a student or his/her parents. The class council (consisting of the entire leadership and teaching staff in charge of a specific group of students) plays an important role in all decisions concerning transitions and certification.

Currently, there are different types of certification depending on the level and orientation of students. There are different certificates for completing the first, second and third stage of secondary education. The qualification certificate (CQ) is delivered to students at the end of sixth and seventh year studies (CQ6 or CQ7) by a qualification jury. Also, students who leave the education system after having completed at least four years of studies in professional or technical education are awarded an attestation of intermediary competences specifying the competencies acquired by the student.

Common external certification examinations for all schools are currently being developed.

*Source:* Blondin and Giot (2011).

**Czech Republic**

Upper secondary education in the Czech Republic is completed by an examination specific to the different strands. These are the school-leaving examination (in general and technical schools), the final examination to acquire the apprenticeship certificate (in secondary vocational schools) and the final examination leading to a final school report.

Traditionally, the content, administration and marking of exit examinations, including the one associated with the apprenticeship certificate, have been the responsibility of individual schools. Teachers in individual schools specify the examination requirements, and assess students according to the assessment criteria stipulated in the respective school's School Education Programme. However, both for the school-leaving examination and the apprenticeship certificate, there have been developments to move towards introducing an element of externality.

Since 1997 there has been the intention to introduce a common, standardised component to the school-leaving examination, centrally administered and marked (with the exception of the parts which cannot be automatically marked, for which specifically trained teachers in the schools attended by the concerned student take responsibility). The main motivation for the introduction of the standardised component was to provide a more objective basis for the access to tertiary education. Following the development of several versions, the common component of the school-leaving examination was administered by the Centre for the Evaluation of Educational Achievement (CERMAT) in spring 2011. The school-leaving examination now has two components: the common (external) standardised part and the profile (school-based) part.

- The centrally prepared standardised common part is offered in two levels of difficulty with the choice up to the student. Both levels of difficulty are supposed to give access to further study conditional upon receiving the school-leaving examination certificate. In 2011, the common part included the following three assessments: Czech language, a foreign language, and an assessment in an optional subject (chosen from mathematics, sciences, information sciences and social sciences).

- The school-based component consists of two or three assessments set in Framework Education Programmes (FEPs) and assessments in two optional subjects. The school-based component is aligned with the character of the school (*i.e.* general or technical). The assessments are written or oral. Students may only take these examinations if they have successfully completed the final year of school. The passing of the school-leaving examination is a pre-requisite for admission to a higher education institution or a tertiary professional school.

Students "pass" the final examination if they successfully pass all the assessments in non-optional subjects in both the common and school-based examinations. There is no partial credit awarded for partial success, although students have the right to take a corrective examination in the subjects they failed.

The final examination to acquire the apprenticeship certificate is composed of practical and theoretical parts in given vocational fields, typically including a written component, an oral test and an examination in practical training. The final examination is vocational in nature and students must demonstrate how well prepared they are to perform the relevant skills for the specified occupations. Since 2005 the National Institute of Technical and Vocational Education has been working on a new final examination with the development of uniform content/tests (common assignments) to be used by schools in the examination of the various vocational programmes so a certain level of standardisation is reached. The objective is that the standardised content matches standards of individual qualifications in the National Qualifications Framework currently being developed.

*Source:* Santiago *et al.* (2012a)

**Luxembourg**

Assessment for qualification and certification in upper secondary school takes place in the final year (Year 13) of school. Standards are centrally set by the MENFP. Continuous assessment of the final year counts for one-third of the final grade in each subject. The final examination thus weighs two-thirds. This final examination is further divided into a written and oral part. The oral exam counts for 25%. Students who successfully pass continuous and final assessment are awarded the *diplôme de fin d'études secondaires* (Ministère de l'Education Luxembourg, 2006).

As an alternative to general secondary education, students may follow technical secondary education in specialised schools (Ministère de l'Education Luxembourg, 2006). The certification of the "technical"

stream entitles for the admission to certain higher education institutions and for entry into the job market, while the "professional" stream prepares for different qualification levels in the particular profession.

*Source:* MENFP (2006).

