A Critical Review of National Policies on Formative Assessment

Introduction

Educational policy is considered to have an indirect effect on student achievement by changing school stakeholders' actions toward improving teaching and assessment practice (Kyriakides, Creemers, Antoniou, Demetriou, & Charalambous, 2015). One of the aims of our Erasmus+KA3 project entitled "Promoting Formative Assessment: From Theory to Policy and Practice (FORMAS)", is to generate policy guidelines that promote formative assessment, since research suggests that formative assessment practices have a positive impact on student achievement (Creemers & Kyriakides, 2015; Hattie & Temperley, 2007; Herman, Osmundson, Ayala, Schneider, & Timms, 2006; Wiliam, Lee, Harrison, & Black, 2004). Specifically, the project attempts to encourage policy makers to reform assessment policies and to establish teacher support mechanisms for the effective implementation of formative assessment. In order to do so, a critical analysis of current national policies of the four participating countries (i.e. Cyprus, Belgium, Greece and the Netherlands) related to student assessment was conducted. In this report, the framework and the methodology for conducting the data collection and analysis are presented. Then, the main results of the analysis are presented. Both within- and across-country analyses are conducted to identify similarities and differences among the four countries. The report ends with a critical discussion of the results in relation to the main findings of research on formative assessment in order to identify strengths and limitations of current assessment policies.

1. The Framework used for the Analysis of Policy

The term policy refers to a course or principle of action adopted or proposed by an organization or individual (Cohen & Hill, 2001; Kyriakides et al., 2015; Kyriakides, Georgiou, Creemers, Panayiotou,

& Reynolds, 2018). In education, policy-makers are expected to introduce and encourage specific actions that school stakeholders (e.g., management teams, teachers, students, and parents) should undertake to promote student learning. These actions are proposed in educational laws and/or official documents published by the national/local governing bodies and define regulations, as well as, the roles of different stakeholders towards the effective functioning of schools. At the same time, relevant mechanisms are developed to support the effective implementation of the proposed policy guidelines and regulations.

In this project, the educational policy in relation to student assessment in secondary education Mathematics in the four participating countries is investigated. Recognizing the role of teacher training and Teacher Professional Development (TPD) and how these may affect teachers' practice (Desimone et al., 2002; Guskey & Sparks, 2004), policies related to initial and in-service training on assessment are also examined. In addition, it is acknowledged that evaluation mechanisms examining policy implementation are also necessary since they guide decisions for the continuation, termination or adjustment of implemented policies based on their effectiveness (Scheerens, 2016). Therefore, in this project national educational policies related to the following three aspects are taken into account: *a) student assessment, b) teacher training and TPD in student assessment,* and *c) evaluation mechanisms in regard to the policy on assessment and TPD.*

The first aspect addresses policies that refer to the assessment of students. Specifically, we examine policies that refer to the purpose of assessment (i.e. formative and summative), assessment techniques, reporting procedures to relevant stakeholders, the provision of supportive resources, and the existence (if any) of national assessment mechanisms. The second aspect addresses teacher training and professional development in student assessment. Thus, we examine policies that refer to training in assessment as part of the programs of study leading to a graduate of post-graduate teaching degree, in-service professional development and the support offered to schools by advisory bodies for the implementation of student assessment. In addition, teacher collaboration within and between schools in relation to assessment is examined. Finally, the third aspect refers to policies relevant to mechanisms developed to evaluate the national policies (including policy reforms) on assessment and TPD. Specifically, first we examine if an evaluation mechanism is present. If such a mechanism exists then

2

we look at how frequently and at which stage of the policy implementation the evaluation is performed, the sources of data and processes used, the focus of the evaluation, as well as, how the results are used and for whom.

In addition, for each of the three aspects presented above, both quantitative and qualitative characteristics of the policies are taken into account. To achieve this, we draw from the dynamic model of educational effectiveness (Creemers & Kyriakides, 2008) which recognizes student assessment as a factor that has an impact on student learning outcomes. The model suggests that student assessment can be defined and measured in relation to five dimensions: *frequency, stage, focus, quality, and differentiation*. First, the number of relevant policies is identified in order to examine the *frequency* that each aspect is addressed in each educational system. Second, the *focus* of each policy is examined by looking into its specificity, as well as, the purpose or purposes it aims to serve. Third, the *stage* at which the policy takes place is investigated. This way, it is possible to identify when a policy is established and whether and when at some point it has been reviewed, adjusted or terminated. Fourth, it is examined whether policies are in line with the literature. This way, conclusions concerning the *quality* of a policy can be drawn. Finally, the extent to which a policy is *differentiated* regarding the group of teachers (i.e. newly appointed) or students (i.e. special education students) it involves is taken into account. The use of this measurement framework allows us to have a more comprehensive view of the existing policies. At the same time, it helps us to provide more specific suggestions and policy guidelines that can promote formative assessment.

2. Methodology

Two different sources of data on national policy on student assessment in the subject of Mathematics in secondary education were collected through: **(a) content analysis of policy documents** and **(b) interviews with policymakers who have a key role in promoting assessment policies**. In this way, we were able to measure not only the official policy as it is stated in documents and guidelines sent to schools, but also the actions taken by policymakers in their attempt to promote the implementation of the national policy at the school level.

2.1. Content analysis of policy documents

A content analysis was conducted of all official educational documents of each country (i.e. Cyprus, Greece, Belgium and the Netherlands) that included regulations and guidelines on aspects related to: a) student assessment, b) teacher training and TPD in assessment, and c) evaluation mechanisms in regard to the policy on assessment and TPD. The specific dimensions examined for each aspect have been presented above and are also shown in Table 1.

For each aspect, a profile was established by studying the educational laws and all regulations and guidelines communicated to schools (see Country's Policy Profile at https://ucy.ac.cy/formas/documents/Deliverable_2.1_Analysis_of_Policies/Deliverable_2.1_Policy_Profile.pdf). The suggestions/regulations were documented and evaluated in order to generate scores per each dimension. Specifically, the research team had to examine each policy by taking into account: (a) the extent to which schools were encouraged to undertake relevant actions, or were required to do so, or even were held accountable; (b) the stakeholders addressed (i.e., secondary and/or primary education); (c) the subjects addressed (i.e. Mathematics only and/or other subjects) (d) the extent to which it was in line with the literature; (e) the extent to which there was any kind of differentiation in the suggestion and the relevant actions that emerged from it; and (f) the time period that was initially established and the changes thereafter. The number of policy guidelines per aspect was also taken into account in order to find out the emphasis that each system gives to each aspect. In regard to the evaluation policy, the views of each country team about the national evaluation mechanisms that exist in each country were examined. Both, partners from the University Institutions and the Ministries were expected to provide their views and reach consensus. Based on the common views that derived, the seven questions included in the profile were answered. The scores for each dimension of each aspect were entered in an SPSS data file. In the SPSS sheet as many columns as types of actions were created, to measure the frequency dimension (see Excel File used for Entering Data from Country's Policy Profile at https://ucy.ac.cy/formas/en/resources with 1-2 examples per country). The completed policy documents profiles per country are available in their original form on our project's website (see https://ucy.ac.cy/formas/en/resources). Next, the methodology for conducting the interviews with policymakers is presented.

