This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Proceedings of the International Conference

"Developing Educational Policies to Promote Formative Assessment: The Contribution of Educational Research"

April 21 & 22, 2021
Online via Zoom (operated by the University of Cyprus)

held for the purposes of the project
Promoting Formative Assessment: From Theory to Policy and Practice (FORMAS)

under the

Erasmus+ Programme Key Action 3, Support for Policy Reform Forward looking cooperation projects

University of Cyprus

ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ, ΠΟΛΙΤΙΣΜΟΥ, ΑΘΛΗΤΙΣΜΟΥ ΚΑΙ ΝΕΟΛΑΙΑΣ

HELLENIC REPUBLIC
National and Kapodistrian University of Athens

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Υπουργείο Παιδείας, Έρευνας και Θρησκευμάτων

UNIVERSITY OF TWENTE.

Ministry of Education, Culture and Science of the Netherlands

University of Antwerp

Vlaanderen
is onderwijs & vorming

Contents

Welcome Speech by the Coordinator of the Project

Leonidas Kyriakides

Department of Education, University of Cyprus, Nicosia, Cyprus

Email: kyriakid@ucy.ac.cy

Dear Distinguished Guests and Colleagues

Good afternoon.

On behalf of all the members of our project entitled "Promoting Formative Assessment: From Theory to Policy and Practice", I would like to welcome you to our online international conference. In our conference we will share with you the main results and intellectual outcomes of our three-year project which has received support from the Key Action 3 of the Erasmus+ Programme. This project involves eight organizations from four European countries: Belgium, Cyprus, Greece, and the Netherlands.

The ultimate aim of our project was to generate policy guidelines for promoting formative assessment by considering also the policy on Teacher Professional Development (TPD) in assessment. To achieve this general aim, we initially conducted a critical review of the current policies on assessment in the four participating countries. We also made use of the literature in order to define professional standards in assessment. Proposing standards only makes sense when you are in a position to measure the skills of teachers in relation to these standards and use this measurement for defining teachers' professional needs. As a consequence, valid instruments measuring teachers' professional needs in assessment were also developed. Tomorrow, you will have the chance to learn about the framework that was used to measure teachers' skills in conducting assessment and about the results of the first phase of our project that helped us establish valid instruments to measure teachers' skills in assessment and identify professional standards.

Since our research consortium argues that policy reforms should be based on a theory which has been empirically tested, this project has not only proposed standards and a specific approach to TPD on assessment but also developed a specific course addressing the professional needs of teachers in assessment. We also attempted to evaluate its impact on improving assessment skills and through that on promoting student learning outcomes (cognitive and meta-cognitive).

For this reason later today results of the main phase of the project which refer to the impact of the proposed course will be presented. In addition, during this conference, teachers coming from different countries who participated in this TPD course will share their experiences with us. I am very grateful to them for helping us understand the challenges that they had to face in attempting to improve their assessment practice.

At this point let me remind you that the growing accountability framework, the need for higher learning outcomes and the recognition of formative assessment as a key factor for teacher effectiveness have resulted in an increased need for teacher competency in the area of student assessment. Although teachers spend a large amount of time in assessment-related activities, and hold positive views towards formative assessment, they appear to use assessment mainly for summative reasons. At the same time one can see that professional development in assessment appears to be a controversial issue in the literature and very few studies searched for the impact of different approaches on improving assessment practice. Therefore, the presentations of teachers may help us not only identify the strengths and possible limitations of the dynamic approach but may also reveal any differential and country specific effects of the intervention.

I also look forward to the presentations of country-specific projects run from different researchers of the partner organizations which support us in our attempt to develop policy guidelines that are context specific. In order to develop policy guidelines that are context specific and can contribute in establishing support mechanisms promoting formative assessment, we will also have the chance to listen tomorrow to the views of different school stakeholders and policy makers coming from two of the participating countries. Their views may also help us evaluate the interpretive validity of the intellectual outputs of this project. We are therefore very grateful to the representatives of the Ministry of Education in Cyprus and of the Flemish Pupils' Council in Belgium.

At this point, I would like to express my sincere gratitude to all teachers and school stakeholders who participated in this project and especially those who will share their views with us during this conference. My sincere thanks also go to Professor Leonor Santos, from the University of Lisbon in Portugal, who will act as our critical friend and discussant and I am sure that her presentation will help us develop further the outcomes of this project.