**New Zealand**

Assessment for qualification and certification in upper secondary in New Zealand is regulated in the New Zealand Qualifications Framework (NZQF). The responsibility for the award of certificates within the NZQF lies with the New Zealand Qualifications Authority (NZQA).

The main qualification in secondary education is the National Certificate of Educational Achievement (NCEA), in which students are assessed against a range of National Standards. Years 11, 12, and 13 of upper secondary education typically correspond to NCEA Levels 1, 2 and 3, but it is possible for students to take NCEA assessments earlier in their secondary schooling. The accomplishment of level 3 is a requirement for entry to tertiary education (New Zealand Ministry of Education, 2010).

Standards contributing to NCEA are listed on the Directory of Assessment Standards. The standards specify learning outcomes and assessment criteria. There are two types of standards: unit standards and achievement standards. Unit standards are vocationally-based and mostly used in workplace training and the tertiary sector. Achievement standards are academically-based and focused on the secondary school curriculum. The Directory of Assessment Standards contains over 26 000 unit standards and about 850 achievement standards. Schools can design and offer their own courses mixing unit standards and achievement standards.

Assessment for secondary qualification can involve both internal and external assessment approaches. All unit standards are internally assessed. In 2010, assessment of students was carried out using approximately one third unit standards, one third internally assessed achievement standards and one third externally assessed achievement standards. External assessment is conducted by NZQA via national examinations (or by portfolio for certain subjects). Internal assessment is largely based on coursework and classroom-based assessment. An external moderation system is in place to ensure the dependability of internal assessments in Years 11-13.

Students are graded for their achievement on each standard. Four grade categories are available for achievement standards: Not Achieved, Achieved, Achieved with Merit and Achieved with Excellence. Most unit standards have only two grade categories (Not Achieved and Achieved) but there is current work on developing Merit and Excellence grades for some unit standards.

*Source:* Nusche *et al.* (2012).

**Norway**

In Norway, assessment for graduation from upper secondary school is based on a mix of continuous teacher-based assessment and final examinations. Final marks are a key criterion for admission to higher education. Continuous assessment is supposed to provide a broad reflection of a student's achievements through the application of various assessment tools. A student's order and conduct are assessed separately but contribute to certification. Most subjects are only assessed continuously (Norwegian Directorate for Education and Training, 2011).

All students sit a limited number of examinations. In the first year of upper secondary, 20% of all students are sampled to participate in a central examination or a local oral examination in one subject. In the second year of upper secondary education, all students who are qualifying for higher education

participate in one central written examination or a local oral examination in one subject. Students in the vocational programme participate in an interdisciplinary practical exam in one subject and 20% are selected for a central examination or a local oral examination in a common core subject. In the third year of upper secondary education, all students qualifying for higher education sit for a central examination in Norwegian (or Sami, if first language) and all students are selected for a central written examination in two subjects and for a local oral examination in one subject.

The external examiners scoring the written tests are practicing teachers who are recommended to the County Governor by their principals. They participate in a yearly seminar that intends to professionalise their grading and contribute to a common understanding of assessment criteria. The oral examinations are implemented by the subject teacher together with a teacher from another school.

On the school-leaving certificate, grades for final and continuous examination are displayed separately. For the calculation of an overall achievement grade for certification both grades have the same weight. At the end of an upper secondary programme qualifying for higher education, a student's school leaving certificate will contain just over 20 overall achievement marks and five or six examination marks.

*Source:* Nusche et *al.* (2011b)

**Portugal**

For access to higher education, students in Portugal have to take part in national examinations, which take place in the last year the subject is being taught (Eurydice, 2009).

In upper secondary education, students in Portugal may choose between different programmes. The programmes on offer in secondary education are diversified into scientific and humanities courses to prepare students for progression into higher education, and technological courses which have a more vocational focus. There are also artistic programmes and professional/vocational courses available. This diversification is significant for assessment, in that the programme of national examinations is associated only with the scientific and humanities courses.