Table 1. The specific dimensions examined for each aspect of the national educational policy of each participating country

Aspects of the National Educational Policy	Dimensions examined for each aspect
Assessment	
1. Purposes of assessment	✓ Frequency (i.e. Number of policy measures)
2. Techniques of assessment	✓ Focus 1 (i.e. No clear Policy/Encouragement/Requirement/Accountability)
3. Reporting results to students, parents etc.	
4. Provision of resources to support teacher assessment	✓ Focus 2 (i.e. For secondary only (lower, upper secondary or both) or both primary and secondary or other)
5. National assessment mechanisms (if any)	
Teacher training on assessment (initial and in-service) and collaboration among teachers	✓ Focus 3 (i.e. Mathematics only or other subjects too)
6. Teacher initial training in terms of the programs of study offered by the universities, leading to teachers' graduate or post graduate diploma	✓ Stage (i.e. Period of establishment/Evaluation and Adjustment of policy)
7. Teacher professional development: in-service training courses, including mentoring system for newly appointed teachers	✓ Quality (i.e. In line with the literature)
8. Support offered by school advisory bodies (this may not be applicable for centralized systems such as Cyprus and Greece)	✓ Differentiation (i.e. Regarding the group of teachers or group of students)
9. Teacher cooperation about assessment within a school (including team teaching, exchange of visits among teachers etc.)	
10. Promoting networking between schools	
11. Evaluation mechanisms in regard to the policy on assessment and teacher professional development	

2.2.   Interviews with policymakers

Semi-structured interviews were conducted with policymakers and other stakeholders holding key positions regarding the promotion of the national policy in student assessment at the school level. In some countries (i.e. the Netherlands and Belgium), different actors from different levels of the educational system are responsible for promoting specific aspects of the policy for assessment whereas in more centralised systems, such as Cyprus and Greece, a specific group of officials (i.e. officers of the Ministry of Education) were able to answer all the questions. For each aspect of assessment policy, we collected interview data from at least three policymakers from each country, who were selected by using a purposive sampling procedure.

Initially, open-ended questions concerned with national policies and actions taken in relation to student assessment were designed together with a checklist that enabled us to generate data on each dimension and aspect mentioned in Table 1. The interview guide developed (see Interview Guide at https://ucy.ac.cy/formas/documents/Deliverable_2.1_Analysis_of_Policies/Deliverable_2.1_Interview _Guide.pdf) included 11 questions, each of which was followed by a number of follow-up and probing questions for clarification purposes. This structure ensured that all aspects and dimensions under investigation were addressed. Guidelines for analysing the interview data were also developed (see Guidelines   for   Analysing   Interview   Data   at https://ucy.ac.cy/formas/documents/Deliverable_2.1_Analysis_of_Policies/Deliverable_2.1_Guidelin es_for_Analysing_Interview_Data.pdf). Specifically, a record sheet for each interviewee was created, including two sheets which were used for documenting the scores for each aspect of the national policy examined. Respecting the anonymity of the interviewees, only the record sheet in its original form and not the transcripts of the interviews are presented. A description of the aspects and dimensions to be evaluated was provided along with examples when considered necessary. Detailed instructions on the coding of interview data were also provided.

The next section presents the main findings of the data analysis from both policy documents analysis and the policy makers interviews in each country in relation to the aspects and dimensions examined.

3. Main Findings

Data elicited from the policy documents analysis and the policymakers' interviews in each country were compared. The four within-country analyses showed that in each country, data elicited from the analysis of the policy documents analysis match the data elicited from the interview data. Thus, support for the internal validity of the study was provided. This section presents the results of the across country analysis of the data collected through both the content analysis of the policy documents and the policymakers' interviews.

3.1. Student assessment

The FORMAS project focuses on student assessment and aims to support secondary education teachers to conduct assessment in ways that support student learning. Thus, the first aspect investigated refers to national policies that address student assessment. Specifically, we examine policies that refer to: *a) the purpose of assessment (i.e. diagnostic, formative, summative), b) assessment techniques, c) reporting procedures to relevant stakeholders, d) the provision of supportive resources,* and *e) the existence (if any) of national assessment mechanisms.*

3.1.1. The purpose of assessment

Student assessment can serve a variety of purposes (Broadfoot, 1992; Brookhart, 2003; Gipps, 1994; Pellegrino et al., 2001; Torrance & Pryor, 1998). Based on the literature, the two main purposes of classroom assessment are the summative and formative purpose of assessment. Summative assessment is used for the recording of the overall achievement of a pupil in a systematic way (DES/WO, 1988). It aims at describing attainment, achieved at certain time, in order for comparisons to be made according to students' level of performance. On the other hand, formative assessment is used to identify students' learning needs and provide targeted support to help them improve. As mentioned above, formative assessment practices have been found to have a positive impact on student achievement (Hattie & Temperley, 2007; Herman, Osmundson, Ayala, Schneider, & Timms, 2006; Wiliam, Lee, Harrison, & Black, 2004). In addition, research, so far, has also shown that achieving both purposes with a single mechanism is not feasible (Harlen & James, 1997; Black & Wiliam, 1998; Kyriakides & Campbell,

2003; Kyriakides, Demetriou, & Charalambous, 2006). Therefore, examining whether educational policies support or promote the summative and/or the formative purpose of assessment is considered important.

Looking into policies regarding the purpose that student assessment is expected to serve we first note that all four countries have policies that refer both to the formative and the summative purposes of assessment. The analysis of policy documents revealed that in some documents across the four countries there is no explicit reference to the purpose of assessment and that some policy recommendations, although not officially stated, can be applied for achieving both purposes. Almost all policies are in line with the literature, referring to qualities of good assessment practice as these are recognized in the literature. In Cyprus and Greece, there is a requirement for summative testing at the end of each school year in order for students to move to the next grade. End of school exams are also established in higher levels of secondary education in the Netherlands, however if summative assessment reveals that the results do not meet expectations set, schools are questioned about the causes and are asked to outline their actions to monitor a student's progress. On the other hand, in Belgium, the *class council* is the central evaluation body in secondary education. At the end of the school year the class council decides whether a pupil has sufficiently reached the objectives of the teaching plan in order to move on to the next school year based on all assessment data collected throughout the school year.