Finally, I would like to thank all the participating organizations in this 3-year project: the National and Kapodistrian University of Athens in Greece, the University of

Twente in the Netherlands, the University of Antwerp in Belgium, the Ministry of Education, Culture, Sport and Youth in Cyprus, the Flemish Agency for Higher Education, Adult Education, Qualifications and Study Loans in Belgium, the Ministry of Education, Research and Religious Affairs in Greece, the Ministry of Education, Culture and Science in the Netherlands and the University of Cyprus. My special thanks go to the Minister of Education of Cyprus, Mr. Prodromos Prodromou, who is joining our conference today, and to the Department of Secondary Education of the Ministry of Education of Cyprus, for their efforts and commitment in achieving the highest possible outcomes through this project.

Once again, a very warm welcome to our international conference and I hope that you will find it fruitful and pleasant.

Welcome Speech by the Minister of Education, Culture, Sport and Youth of Cyprus

Prodromos Prodromou

Cyprus Ministry of Education, Culture, Sport and Youth, Nicosia, Cyprus

Email: minister@moec.gov.cy

Distinguished guests and delegates of the research community on school effectiveness, dear Project officer from the Erasmus+ Programme of the European Union, who has supported and co-funded the FORMAS project, the results are to be presented at this international conference, welcome 'virtually' to Cyprus and to this conference. The Erasmus + Programme has supported us in dozens of other projects in which the MOECSY had the opportunity to benefit from its participation, either as coordinator or as a partner, as is in this case. For this enduring support in our efforts for better schools and education, I am expressing our gratitude.

I am also welcoming the representatives, researchers and educators, from the participating countries, Belgium, Netherlands and Greece. Together with researchers from the University of Cyprus, and officers and secondary level teachers from the MOECSY have closely worked together for three years researching the topic of Formative Assessment and seeking ways to inform our policies on effective implementation practices.

Indeed, we are at a point where we are intensifying our efforts towards improving our students' learning attainment. We have not allowed the COVID-19 pandemic to hinter our efforts and actions; instead, we have seized the opportunity to accelerate the implementation of our plans for digital transformation. The country's ranking at the last international TIMSS survey was a relief for us. It has provided an impetus for working harder for even better results at TIMSS and the PISA, and other surveys.

Student assessment is an indispensable and integral practice in the education process. Acknowledging the dynamics of formative assessment, not only at the student level but also as regards teachers and the school level, we have abandoned end of year summative examinations. They used to be the norm for many years for our secondary-level schools. Instead, we are gradually introducing, as from 2019, two separate semester formative examinations. Semester examinations are designed to inform teachers and

schools about each student's attainment, identify learning needs and facilitate reshaping teaching to meet these needs.

Day-to-day student assessment used to rely mainly on written unit tests. According to our new approach for formative assessment, teachers are encouraged to use frequent multifacet interactive, both oral, written and performance assessments. This day-to-day assessment framework is also introduced for primary level schools, where there used to be an absence of a coherence learning attainment assessment policy.

However, we are at the beginning of the transition process from the currently used summative assessment practices to blinding methodologies where both formative and summative assessment approaches will be enacted, each one for its contribution. Further on, many actors in the process of education in our educational system still share the false assumption, as described by Margaret Heritage, professor at UCLA, that formative assessment is "a particular kind of measurement instrument, rather than a process that is fundamental and indigenous to the practice of teaching and learning".

It is, therefore, with great pleasure that we realise that this project has focused on the role of teachers and has developed a framework for empowering them in ways of practical implementation for formative assessment as an integral part of their teaching. Furthermore, it is very encouraging that by using experimental design, this project has also validated that its suggestions for formative teaching have a positive impact on student learning outcomes.

With these thoughts in mind, I welcome this well-organised international conference on "Developing Educational Policies to Promote Formative Assessment". I would also like to thank, once again, the project officer of the Erasmus+ Programme of the EU Emilia Venot, professor Leonidas Kyriakides for coordinating this project, as well as the research community from the Universities of Athens, Twente and Antwerp, the delegates from the Ministries of education in each one of the participating countries and the participating teachers. I wish you the most fruitful, exciting and stimulating discussions and exchange of knowledge for the mutual benefit of our education.

Integrating Research on Teacher Professional Development and Educational Effectiveness: The Dynamic Approach

Panayiotis Antoniou

Department of Education, University of Cyprus, Nicosia, Cyprus

Email: antoniou.panayiotis@ucy.ac.cy

Abstract: This paper refers to the Dynamic Approach towards teacher professional development which attempts to merge research findings on teacher effectiveness and teacher professional development. The theoretical framework and the major features of the DA are presented. It is argued that the DA can be effectively implemented through five steps: Establishing clarity and consensus about aims and objectives, identifying needs and priorities for improvement through empirical investigation, provision of improvement guidelines, reflection opportunities and coaching on effective teaching by the Advisory and Research Team, establishing a formative evaluation mechanism and finally establishing a summative evaluation system. Results of empirical studies providing support to the basic elements and the overall effectiveness of the DA are also presented. Implications of the findings are discussed and suggestions for further research, particularly in exploring the conditions under which the DA could have a long lasting effect on teacher effectiveness, are finally drawn.