External assessment has been applied to the scientific and humanistic courses since 2006. Students are marked three times in each year, on a scale of 1-20 with the third grade in each year determining whether the student will be admitted to the national examination associated with each year and which is taken in the final year of each subject. A mark of ten is required to take the national examination. In terms of the final classification of students, the internal summative assessments for which teachers have responsibility carries 70% of the weighting, with the national examination making up the remaining 30%. However, for the purposes of entry to higher education, the balance is different, with the school-based component reduced to 50% and the external examination increased to 50%. The national examinations are the responsibility of the Ministry of Education.

The external national assessments are not applied to technological, professional/vocational, artistic or other specialised programmes. Successful completion of these latter courses is not subject to any national assessment, but to local, internal assessment only. Students in these courses who wish to progress to higher education have to take the national examinations required as entrance tests for the courses they wish to take.

*Source:* Santiago *et al.* (2012b).

**Sweden**

The successful completion of upper secondary education is a requirement for access to higher education.

Students in upper secondary education in Sweden are required to sit national assessments in the core subjects (Swedish/Swedish as a Second Language, mathematics and English) and in one of the science subjects (biology, physics or chemistry), as allocated by the National Agency for Education. The assessments are administered and graded by the students' own teachers. Tests are also available "on demand" in different subjects, including foreign languages, social science subjects and selected vocational subjects from a test bank run by the National Agency for Education (Nusche *et al.*, 2011).

The Swedish education system relies on criterion-referenced grading. The grade levels in upper secondary education currently are IG (Icke Godkänt - Fail), G (Godkänt - Pass), VG (Väl Godkänt - Pass with distinction), MVG (Mycket Väl Godkänt - Pass with special distinction).

The school-leaving reports at the end of upper secondary report provide a summary statement of student achievements in school. They are based on teachers' continuous assessment in the classroom, which is supported by the compulsory national tests in certain subjects.

*Source:* Nusche *et al.* (2011a).

## REFERENCES

Amrein, A.L. and D.C. Berliner (2002), "High-Stakes Testing, Uncertainty, and Student Learning", *Education Policy Analysis Archives*, Vol. 10, No. 18, *http://epaa.asu.edu/epaa/v10n18/.*

Angrist, J. and V. Lavy (2009), "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial", *American Economic Review*, Vol. 99, No. 4, pp. 1384–1414.

Au, W. (2007), "High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis", *Educational Researcher*, Vol. 36, No. 5, pp. 258-267.

Banks, J., D. Byrne, S. McCoy and E. Smyth (2010), *Engaging Young People? Experiences of the Leaving Certificate Applied Programme*, Research Series No. 15, April, The Economic and Social Research Institute, Dublin.

Betts, J.R. (1995), *Do Grading Standards Affect the Incentive to Learn?*, Working Paper, University of California, San Diego.

Betts, J.R. and R.M. Costrell (2000), "Incentives and Equity Under Standards-Based Reform", *Economics Department Working Paper Series*, Paper 88, University of Massachusetts Amherst.

Bishop, J.H. (1997a), "The Effect of National Standard and Curriculum-Based Exams on Achievement", *CAHRS Working Paper Series,* No. 143.

Bishop, J.H. (1997b), "The Effect of Curriculum-Based External Exit Exam Systems on Student Achievement", *CAHRS Working Paper Series*, Paper 156.

Bishop, J.H. (1999), "Nerd Harassment, Incentives, School Priorities and Learning", in S.E. Mayer and P.E. Peterson (eds.), *Earning and Learning: How Schools Matter*, Brookings Institution Press, Washington.

Bishop, J.H. (2006), "Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn – Externalities, Information Problems and Peer Pressure", *Handbook of the Economics of Education*, Vol. 2, pp. 910-944.

Bishop, J.H. and F. Mane (2001), "The Impacts of Minimum Competency Exam Graduation Requirements on High School Graduation, College Attendance and Early Labor Market Success," *Labour Economics*, Vol. 8, pp. 203-222.

Bishop, J.H., J.Y. Moriarty and F. Mane (1997), "Diplomas for Learning, not Seat Time: The Impact of New York Regents Examinations", *CAHRS Working Paper Series*, Paper 171.