In addition, in all four countries, there are policies that encourage the use of assessments that help students identify their learning progress so that they can make efforts to improve. However, most policies that refer to the formative purpose of assessment are not required to be implemented by schools but are stated as suggestions which are good to be implemented. No specific actions on behalf of the school, teacher or student are mentioned or required. This is also confirmed by policy makers in all four countries. On the other hand, policies that refer to summative assessment are mostly set as requirements and suggest specific actions that need to be performed. Interviewees from both Greece and Cyprus stressed that even though a number of initiatives to promote formative assessment can be identified, only summative assessment practices are officially established. In the Netherlands and Belgium where

a more autonomous system is established, again teachers are encouraged to use assessment formatively but they are not held accountable over it.

Looking into the other dimensions examined, one can see that almost all policies under this aspect refer to both primary and secondary education. In addition, none of them is specific to the subject of Mathematics but refer to other subjects at the same time. This applies for all aspects examined in the policy analysis and this finding will be discussed further in the final remarks section.

Based on the above, questions are raised as to whether schools actually implement assessments that promote learning or whether they inevitably shift their emphasis on the required summative assessment of students. Since summative assessment appears to have more political power, it is possible that formative practices are ignored. As mentioned by a stakeholder in the Netherlands, the end of year examinations influence teachers' other assessment practices since schools appear to adapt to the form and content of these exams. Similar comments were also made by policy makers in all countries. The analysis of both policy documents and interviews, showed that the differences between the two purposes of assessment are not appropriately acknowledged. Research argues that each purpose suggests a quite different approach to assessment in regards to the instruments to be used, assessment administration, as well as, the recording and reporting of assessment data. It is also argued that achieving both purposes with one mechanism is not possible. Thus, developing policy measures that specifically define the purpose of assessment is necessary in order to ensure their effective implementation.

3.1.2. Assessment techniques

Assessment techniques refer to the evaluation methods employed to assess students' learning (e.g. written assessment, oral assessment, performance assessment). Assessment techniques hold an important role in ensuring the quality and effectiveness of assessment, since they usually have an influence on how and what students learn. Current thinking in assessment recognizes that a variety of assessment techniques needs to be employed, as learning is multidimensional and cannot be adequately measured by a single technique (Brookhart, 2003; Gipps, 1994).

Examining the policies that make a reference *to assessment techniques*, one can notice that whereas all countries have policies that refer to assessment techniques that can be used for student

assessment, almost half of them (i.e. 12 out of 23) are stated as suggestions and are not required to be used. Furthermore, it is not clear whether these techniques are expected to be used for formative or summative purposes. For example, in the Netherlands, a Diagnostic Interim Test (DIT) was developed to be administered at the end of the lower grades of secondary education for the subjects Dutch, English and Mathematics. This test is expected to provide information on the stronger and weaker points of each student and with this information, teachers and students can work at their learning process. However, this policy has no continuity (i.e. it was administered only once in 2017) and based on the information provided by the country's policy makers during the interviews, the tool is difficult to be used for formative purposes, since at the moment of the administration of the DIT coincides with a high-stakes decision (i.e. at which level can the student be situated in the upper grades of secondary education). At the same time, there is also a policy requirement for a secondary school examination, which consists of a central (developed by Cito-National Institute for Educational Measurement) and school-based part and is done for summative purposes. Similarly, in Cyprus, a diagnostic test in Language, Mathematics and Science was administered to all first-grade secondary students in September 2017. However, the test was only administered once and its use was restricted in identifying students under risk. Although data were available, it is not clear whether these have been used for formative purposes for all students. In Belgium, the assessment and evaluation policy belong to the complete autonomy of the school. However, at the end of the technical, art and vocational education, schools are obliged to organize an integrated exam in which specific competences are tested, but the purpose of this assessment is not clearly defined and seems to be in line with the summative purpose of assessment. In Greece and Cyprus, even though authentic assessment techniques such as performance tests and student portfolios are encouraged as part of formative assessment, many policy documents (i.e. more than 60% of them) refer to the required used of written assessment for summative purposes.

All the above show that, in all four countries, teachers are mostly free to choose the assessment techniques to be used for assessing student learning and whether their assessment will be based on the use of one or a combination of assessment techniques. As mentioned above, not all techniques are appropriate to assess all learning objectives and the literature suggests that a variety of assessment techniques should be employed to ensure the internal validity of student assessment. Therefore, it is

important that policy measures provide more specific directions regarding the type and variety of assessment tools to be used for assessing student learning. Finally, the cases of both Cyprus and the Netherlands confirm that attempts to use assessment in order to identify students' learning needs are made. However, these attempts do not have a clear formative orientation and have no continuity which hinders their positive impact.

3.1.3. *Reporting results to stakeholders*

The communication of assessment results bridges the gap between the recorded data, their analysis and interpretation and their use by the involved participants. Reporting procedures deliver assessment results into the hands of the various intended users of the information in a timely and understandable manner (Roeber, 2003) and enhance the continuity and quality of students' learning experience (Berry, 2008). They also provide all intended users of assessment with knowledge of results that can be later used to make adjustments to support teaching and learning.

When analyzing policies related to the reporting of assessment results, one can notice the differences between countries with a centralized system (i.e. Greece and Cyprus) and countries with a decentralized system (i.e. Belgium and the Netherlands). In Greece and Cyprus, mathematics teachers are responsible to communicate assessment results for their subject to intended users. In addition, there are specific policies determining when and how results of summative assessment are communicated to stakeholders. For example, in Cyprus policy determines the grading scale to be used, the percentile given to oral and written performance in order to determine the student's final grade, as well as, when the results are communicated to students and parents. Once again, the emphasis is on summative assessment as no policy in both countries refers to reporting assessment for formative purposes. On the other hand, in Belgium and the Netherlands whereas the reporting of assessment results is a requirement, the purpose of reporting is not clarified and schools are free to decide how this will be done. For example, in the Netherlands, a mentor is assigned for each classroom and is expected to act as the connection with parents. He/she is usually also teaching in the particular classroom and therefore is able to report information based on student's results in different subjects, their progress and in relation to the specific classroom in terms of performance and behaviour.

Based on the above, the lack of a formative orientation in reporting assessment results can be identified. Whereas, providing constructive feedback to all stakeholders, and especially the student, is a prerequisite for assessment to aid learning, this does not appear to be recognized at policy level. Reporting assessment information is the first step to enable the use of assessment information, since no one can act on data they are not aware of. In addition, the methods, content and quality of this communication need to be addressed to ensure that assessment results are communicated in a way that allows adjustments to be made to support teaching and learning.

3.1.4. *Provision of resources to support student assessment*

Next, policies referring to the provision of resources that support teachers in the implementation of student assessment are examined. Resources in this report refer to any kind of support provided to teachers in order to support their practice in student assessment. Such resources may include amongst others item banks, battery of tests and financial support. Recourses are considered important as they contribute to achieving a system's educational objectives to the fullest by enhancing the support system provided for policy implementation (Hanushek, 1986, 1989; Hedges, Laine, & Greenwald, 1994). Adding to this, an optimal distribution and use of resources is needed in order for both efficiency and equity objectives to be achieved (Kyriakides, Creemers, & Charalambous, 2018).