Introduction

Teacher training and professional development are considered essential mechanisms for deepening teachers' content knowledge and developing their teaching practices in order to teach to high standards (Darling-Hammond & McLaughlin, 1995; Cohen & Hill, 2001). Over the last years, the demand for improving the quality of teaching and learning and the demand for increasing accountability have put issues related with effective professional development high on the agenda of educators, researchers and policy makers. The underlying rationale is that high quality teacher professional development could facilitate improvement of teaching practices, which could in turn translate into higher levels of student achievement (Borko et al., 2010; Desimone, 2009).

Despite the recognition of the importance of teacher professional development, most training opportunities remain fragmented, poorly aligned with curricula, and inadequate to meet teachers' needs and priorities for improvement (Borko, 2004; Cohen & Hill, 2001). In this context, each year, schools, districts, and educational systems spend a considerable amount of money and resources on in-service seminars and other forms of professional development, which are intellectually superficial and don't take into account the knowledge base of effective teaching and how teachers could better learn and implement such practices (Kyriakides et al, 2009; Ball & Cohen, 1999). This is exactly

why, there is now more than ever the need to support and guide teachers to effectively respond to the growing demands of increasing accountability and of raising student learning standards by developing effective professional development programs that could promote changes in classroom practices (Spillane, 1999; Ball & Cohen, 1999).

The Dynamic Approach (DA), proposed in this paper, aims to promote improvements in teacher pedagogical knowledge and teaching skills. This is important to clarify since different professional development programs may have various aims and objectives related with teacher knowledge, perceptions and practices (see Shulman, 1987 for a review). Pedagogical knowledge goes beyond knowledge of subject matter *per se* to the dimension of subject matter knowledge for effective teaching. Thus, it relates to teacher behaviour in the classroom that could maximise student learning gains. The question that may arise at this point, however, relates to the content of teacher professional development programs, i.e., which skills should be targeted, who is to decide and why? Despite the large body of literature on professional development, surprisingly little attention has been paid to the actual content of the professional development activities (Garet et al., 2001). This content can be derived from a variety of sources, such as the various task analyses of teaching, attempts to specify the attributes of the teacher as professional or even competences specified by external agencies. Nevertheless, this paper supports that we need to be in a position to justify this selection on the basis of research findings. From this perspective, it is argued that we need to utilise and reflect on the knowledge base of the Educational Effectiveness Research (EER) describing teaching practices, strategies and actions that were found to have a positive impact on student outcomes. This is important as identifying specific practices fundamental to supporting student learning is at the heart of building an effective system for the professional training and development of teachers (Ball & Forzani, 2011).

Merging findings from research on teacher professional development with research on teacher effectiveness

Educational Effectiveness Research addresses the question of what works in education and why and attempts to identify factors situated at different levels that are associated with student achievement (Scheerens & Bosker, 1997). During the last thirty-five years specific types of teacher behaviour in the classroom were found to be associated with student achievement (e.g., Muijs & Reynolds, 2001). Although one would have expected

strong associations between research on teacher professional development and on teacher effectiveness, research in the two fields has been conducted apart from and without much reference to one another. This mutual isolation is particularly unfortunate for one attempting to draw implications for teacher education and professional development from EER and visa versus. A similar argument related with merging findings of research on teacher effectiveness and teacher professional development has already been implied but was not developed further either for research or for policy purposes (see Gage, 1978; Katz and Raths, 1984). Three decades after these publications, very similar conclusions were drawn by the AERA panel on research in teacher education (Cochran-Smith & Zeichner, 2005). Few investigators of training methods rationalize their selection of teaching skills in terms of research on teaching effectiveness and very few evaluate the impact of the teaching skills they develop on such dependent variables as student learning. At the same time, researchers of teacher effectiveness spend little time speculating about the methods that might be used to develop teaching skills that were found to be associated with student outcomes.

The development of the DA is based on the argument that research on teacher training and development should increasingly take into account the results of research on teacher effectiveness, addressing skills that are found to contribute to student learning. By establishing links between EER and research on teacher professional development, both fields could have mutual benefits. Particularly, research on teacher professional development could expand its research agenda by taking into consideration the impact of effective programs on student outcomes and at the same time EER could identify the extent to which its validated theoretical models can be used for improvement purposes. In this way, stronger links between research, policy and improvement of teaching practice could be established (Antoniou & Kyriakides, 2011). From this perspective, the dynamic model of EER (Creemers & Kyriakides, 2008), which is considered to be the latest development in the field (Sammons, 2009), could contribute in establishing a theory-driven and evidence-based approach to teacher professional development.