Blondin, C. and B. Giot (2011), *Cadres d'évaluation en vue d'améliorer les résultats scolaires : Rapport de base pour la Communauté française de Belgique*, *www.oecd.org/dataoecd/63/21/49048280.pdf.*

Bolstad, R. (2006), "Questions for a Twenty-First Century Senior Secondary Curriculum", *Curriculum Matters*, Vol. 2, pp. 105-127.

Bonesrønning, H. (1999), "The Variation on Teachers' Grading Practices: Causes and Consequences", *Economics of Education Review*, Vol. 18, pp. 89-105.

Bouder, A., F. Dauty, J.-L. Kirsch and P. Lemistre (2008), "Readability of Qualifications: A Question as Old as Europe?", in P. Descy and M. Tessaring (eds.), *Modernising Vocational Education and Training – Fourth Report on Vocational Training Research in Europe: Background Report*, Vol. 3, Office for Official Publications of the European Communities, Luxembourg.

Bundesministerium für Unterricht, Kunst und Kultur (2008), "Allgemeinbildende höhere Schulen", *www.bmukk.gv.at/schulen/bw/abs/ahs.xml*.

Cedefop (2010), *Changing Qualifications: A Review of Qualifications Policies and Practices*, Publication Office of the European Union, Luxembourg.

Cumming, J.J. and G.S. Maxwell (2004), "Assessment in Australian Schools: Current Practice and Trends", *Assessment in Education*, Vol. 11, No. 1, pp. 89-108.

Dalbert, J., U. Schneidewind and A. Saalbach (2007), "Justice Judgements Concerning Grading in School", *Contemporary Educational Psychology,* Vol. 32, pp. 420-433.

Danish Ministry of Education (2010), *Upper Secondary Education: General Facts*, *www.eng.uvm.dk/Fact%20Sheets/Upper%20secondary%20education/General%20facts.aspx*.

Danish Ministry of Education (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Denmark*, OECD, Paris, *www.oecd.org/edu/evaluationpolicy*.

Dee, T.S. and B.A. Jacob (2006), "Do High School Exit Exams Influence Educational Attainment or Labor Market Performance?", *NBER Working Paper Series,* No. 12199, National Bureau of Economic Research, Massachusetts, *www.nber.org/papers/w12199.pdf*.

Dryler, H. (1999), "The Impact of School and Classroom Characteristics on Educational Choices by Boys and Girls: A Multilevel Analysis", *Acta Sociologica*, Vol. 42, pp. 299-318.

Eurydice (2009), *Key Data on Education in Europe 2009*, Eurydice, Brussels.

Fan, R.M. and S.C.N. Chan (1999), "Students' Perceptions of Just and Unjust Experiences in School", in S. Gibbs (ed.), *Education and Child Psychology*, Vol. 16, No. 4, pp. 32-50.

Field, S., M. Kuczera and B. Pont (2007), *No More Failures: Ten Steps to Equity in Education*, OECD, Paris.

Figlio, D.N. and M.E. Lucas (2003), "Do High Grading Standards Affect Student Performance?", *NBER Working Paper Series*, No. 7985, *www.nber.org/papers/w7985*.

Flemish Ministry of Education and Training and the University of Antwerp Edubron Research Group (2010), *Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for the Flemish Community of Belgium*, *www.oecd.org/dataoecd/61/49/46974684.pdf*.

Fullarton, S. (2001), "Does VET in Schools Make a Difference to Post-School Pathways?", *Understanding Youth Pathways: Research Conference 2001,* Australian Council for Educational Research (ACER), Melbourne, pp. 40-42.

Fullarton, S., M. Walker, J. Ainley and K. Hillman (2003), "Patterns of Participation in Year 12", *Australian Council for Educational Research Longitudinal Surveys of Australian Youth*, Research Report 33.

Graham, A. and T. Husted (1993), "Understanding State Variation in SAT Scores", *Economics of Education Review*, Vol. 12, No. 3, pp. 197-202.

Grubb, N., S. Field, H. Marit Jahr and J. Neumüller (2005), *Equity in Education Thematic Review: Finland Country Note*, OECD, Paris, *www.oecd.org/dataoecd/49/40/36376641.pdf.*

Guskey, T. R. (2000), "Grading Policies that Work Against Standards…and How To Fix Them", *NASSP Bulletin*, Vol. 84, No. 620, pp. 20-29.