The analysis shows that besides Greece, the other three countries have little or no policies referring to supportive resources. In the case of Greece, a number of centrally developed resources are identified, however, these are only listed as suggestions, their purpose is not clearly defined and schools are not required to use them. In Cyprus, performance indicators per subject are available and teachers are expected to plan, implement and evaluate their teaching based on them. However, all policy makers in their interviews recognized that this policy needs further improvement since teachers are not confident on how to use these performance indicators to aid their teaching. In addition, the large number of indicators identified makes it even more difficult for teachers to use them to plan their teaching and assessment. Similarly, in Belgium a policy developed in 2014 refers to evaluation indicators that aim to help teacher develop a clearer view of assessment. In the Netherlands, the DIT, as well as, the Cito

LVS tests are stated as a resource that can be used to support teachers, but again it is up to the schools to decide whether to use them and for what purpose.

The above results seem to show that teachers are not sufficiently supported to use assessment for formative purposes. Given that implementing assessment for formative purposes is more difficult to be achieved, the lack of supporting recourses makes it even more difficult for teachers to conduct assessment for formative rather than summative reasons. Although some supporting mechanisms can be identified in all four countries, these are not set as requirements and therefore it is up to the teachers to decide whether they will make use of them and how. Student assessment is a process that involves multiple steps and procedures and expecting that teachers are able and/or willing to implement it correctly on their own initiative is not so promising. It is evident that policy measures that provide adequate support in the form of various resources (i.e. item banks, financial support, assessment tools) are necessary in order for teachers to effectively use assessment in their everyday classroom practice.

3.1.5. National assessment mechanisms

Finally, policies related to national assessment mechanisms are investigated. In Cyprus and Greece, the only nation-wide assessment mechanism refers to the end of secondary school examinations. These are University entrance exams and are a requirement for all 6th grade students. Policymakers in both countries confirmed that the high stakes character of these exams, makes teachers focus their efforts throughout the school year on how to help their students achieve the highest possible performance. In Belgium, the only nation-wide mechanisms since 2002 is the National Assessment Program ('Peilingen'), which aims to collect and provide reliable information on the validity and feasibility of the attainment targets, on the outputs of the school system regarding those attainment targets, and on the effectiveness of processes used in the current school system to achieve these outputs. Although it is a policy that can be used for formative purposes, it is not a requirement. On the other hand, in the Netherlands, a number of nation-wide assessment mechanisms are established. First, in 2017 a policy that requires the school to collect information about the knowledge and skills of their students from the moment that they enter the school was developed. The school is required to compare this information with the expected developments. This information and comparison make it possible to adapt education

to the educational needs of individual students and/or groups of students based on their ability. When students seem to not benefit from their education, the school is expected to identify the reason and provide a possible explanation. They decide what should happen to eradicate certain backlogs and the students get the supervision that they need to adequately go through the educational program. For students with backlogs, education has been structured in such a way that noticeable attention can be spent on eradicating these backlogs. This policy clearly serves the formative purposes of assessment since it uses assessment results to adjust teaching and support student learning. In addition, since 2017 each school is required to develop a Programme of Assessment (PTA) and examination regulations that follow the country's' relevant law. This document states the processes of student assessment applied in each school.

Based on the analysis of policies regarding the first aspect examined, student assessment, we reach the conclusion that student assessment is recognized as important in all participating countries. Teachers are expected to assess their students for both summative and formative purposes. However, even though formative assessment practices appear to have positive impact on student learning (Kyriakides, 2005; Kyriakides & Creemers, 2008), formative assessment practices are not set as requirements. Policies referring to the use of assessment techniques and the reporting of assessment information to intended users hold teachers accountable only when summative assessment is concerned. Taking into account the lack of supportive resources for student assessment, questions are raised as to whether and how formative assessment practice is implemented.

3.2. *Teacher training and TPD in student assessment*

The second aspect investigated policies addressing teacher training and TPD in student assessment. Teacher training and professional development are considered important components of any effort to create effective schools (Darling-Hammond, 2000; Guskey, 2003; Smith & O'Day, 1991). Thus, we examine policies that refer to training in assessment as part of the programs of study leading to a graduate or post graduate teaching degree, in-service professional development and the support offered to schools by advisory bodies for the implementation of student assessment. Teacher collaboration within and between schools in relation to student assessment is also examined since research suggests

that professional learning of teachers takes place not only in organized trainings but also through a variety of less formal interactions (i.e., communities of practice, professional learning communities and teacher networks) (Richter et al., 2014).

3.2.1. Teacher initial training

Teacher initial training refers to training offered to prospective teachers as part of the programs of study leading to a graduate or post graduate teaching degree. Initial training is expected to adequately prepare teachers to effectively undertake their complex role in real school settings (Darling-Hammond, 2006). Given the large amount of time teachers spend in assessment-related activities, pre-service teachers' training is expected to provide, amongst others, the necessary training in student assessment. Here, the focus is on examining how official policy requires or supports teacher training in assessment before teachers enter the profession.

First, we notice that none of the four countries has a formal policy that defines assessment training as a prerequisite for teacher appointment. Interviews with policy makers in all countries revealed that educational authorities may have connections and common projects with higher educational institutions, however they have no saying in the curriculum or content development of undergraduate or postgraduate courses leading to a teachers' graduate or post graduate degree. This creates questions as to whether teachers are ready to assume their role as assessors, especially when they are newly appointed. Whereas, assessing students is included in their duties, no policies that require them to be knowledgeable and skilful in assessment are present. Requiring teachers to have developed a minimum set of assessment skills before entering the profession, could, therefore, be considered important. The fact that no initial training in assessment is required for teacher appointment, highlights further the role of in-service training in assessment.

3.2.2. Teacher in-service training and professional development

TPD is considered an essential mechanism for deepening teachers' content knowledge and developing their teaching practices in order to teach to high standards (Borko, 2004). Research suggests that teachers can improve the quality of their practice, by undertaking appropriate interventions and

participating in effective professional development programs (Creemers, Kyriakides, & Antoniou, 2013).

Looking into policies in respect to TPD in assessment (i.e. in-service training courses, mentoring systems for newly appointed teachers etc.) we notice that in all four countries higher institutions often offer TPD opportunities related to assessment for in-service teachers, but these are usually on a volunteer-only basis and are usually not linked to career development.

In Greece, for the years 2010 until 2011 newly appointed teachers were required to attend an introductory training course which included amongst others training in student assessment. However, this policy was discontinued. In Cyprus, there is a requirement for all newly appointed secondary teachers to attend a training program in order to develop their educational capacity before their appointment. Amongst the courses offered there is also a course focused on student assessment, however, this is set as an elective and not all teachers have to complete it. In 2018, the Cyprus Ministry of Education and Culture placed a stronger emphasis on student assessment and all secondary education teachers were required to attend a seminar on authentic assessment practices organized by the subjects' coordinators. Policy makers however, in their interviews emphasized that although this was a step towards the right directions, the duration and focus of this training was not adequate. In Belgium since 2009, all schools are obliged to create an annual plan for in service training. This plan takes into account every training effort that focuses on the development of teacher's knowledge, competences and attitudes. Student assessment can be part of this plan, however, this is not a requirement. In the Netherlands, in 2018, a policy regarding TPD was introduced as a requirement. Teachers were expected to account for their development in teaching by entering relevant information in the Teacher Register. However, the implementation of the Teacher Register was discontinued the same year (see https://www.vo-raad.nl/themas/36/onderwerpen/330 for more details).

TPD is a powerful tool for fostering improvements in teaching (Kennedy, 2016). However, the analysis of both policy documents and interviews in all four countries shows that most training opportunities offered are on a volunteer basis. This raises questions as to whether all teachers and especially the ones who lack the basic skills in assessment are receiving the necessary training. Thus, the need for more structured and focused policies in teacher professional development is assessment is

highlighted. In addition, even when professional development opportunities are offered these are usually short in duration and have no continuity as is the case in most European countries. Research has shown that on–off professional development workshops are not typically aligned with the participants' existing practices, needs and priorities for improvement and do not reliably lead to changes in teacher practice (Antoniou & Kyriakides, 2013; Loucks-Horsley et al., 1998). Therefore, the duration of professional development programs, as well as, the span of time over which these programs are spread (Desimone, 2009), also need to be taken into consideration when designing policy measures that aim to support assessment practice through professional development interventions. Finally, it is argued that in order for professional development to be considered effective it should provide structured professional learning that results in changes in teacher practices and improvements in student learning outcomes (Darling-Hammond, Hyler, & Gardner, 2017). This underlines the importance of examining the impact of professional learning initiatives on both teacher practice and student learning.

3.2.3. *Support offered to teachers regarding student assessment by school advisory bodies*

The next aspect under examination, was the support offered to teachers regarding student assessment by school advisory bodies. Based on the analysis of the national policies, this aspect was not applicable for Greece. In Cyprus, official in-service teacher is offered primarily by the Pedagogical Institute of Cyprus. The Pedagogical Institute of Cyprus offers professional training courses through a series of optional seminars. Since 2015, the Cyprus Ministry of Education and Culture introduced a policy measure about Professional Learning. This measure requires all schools to develop a plan for school-based professional development based on the particular needs of the teachers and in relation to the particularities of the school unit. Each school chooses their area of priority and designs actions for the development of teachers in relation to the needs identified. The Professional Learning program is implemented with the support of the Pedagogical Institute. In the Netherlands, in 2015 the CvTE has facilitated a number of schools in researching the added value of Diagnostic Interim Test (DIT) The DIT was originally developed to serve summative purposes. However, later on it was suggested that the tool can also be used for formative purposes. The CvTE provided support to students and teachers by helping them identify how the tool fits in the cycle of formative evaluation and in what way the schools

can use the diagnosis provided to improve their practice. However, this support was only offered to the 24 schools participating in the pilot study and was then discontinued. As a policy maker in the Netherlands further explained, the fact that the tool was not primarily designed to serve formative purposes made its formative use difficult to be achieved. In addition, since November 2018 teachers are entitled to financial aid (€600 per fulltime teacher) and 83 hours of TPD and of per school year. The hours and the budget are only to be spent on activities that are assigned by the supervisor, but only if the teacher agrees with those and there is no specific mention to training focused on student assessment. In Belgium, each educational umbrella organization has its own School Advisory Service ('Pedagogische begeleidingsdienst'), which offers professional internal support to schools if the ask for it. Schools can call on them for educational and methodological advisory services (e.g. innovation projects, self- evaluation projects, support initiatives). School advisors work across schools for the in-service training and support of school principals and school teams. They also support and stimulate quality assurance mechanisms in the schools. All three policy makers in their interviews emphasize the shift from supporting individual teachers towards support at the system level. They also mention that several in service trainings are organized by the school advisory services to develop teachers' assessment and reporting policy, based on school advisory initiatives and on demand of school leaders. However, there is no clear policy on this.

It is argued that when decision on professional development is primary school-based and without evaluating the training needs of teachers, it is more possible that the program is adjusted to what the school considers as important without taking into consideration the knowledge base provided by the relevant literature, as well as, the needs of participating teachers. Thus, policy measures that support the use of school-based approaches to TPD should take into account that beside the awareness of critical contextual characteristics, the necessary knowledge base is required to ensure that initiatives are research based and thus are more likely to have a positive contribution to teachers' practice.

3.2.4. Teacher collaboration within and between schools

"Collaboration can be defined as joint interaction in the group in all activities that are needed to perform a shared task" (Vangrieken, Dochy, Raes, & Kyndt, 2015, p. 23). High quality teacher collaboration

can have positive effects on teachers and their students and assessment-focused collaboration is recognized by teachers as a very helpful tool for improving their practice (Ronfeldt, Farmer, McQueen, & Grissom, 2015). However, examining the aspect of teacher collaboration within and between schools in relation to student assessment data analysis shows that none of the four countries has any relevant requirements. In Belgium and the Netherlands, teachers and schools are encouraged to cooperate in order to improve their educational practice. Still, this is not set as a requirement but mostly as a suggestion of good practice. More specifically, in the Netherlands in 2015 a description of eight best practice examples of teacher and school cooperation was created. In addition, since 2016 there is a policy regarding school networking in relation to formative assessment aiming to facilitate the use of formative assessment in the classroom. In five network meetings, schools are expected to work together on issues related to assessment. The main purpose of these meetings is for schools to learn from each other and in the period between meetings differentiated supervision and advice is provided by the SLO. In Greece, since 2016 teachers of different subjects are encouraged to cooperate in order to develop a multidisciplinary educational project, however there is no clear focus on assessment. However, the cooperation encouraged is across different subjects, and no policy is identified in regard to the collaboration of same subject teachers and particularly mathematics teachers. In Cyprus, there are no established policies regarding teacher and school cooperation and networking regarding student assessment. It is, however, important to note that all Cypriot policymakers interviewed confirm that these are important elements for improving teacher quality.

Creating the opportunities for teacher collaboration and school networking is difficult to be achieved. In order for collaboration to be effective certain conditions need to be fulfilled. For example, scheduled common planning time (Main, 2012), consistent meeting times (Saunders et al., 2009) and collaborative policies and accountability systems (George & Alexander, 2003). Therefore, although collaboration at both teacher and school level are recognized as important by policymakers in all four countries, the conditions necessary for high level collaboration to be achieved are not created.