Hambleton, R.K., R.M. Jaeger, B.S. Plake and C. Mills (2000), "Setting Performance Standards on Complex Educational Assessments", *Applied Psychological Measurement*, Vol. 24, No. 4, pp. 355-366.

Harlen, W. (2007), "Criteria for Evaluating Systems of Student Assessment", *Studies in Educational Evaluation*, Vol. 33, pp. 15-28.

Hart, J. and C. Howieson (2004), "Unitisation: Benefits and Issues", *Research and Information Services Bulletin*, No. 9, SQA, Glasgow and Dalkeith.

Hodgson, A., C. Howieson, D. Raffe, K. Spours and T. Tinklin (2004), "Post-16 Curriculum and Qualifications Reform in England and Scotland: Lessons from Home International Comparisons", *Journal of Education and Work*, Vol. 17, No. 4, pp. 441-465.

Holme, J.J., M.P. Richards, J.B. Jimerson and R.W. Cohen (2010), "Assessing the Effects of High School Exit Examinations", *Review of Educational Research,* Vol. 80, No. 476, American Educational Research Association, Washington, pp. 476-526.

INCA (International Review of Curriculum and Assessment Frameworks Internet Archive) Australia (2008), "Australia: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive*, *www.inca.org.uk/2209.html*.

INCA Canada (2007), "Canada: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive, www.inca.org.uk/1218.html*.

INCA England (2010), "England: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive*, *www.inca.org.uk/1317.html*.

INCA France (2008), "France: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive*, *www.inca.org.uk/1383.html*.

INCA Germany (2005), "Germany: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive*, *www.inca.org.uk/1434.html*.

INCA Hungary (2006), "Hungary: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive*, *www.inca.org.uk/1501.html*.

INCA Ireland (2011), "Ireland: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive*, *www.inca.org.uk/1854.html*.

INCA Italy (2008), "Italy: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive*, *www.inca.org.uk/1612.html*.

INCA Japan (2007), "Japan: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive*, *www.inca.org.uk/1483.html*.

INCA Korea (2008), "Korea: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive*, *www.inca.org.uk/1403.html*.

INCA Netherlands (2007), "Netherlands: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive, www.inca.org.uk/1342.html*.

INCA New Zealand (2008), "New Zealand: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive*, *www.inca.org.uk/1285.html*.

INCA Spain (2008), "Spain: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive*, *www.inca.org.uk/874.html*.

INCA Sweden (2009), "Sweden: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive*, *www.inca.org.uk/692.html*.

INCA Switzerland (2009), "Switzerland: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive*, *www.inca.org.uk/558.html*.

INCA USA (2007), "USA: Assessment Arrangements", *International Review of Curriculum and Assessment Frameworks Internet Archive, www.inca.org.uk/397.html*.

Jacob, B.A. (2005), "Accountability, Incentives and Behavior: the Impact of High-Stakes Testing in the Chicago Public Schools", *Journal of Public Economics,* Vol. 89, pp. 761-796.

Jurges, H. and K. Schneider (2010), "Central Exit Examinations Increase Performance…But Take the Fun out of Mathematics", *Journal of Population Economics,* Vol. 23, pp. 497-517.

Köditz, V. (2009), *How Should Europe Handle the Globalisation? An Educational Perspective: Part III Upper Secondary Education and Vocational Training Research Report*, Foundation for European Progression Studies, Cologne.

Le Métais, J. (2002), "International Developments in Upper Secondary Education: Context, Provision and Issues", *INCA Thematic Study No.8*, National Foundation for Educational Research (NFER), Slough.

Lee, H. and G. Lee (2001), "The National Certificate of Educational Achievement (NCEA): Fragile – Handle With Care", *New Zealand Annual Review of Education*, Vol. 10, pp. 5-38.

Lillard, D.R. and P.P. DeCicca (2001), Higher Standards, More Dropouts? Evidence Within and Across Time", *Economics of Education Review*, Vol. 20, pp. 459-473.