This section looked at the policies addressing teacher pre- and in-service service training, as well as, teacher and school collaboration in regards to assessment. Based on the above, we reach to the conclusion that although a number of formal and informal opportunities for teacher training in

assessment are available, some teachers may not sufficiently be educated on how to assess their students. Data analysis showed that in all four countries teachers are encouraged to take advantage of given opportunities; however, no systematic training to ensure quality of assessment practice (and especially assessment in mathematics) is evident.

3.3. Evaluation mechanisms in regard to the policy on assessment and TPD

The increasing emphasis on evidenced–based policy making highlights the need for the establishment of mechanisms that evaluate the relevance, impact, effectiveness, efficiency, and sustainability of the interventions at policy level. Evaluation mechanisms provide credible, reliable and useful information that can be used to generate and/or improve policy making (Sanderson, 2002). Thus, the third aspect examined in this report refers to policies relevant to mechanisms developed by educational systems in order for them to be able to evaluate the existing policies on assessment and teacher professional development.

Specifically, if an evaluation mechanism is present then it is examined how frequently an evaluation is undertaken, as well as, the number of sources of data used (i.e., views of different stakeholders) in evaluating a national policy. The stage of policy implementation at which evaluation is performed is also examined (e.g. at the end of implementation, at regular periods during implementation). This helps us investigate the continuity of the evaluation and if adaptations based on evaluation results can be made. Furthermore, the quality of the processes employed is taken into consideration in order to examine whether data elicited can be used to help policy makers make decisions that may improve national policy. Finally, how the results are used and for whom, is investigated. This allows us to identify whether the evaluation mechanism is used for formative or summative purposes and if stakeholders have access to information that can help them make changes for improvement.

Based on the analysis of the policies, no evaluation mechanisms in regard to the policy on assessment and TPD exist in three out of the four participating countries (i.e., Cyprus, Greece, and Belgium). More specifically, in Belgium, the principle of educational freedom grants schools the right to develop a school-specific vision on quality of education within the boundaries set by the Flemish

government and the Flemish parliament. This vision is elaborated in a school-specific curriculum and school-specific teaching and assessment practices. Therefore, the assessment and evaluation policy belongs to the complete autonomy of each school. As a result, no national evaluation mechanisms exist. Furthermore, in Cyprus and Greece, despite the number of policies defining how and when assessment is taking place (when it comes to summative assessment) there are no established evaluation mechanisms of these policies even those that refer to summative assessment. For example, in both countries there are exams at the end of the 12th grade in order for students to acquire a position at a local university. Despite the high stakes character of these exams and their impact on students' lives, there is no meta-evaluation mechanism of these exams to monitor and evaluate their implementation. In addition, schools are expected to implement suggested/required policies but have no responsibility to monitor their implementation. As mentioned in the interviews with the policymakers in the two countries (i.e. three policy makers per country), policies in the two countries can be terminated, replaced or altered but this is usually based on revised or new educational policy, political priorities, or educational trends (as in the case of formative assessment). No evaluation of the policies' implementation or impact on learning is performed.

The only country out of the four participating countries that has some policies related to evaluation mechanisms is the Netherlands, even though not all policy makers stated this in their interviews. Based on the analysis of the policy documents, when it comes to new laws, the Dutch law mandates that these are evaluated five years after introduction. Therefore, any new law related to student assessment and/ or TPD is expected to have an evaluation clause. This also stands for separate policy initiatives introduced by ministerial order, often half way through the term during which the ministerial order is in effect, and at the end. For this purpose, a separate monitoring program will be ordered in most cases, usually carried out by an independent research institution. This program will usually examine whether the policy initiative has the desired effect and if the policy rule in general is effective. In addition, provisional and final evaluations are written, which act as the basis for termination, renewal as is, renewal after alteration, or it can form the basis for changing a law. Subsidized policies are usually monitored informally year-round by the public servant at the Ministry who is tasked with the functioning of the policy measure. A yearly activity report and financial overview is a mandatory part

of the subsidy process. If a monitoring program is set up, it depends on the methodology how often data is collected, but generally the ministry expects (preliminary) results once a year. When examining the validity of evaluation instruments used, policy documents analysed, show that experiences from independent scientific studies and from previous (subsidized) research projects are used for the design and implementation of the evaluation mechanisms. When a monitoring program for a specific policy measure is necessary, tenders (i.e. academic and/or institutions) are involved and are expected to undertake the program following sound scientific methods. In addition to this, larger monitoring programs are usually overseen by a committee which includes scientific, educational and policy experts.

When it comes to the sources of data used to evaluate a policy, policy documents suggest that the Ministry of Education is in constant dialogue with stakeholders. Stakeholders include teacher organizations, the organizations of the various curricular disciplines, the interest organization of schools and school leaders (Vo-Raad), student organizations (LAKS) and parents (Ouders & Onderwijs). Exactly who is consulted may vary depending on the specific policy measure. However, the perspectives of schools and school leaders, teachers, students and parents are usually included. In addition, since the Dutch secondary education concludes with final central exams for most subjects a common practice is to measure the effects of a policy on the grades of students who have been affected by that specific policy measure. When students' grades are not considered a suitable measure, other measures depending on the policy are used including student motivation, mastery and wellbeing. Finally, as mentioned in the policymakers' interviews the system recognizes public and political acceptance as a prerequisite for the effective implementation of a policy. Thus, if a measure is very popular or unpopular is also taken into consideration.

From the above it is evident that the Netherlands recognize that evaluation can offer a valuable insight on the implementation of a policy. However, evaluation mechanisms established refer only to policies for which schools or teachers are held accountable for. On the other hand, as the analysis of data showed, most assessment related policies are stated as encouragements and schools and teachers are given the autonomy to formulate and implement base on their own needs and judgment without being held accountable. This means that the evaluation clause in these cases is applicable.

Looking into the third and final aspect examined, we reach to the conclusion that besides the Netherlands no other participating country (i.e., Cyprus, Greece and Belgium) has established mechanisms to evaluate their policy on assessment and TPD. This is recognized as problematic given that evaluation of policy reforms may give important insights on how the reform policy is implemented and provides data based on which a policy can be continued, terminated, or improved to achieve its purpose. The FORMAS project attempts to encourage and support the use of assessment for formative purposes; however, this formative orientation is not evident at a policy level. Whereas policy makers recognize the importance and value of formative assessment, they do not seem to engage in formative evaluation of their own policies.

Final Remarks

The FORMAS project attempts to encourage policy makers to reform assessment policies and to establish teacher support mechanisms for the effective implementation of formative assessment. In order to do so, a critical analysis of current national policies of the participating countries related to student assessment was conducted. The across-country analyses conducted based on the policy documents and the policy makers' interviews, helped to identify similarities and differences among the four participating countries.

First, it was found that in many policy documents in all four countries there is no explicit reference to the purpose of assessment, suggesting that the differences between the two purposes of assessment are not appropriately acknowledged. This is considered as problematic since the literature argues that achieving both purposes using a single mechanism is not feasible. Thus, it is important that when policy measures in relation to assessment are introduced the purpose that assessment is expected to serve to be clearly stated. Clarifying the purpose of an assessment is necessary in order for the appropriate procedures, methods and tools to be used. It is also needed in order to avoid possible misconceptions both in terms of teacher practice and in terms of support given to teachers. Finally, clarifying the purpose is important in order for all stakeholders to be able to evaluate both the effectiveness of a policy measure, as well as, the quality of its implementation.

Despite the fact that most policies examined do not make an explicit reference to the purpose of assessment, in all four countries there are policies that indicate that the formative purpose of assessment is highly appreciated and is recognized as a means to improve student learning and address students' differentiated learning needs. Especially over the last years, a number of policies with a formative orientation can be identified in all four countries. However, most of them are merely suggestions and it is up to the schools or teachers to decide whether they will act upon them. In addition, these policies usually outline what the literature suggests as good assessment practice but no specific actions or steps are described that can give a more detailed guidance and support to intended users. Specifically, support in the form of professional development, advisory guidance or recourses is available only on a volunteer basis and usually not in a structured framework. On the other hand, policies on summative assessment are usually far more specific and most times have a requirement clause. Given the connection between summative assessment, accountability and high stakes decisions, it is possible that teachers place more emphasis on using assessment for summative rather than formative purposes. In addition, implementing formative assessment is more challenging as it requires more time and effort on their behalf. Indeed, research suggests that whereas teachers appear to hold positive views towards assessment that aids learning (Brown, 2004; Sach, 2012), their practice still appears to be outcome - oriented (Earl & Katz, 2000; Kahn, 2000; Lock & Munby, 2000; Christoforidou & Xirafidou, 2014). In order for teachers to shift their emphasis on formative assessment, it is not enough for policy measures to acknowledge the formative purpose of assessment and its impact on student learning, but the necessary support to teachers should also be provided. The analysis showed that no policies that require them to be knowledgeable and skilful in assessment are present. Furthermore, most training opportunities offered are on a volunteer basis. As the literature confirms, teachers report insufficient training in classroom assessment both during their teacher preparation programs and their in-service professional development (DeLuca & Klinger, 2010). The challenges of availability and adequacy of resources that support effective assessment practice is thus highlighted. This calls for the introduction of policy measures that outline professional standards in relation to assessment practice and especially the implementation of formative assessment in mathematics. In

addition, it is not enough for policy measures to acknowledge the importance of formative assessment practice but they should also find ways to hold teachers accountable for their implementation.

Another issue raised has to do with the training of teachers in assessment. The analysis highlighted that policies that ensure that all teachers are sufficiently trained through appropriate professional development interventions both in terms of content and duration are needed. Although teachers have unique strengths and weaknesses, they most often receive "one-size-fits-all" professional development, such as one-time workshops, that are less likely to improve the quality of their assessment practice, as well as learning. The FORMAS project aims to develop and test a comprehensive dynamic approach to TPD in order to support teachers in conducting assessment for formative purposes. This approach recognizes that first teachers' professional needs in relation to assessment need to be identified and thus a comprehensive framework for measuring teacher skills has been developed (see Deliverable 2.2). Based on this framework, a tool for measuring teachers' skills in assessment (see Deliverable 3.3) has been designed and will be administered before the professional development intervention in order to identify teachers' professional needs in relation to assessment. This will allow us to provide a more focused and targeted professional development program that will match the identified needs of each group of teachers. Moreover, effective professional development is expected to result in changes in teacher practices and student learning (Darling-Hammond, Hyler, & Gardner, 2017). For this reason, the FORMAS project will conduct an experimental study which will examine the impact of the dynamic approach to TPD on improving teachers' assessment skills and on promoting students' cognitive and metacognitive learning outcomes. Based on the results of the study, we will be able to provide specific and evidence-based policy guidelines and relevant supporting mechanisms that may help the effective implementation of formative assessment practices. The project, through this study, aims also to develop awareness among policymakers for introducing an evidence-based and theory-driven approach in the formation and implementation of assessment related policies.

Furthermore, the analysis of the available policies in the four participating countries revealed the lack of focus of available policies in regards to specific subjects or age group. Throughout the analysis, it was found out that almost all assessment policies did not make a particular reference to the subject of mathematics. This is an important finding since current views of effective mathematic

instruction value the complexity of mathematics (Boaler, 2008) and require teachers to have developed the necessary skills to assess students' conceptual understanding, as well as, their problem-solving and reasoning abilities (Suurtamm et al., 2010). Thus, policies that have a particular focus on mathematics assessment are considered necessary.

Finally, data analysis revealed the absence of evaluation mechanisms in three out of the four participating countries (i.e. Cyprus, Greece and Belgium). In the Netherlands evaluation mechanisms are present in the case of educational laws or policies that have an accountability element. However, no evaluation mechanisms for formative oriented policies were identified. This is because most policies that refer to student formative assessment come as recommendations and it is up to the schools and teachers to decide whether they will follow these recommendations or not. Monitoring and evaluation mechanisms are considered critical tools as they influence decision-making on the basis of lessons learned. Based on the results of the evaluation, decisions to improve, reorient or discontinue an intervention or policy are made. Especially when it comes to education the evaluation of interventions and/or policies is necessary since these have an impact on students and their learning. Therefore, the introduction of evaluation mechanisms for policies related to assessment is required.

Research reveals that formative assessment practices have a positive impact on student achievement (Hattie & Temperley, 2007). However, in order for formative assessment to become an actual practice among teachers, support at system level should be provided. Introducing new or adjusting existing policies based on available an evidence based and a theory driven approach is required if we wish to ensure that assessment practice is appropriately supported and that it has a positive impact on students' learning.

References

Antoniou, P., & Kyriakides, L. (2013). A Dynamic Integrated Approach to Teacher Professional Development: Impact and Sustainability of the Effects on Improving Teacher Behavior and Student Outcomes. *Teaching and Teacher Education, 29*(1), 1-12.

Berry, R. (2008). *Assessment for Learning*. Hong Kong: Hong.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, *5*(1), 7-74.

Boaler, J. (2008). *What's math got to do with it?* London: Penguin Press.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational researcher*, *33*(8), 3-15.

Broadfoot, P. (1992) Exploring the forgotten continent: a traveller's tale. *Scottish Educational Review, 26*(2), 88-96.

Brookhart, S.M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice, 22*(4), 5–12.

Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, *11*(3), 301-318.

Christoforidou, M., & Xirafidou, E. (2014). Using the dynamic model to identify stages of teacher skills in assessment. *Journal of Classroom Interaction*, *49*(1), 12-25.