Looney, J. (2009), "Assessment and Innovation in Education", *OECD Education Working Papers,* No. 24, OECD Publishing, Paris.

Looney, J. (2011). "Integrating Formative and Summative Assessment: Progress Towards a Seamless System?", *OECD Education Working Papers,* No. 58, OECD Publishing, Paris.

Lund, S. (2008), "Choice Paths in the Swedish Upper Secondary Education: a Critical Discourse Analysis of Recent Reforms", *Journal of Education Policy,* Vol. 23, No. 6, Routledge, London, pp. 633-648.

Martorell, F. and D. Clark (2010), "The Signaling Value of a High School Diploma" *Rand Education Working Paper Series*, 780.

MENFP (Ministère de l'Éducation nationale et de la Formation professionelle) (2006). *Reglement grand-ducal du 14 juillet 2005, determinant l'évaluation et la promotion des élèves de l'enseignement secondaire technique et de l'enseignement secondaire.*

Mercurio, A. (2006), "Teacher-Based Assessment at the Upper Secondary Level of Education: An Exploration of the Interaction between 'Teachers as Assessors' and 'Teachers as Moderators'", paper presented at the 32nd International Association for Educational Assessment (IAEA) Conference, Singapore, *www.iaea2006.seab.gov.sg/conference/programme.html*.

Mercurio, A. (2008), "Re-imagining School-Based Assessment at the Upper Secondary Education Level", paper presented at the 34th IAEA Annual Conference, Cambridge, *www.iaea2008.cambridgeassessment.org.uk/ca/digitalAssets/180434_Mercurio.pdf.*

Miguel, A. and E. Larson (2006), *Efficient and Effective Grading of Student Work*, American Society for Engineering Education, *http://fac-staff.seattleu.edu/elarson/web/Research/asee06.pdf.*

Montmarquette, C. and S. Mahseredjian (1989), "Could Teacher Grading Practices Account for Unexplained Variation in School Achievements?", *Economics of Education Review*, Vol. 8, No. 4, pp. 335-343.

Mueller, M. and M.H.J. Wolbers (2003), "Educational Attainment in the European Union: Recent Trends in Qualification Patterns", in W. Mueller and M. Gangl (eds.), *Transitions from Education to Work in Europe: The Integration of Youth into EU Labour Markets*, Oxford University Press, Oxford, pp. 23-62.

New Zealand Ministry of Education (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report New Zealand*, OECD, Paris, *www.oecd.org/edu/evaluationpolicy*.

Norwegian Directorate for Education and Training (2011), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Norway*, OECD, Paris, *www.oecd.org/edu/evaluationpolicy*.

Nusche, D., G. Halász, J. Looney, P. Santiago and C. Shewbridge (2011a), *OECD Reviews of Evaluation and Assessment in Education: Sweden*, OECD, Paris.

Nusche, D., L. Earl, W. Maxwell and C. Shewbridge (2011b), *OECD Reviews of Evaluation and Assessment in Education: Norway*, OECD, Paris.

Nusche, D. D. Laveault, J. MacBeath and P. Santiago (2012), *OECD Reviews of Evaluation and Assessment in Education: New Zealand*, OECD, Paris.

Nusche, D. (forthcoming), "Summative Assessment: What's in it for Students?", *OECD Education Working Papers*, OECD Publishing, Paris.

OECD (2000), *From Initial Education to Working Life: Making Transitions Work,* OECD, Paris.

OECD (2005), *Economic Survey Belgium - Vol. 2005/5*, OECD, Paris.

OECD (2007), *Qualifications Systems: Bridges to Livelong Learning*, OECD, Paris.

OECD (2010a), *Learning for Jobs*, OECD, Paris.

OECD (2010b), *Education at a Glance: OECD Indicators 2010*, OECD, Paris.

Ou, D. (2010), "To Leave or Not to Leave? A Regression Discontinuity Analysis of the Impact of Failing the High School Exit Exam", *Economics of Education Review*, Vol. 29, pp. 171-186.