Cohen, D.K., & Hill, H.C. (2001). *Learning policy*. New Haven, CT: Yale University Press.

Creemers, B., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London and New York: Routledge.

Creemers, B.P.M., & Kyriakides, L. (2015). Developing, testing and using theoretical models of educational effectiveness for promoting quality in education. *School Effectiveness and School Improvement, 26*(1), 102-119.

Creemers, B. P., Kyriakides, L., & Antoniou, P. (2013). A dynamic approach to school improvement: Main features and impact. *School Leadership & Management*, *33*(2), 114-132.

Darling-Hammond, L. (2000). Teacher quality and student achievement: a review of state policy evidence. *Education Policy Analysis Archives, 8*(1), http://epaa.asu.edu/epaa/v8n1/

Darling-Hammond, L. (2006). Securing the right to learn: Policy and practice for powerful teaching and learning. *Educational Researcher, 35*(7), 13-24.

Darling-Hammond, L., Hyler, M. E., Gardner, M. (2017). *Effective Teacher Professional Development*. Palo Alto, CA: Learning Policy Institute.

DeLuca, C., & Klinger, D.A. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice, 17*(4), 419-438.

DES/WO (1988) *National Curriculum Task Group on Assessment and Testing—a report.* London: DES.

Desimone, L. M. (2009). Improving impact studies of teacher's professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181-199.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational evaluation and policy analysis*, *24*(2), 81-112.

Earl, L., & Katz, S. (2000). Changing classroom assessment: Teachers' struggles. In N. Bascia & A. Hargreaves (Eds.), *The sharp edge of educational change* (pp. 97-111). London: Routledge.

George, P.S., & Alexander, W.M. (2003). *The exemplary middle school*. Wadsworth Publishing Company.

Gipps, C. (1994). *Beyond testing*. London: Routledge Falmer.

Guskey, T. R. (2003). What makes professional development effective? *Phi delta kappan*, *84*(10), 748-750.

Guskey, T. R., & Sparks, D (2004). Linking professional development to improvements in student learning. In E.M. Guyton & J.R. Dangel (Eds.), *Research Linking Teacher Preparation and Student Performance: Teacher Education Yearbook XII*. Lanham, MD: Rowman & Littlefield Education.

Hanushek, E.A. (1986). The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature, 24*, 1141-1177.

Hanushek, E.A. (1989). The impact of differential expenditures on student performance. *Educational Research, 66*(3), 397-409.

Harlen, W., & James, M.J. (1997). Assessment and learning: differences and relationship between formative and summative assessment. *Assessment in Education, 4*(3), 365-380.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112.

Hedges, L.V., Laine, R.D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes (An exchange: Part 1). *Educational Researcher*, *23*(3), 5-14.

Herman, J.L., Osmundson, E., Ayala, C., Schneider, S., & Timms, M. (2006). *The nature and impact of teachers' formative assessment practices. CSE Technical Report #703.* National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Khan, E. A. (2000). A case study of assessment in a grade 10 English course. *The Journal of Educational Research, 93,* 276-286.

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of educational research, 86*(4), 945-980.

Kyriakides, L. (2005). Extending the comprehensive model of educational effectiveness by an empirical investigation. *School Effectiveness and School Improvement*, *16*(2), 103–152.

Kyriakides, L., & Campbell, R.J. (2003). Teacher Evaluation in Cyprus: Some conceptual and methodological issues arising from Teacher and School Effectiveness Research. *Journal of Personnel Evaluation in Education, 17*(1), 21-40.

Kyriakides, L., & Creemers, B.P.M. (2008). Using a multidimensional approach to measure the impact of classroom level factors upon student achievement: a study testing the validity of the dynamic model. *School Effectiveness and School Improvement*, *19*(2), 183-205.

Kyriakides, L., Creemers, B.P.M., & Charalambous, E. (2018). *Equity and Quality Dimensions in Educational Effectiveness.* Dordrecht, the Netherlands: Springer.

Kyriakides, L., Creemers, B.P.M., Antoniou, P., Demetriou, D., & Charalambous, C. (2015). The impact of school policy and stakeholders' actions on student learning: A longitudinal study. *Learning and Instruction, 36,* 113-124.

Kyriakides, L., Demetriou, D., & Charalambous, C. (2006). Generating criteria for evaluating teachers through teacher effectiveness research. *Educational Research, 48*(1), 1-20.

Kyriakides, L., Georgiou, M.P., Creemers, B.P.M., Panayiotou, A., & Reynolds, D. (2018). The impact of national educational policies on student achievement: a European study. *School Effectiveness and School Improvement*, *29*(2), 171-203.

Lock, C.L., & Munby, H. (2000). Changing assessment practices in the classroom: A study of one teacher's change. *The Alberta Journal of Educational Research*, *46*(3), 267-279.

Loucks-Horsley, S., Hewson, P.W, Love, N. & Stiles, K. (1998). *Designing professional development for teachers of science and mathematics.* Thousand Oaks, CA: Corwin.

Main, K. (2012). Effective middle school teacher teams: A ternary model of interdependency rather than a catch phrase. *Teachers and Teaching, 18*(1), 75-88.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Richter, D., Kunter, M., Klusmann, U., Lüdtke, O., & Baumert, J. (2014). Professional development across the teaching career: Teachers' uptake of formal and informal learning opportunities. In *Teachers' Professional Development* (pp. 97-121). Brill Sense.

Roeber, E.D. (2003). *Assessment models for No Child Left Behind.* Denver, CO: Education Commission of the States.

Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal, 52*(3), 475-514.

Sach, E. (2012). Teachers and testing: an investigation into teachers' perceptions of formative assessment. *Educational Studi*es, 38(2), 261-276.

Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public administration, 80*(1), 1-22.

Saunders, W.M., Goldenberg, C.N., & G. Gallimore, R. (2009, December). Increasing achievement by focusing grade-level teams on improving classroom learning: A prospective, quasi-experimental study of Title I schools. *American Educational Research Journal, 46*(4), 1006-1033.

Scheerens, J. (2016). *Educational Effectiveness and Ineffectiveness: A Critical Review of the Knowledge Base.* Dordrecht, The Netherlands: Springer.

Smith, M.S., & O'Day, J. (1991). Systemic school reform. In S. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233-268). Philadelphia: Falmer Press.

Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' emerging assessment practices in

    mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy,*

    and Practice, 17(4), 399-417.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: teaching, learning and assessment*

    in the classroom. Buckingham, UK: Open University Press.

Vangrieken, K., Dochy, F., Raes, E., & Kyndt, E. (2015). Teacher collaboration: A systematic review.

    Educational research review, 15, 17-40.

Wiliam, D., Lee, C., Harrison, C., & Black, P. J. (2004). Teachers developing assessment for learning:

    Impact on student achievement. *Assessment in Education: Principles Policy and Practice, 11*(1),

    49-65.