Papay, J.P., R.J. Murnane and J.B. Willett (2008), "The Consequences of High School Exit Examinations for Struggling Low-Income Urban Students: Evidence from Massachusetts", *Working Paper* 14186, National Bureau of Economic Research, Massachusetts, *www.nber.org/papers/w14186.*

Perna, L.W. and S.L. Thomas (2009), "Barriers to College Opportunity: The Unintended Consequences of State-Mandated Testing", *Educational Policy*, Vol. 23, No. 3, pp. 451-479.

QSA (Queensland Studies Authority) (2010), *Annual Report 2009-10: Queensland Studies Authority*, South Brisbane, The State of Queensland, *www.qsa.qld.edu.au.*

Race, P. (2007), *The Lecturer's Toolkit: a Practical Guide to Assessment, Learning and Teaching*, Routledge, New York.

Raffe, D. (2001), "What Are Pathways? Concepts and Evidence from Cross-National Research", *Understanding Youth Pathways: Research Conference 2001,* Australian Council for Educational Research (ACER), Melbourne, pp. 3-5.

Raffe, D. (2003), "Simplicity Itself: The Creation of the Scottish Credit and Qualifications Framework", *Journal of Education and Work*, Vol. 16, No. 3, pp. 239-257.

Randall, J. and G. Engelhard (2010), "Examining the Grading Practices of Teachers", *Teaching and Teacher Education*, Vol. 26, pp. 1372-1380.

Rawlins, P., J. Brandon, J. Chapman, L. Leach, G. Neutze, A. Scott and N. Zepke (2005), "Standards-Based Assessment in the Senior Secondary School: A Review of the Literature", *New Zealand Journal of Teachers' Work*, Vol. 2, Issue 2, pp. 107-115.

Reardon, S.F., A. Atteberry, N. Arshan and M. Kurlaender (2009), "Effects of the California High School Exit Exam on Student Persistence, Achievement, and Graduation", *Stanford University Working Paper*, No. 12, Institute for Research on Education Policy & Practice.

Remesal, A. (2011), "Primary and Secondary Teachers' Conceptions of Assessment: A Qualitative Study", *Teaching and Teacher Education*, Vol. 27, pp. 472-482.

Rosenkvist, M. A. (2010), "Using Student Test Results for Accountability and Improvement: A Literature Review", *OECD Education Working Papers*, No. 54, OECD Publishing, Paris.

Sabbagh, C., N. Resh, M. Mor and P. Vanhuysse (2006), "Spheres of Justice within Schools: Reflections and Evidence on the Distribution of Educational Goods", *Social Psychology of Education,* Vol. 9, pp. 97-118.

Santiago, P., G. Donaldson, J. Herman and C. Shewbridge (2011), *OECD Reviews of Evaluation and Assessment in Education: Australia*, OECD, Paris.

Santiago, P., A. Gilmore, D. Nusche and P. Sammons (2012a), *OECD Reviews of Evaluation and Assessment in Education: Czech Republic*, OECD, Paris.

Santiago, P., G. Donaldson, A. Looney and D. Nusche (2012b), *OECD Reviews of Evaluation and Assessment in Education: Portugal*, OECD, Paris.

Shewbridge, C., E. Jang, P. Matthews and P. Santiago (2011), *OECD Reviews of Evaluation and Assessment in Education: Denmark*, OECD, Paris.

Smyth, E. and C. Hannan (2007), "School Processes and the Transition to Higher Education", *Oxford Review of Education*, Vol. 33, No. 2, pp. 175-194.

Steele, C.M. (1997), "A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance", *American Psychologist*, Vol. 52, No. 6, pp.613-629.

Strachan, J. (2002), "Assessment in Change: Some Reflections on the Local and International Background to the National Certificate of Educational Achievement (NCEA)", *New Zealand Annual Review of Education*, Vol. 11, pp. 245-273.

Swanson, C.B. and D. Lee Stevenson (2002), "Standards-Based Reform in Practice: Evidence on State Policy and Classroom Instruction from the NAEP State Assessments", *Educational Evaluation and Policy Analysis*, Vol. 24, No. 1, pp. 1-27.

Swedish Ministry of Education and Research (2010), *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Sweden*, OECD, Paris, *www.oecd.org/edu/evaluationpolicy*.

Tait, T. (2003), *Credit Systems for Learning and Skills: Current Developments –LSDA Reports*, Learning and Skills Development Agency, London.

Tuck, R. (2007), *An Introductory Guide to National Qualifications Frameworks: Conceptual and Practical Issues for Policy Makers*, International Labour Office, Geneva.

Tyler, J.H., R.J. Murnane and J.B. Willett (2000), "Do the Cognitive Skills of School Dropouts Matter in the Labor Market?", *The Journal of Human Resources*, Vol. 19, No. 4, pp. 749-754.

Victorian Curriculum and Assessment Authority (2011), *Victorian Certificate of Education*, *www.vcaa.vic.edu.au/vce/index.html.*

Vlor (Vlaamse Onderwijsraad) (2010), "Advies over Tweedekansleerwegen voor het Behalen van een Diploma Secundair Onderwijs", Flemish Education Council, Brussels, *www.vlor.be/sites/www.vlor.be/files/rlll-adv-005-1011_0.pdf.*

Volante, L. (2004), "Teaching to the Test: What Every Educator and Policy-Maker Should Know", *Canadian Journal of Educational Administration and Policy*, Vol. 35, *http://umanitoba.ca/publications/cjeap/articles/volante.html*.

Warren, J.R. and M.R. Edwards (2005), "High School Exit Examinations and High School Completion: Evidence from the Early 1990s", *Educational Evaluation and Policy Analysis*, Vol. 27, No. 1, pp. 53-74.

Wikstrom, C. and M. Wikstrom (2005), "Grade Inflation and School Competition: an Empirical Analysis Based on the Swedish Upper Secondary School", *Economics of Education Review*, Vol. 24, pp. 309-322.

William, D. (2010), "What Counts as Evidence of Educational Achievement? The Role of Constructs in the Pursuit of Equity in Assessment", *Review of Research in Education*, Vol. 34, pp. 254-284.

Winters, M.A., J.P. Greene and J. Trivitt (2010), "The Impact of High-Stakes Testing on Student Proficiency in Low-Stakes Subjects: Evidence from Florida's Elementary Science Exam", *Economics of Education Review*, Vol. 29, No. 1, pp. 138-146.

Woessmann, L. (2003), "Schooling Resources, Educational Institutions and Student Performance: The International Evidence", *Oxford Bulletin of Economics and Statistics*, Vol. 65, No. 2, pp. 117-170.

Woessmann, L., E. Luedemann, G. Schuetz and M.R. West (2007), "School Accountability, Autonomy, Choice, and the Level of Student Achievement: International Evidence from PISA 2003", *OECD Education Working Papers*, No. 13, OECD, Paris.

Wood, R.E., A. Mento and E. Locke (1987), "Task Complexity as a Moderator of Goal Effects: A Meta Analysis", *Journal of Applied Psychology*, Vol. 72, No. 3, pp. 416–425.

World Bank (2005), *Élargir les Possibilités et Développer les Compétences des Jeunes: un Nouveau Programme pour l'Enseignement Secondaire*, World Bank, Washington DC.

Young, M.D. (2003), "National Qualifications Frameworks as a Global Phenomenon: A Comparative Perspective", *Journal of Education and Work*, Vol. 16, No. 3, pp. 223-237.

## THE OECD EDUCATION WORKING PAPERS SERIES ON LINE

The OECD Education Working Papers Series may be found at:

- The OECD Directorate for Education website: www.oecd.org/edu/workingpapers
- Online OECD-ilibrary: http://www.oecd-ilibrary.org/education/oecd-education-working-papers_19939019
- The Research Papers in Economics (RePEc) website: www.repec.org

If you wish to be informed about the release of new OECD Education working papers, please:

- Go to www.oecd.org
- Click on "My OECD"
- Sign up and create an account with "My OECD"
- Select "Education" as one of your favourite themes
- Choose "OECD Education Working Papers" as one of the newsletters you would like to receive

For further information on the OECD Education Working Papers Series, please write to: edu.contact@oecd.org.