The dynamic model of EER

The dynamic model is multilevel in nature and refers to factors, associated with student outcomes, operating at four different levels: student, classroom, school and system. The teaching and learning situation is emphasised and the roles of the two main actors (i.e.,

teacher and student) are analysed (Creemers & Kyriakides, 2008). Particularly, at the classroom /teacher level the dynamic model refers to eight factors which describe teachers' instructional role: *orientation, structuring, questioning, teaching-modelling, applications, management of time, teacher role in making classroom a learning environment*, and *classroom assessment.* These eight factors do not refer only to one theory of teaching and learning such as the direct teaching or constructivism (Joyce, Weil, & Calhoun, 2000), but an integrated approach in defining effective teaching and student learning is adopted.

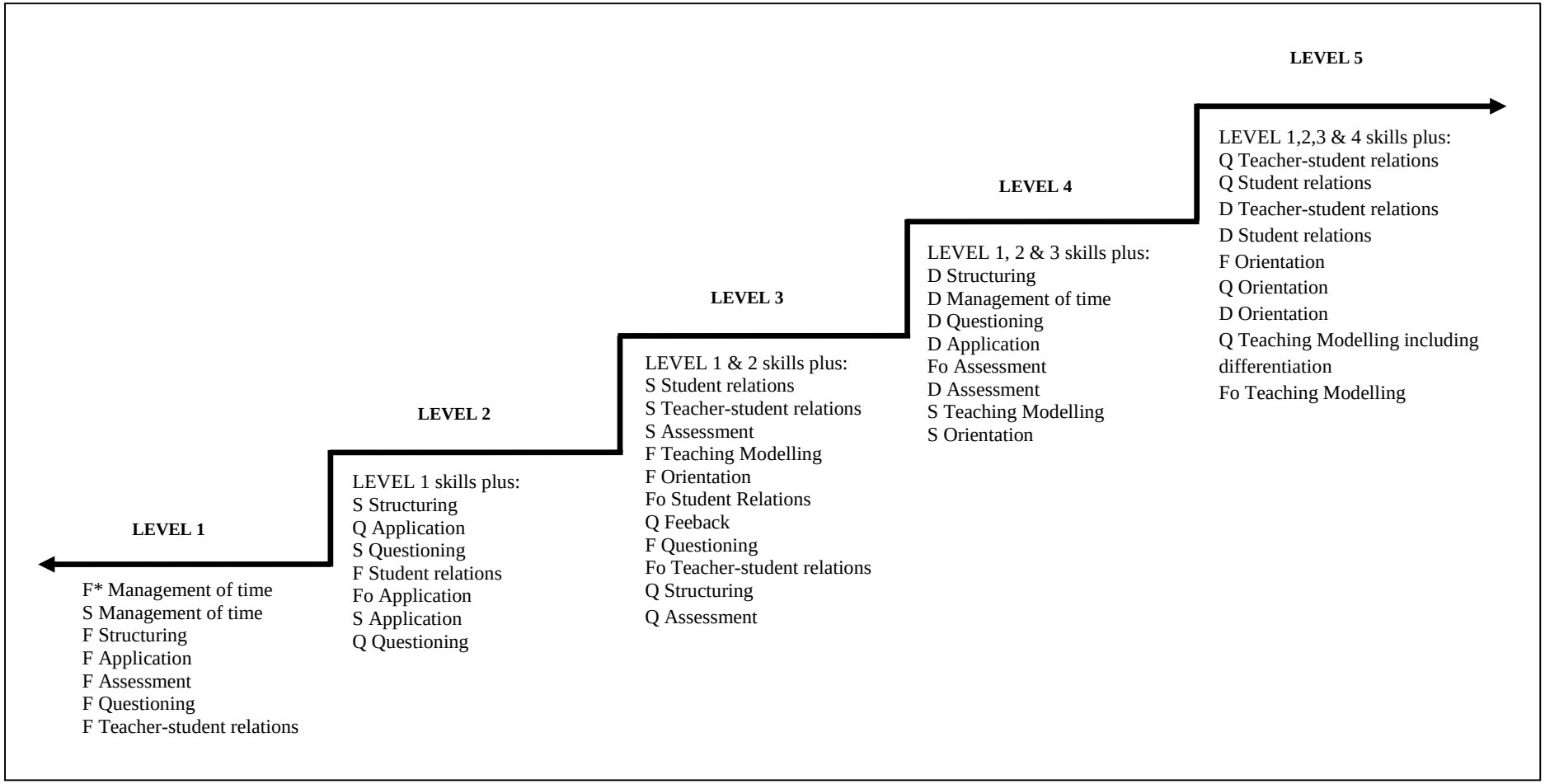
In addition, an essential difference between this model and those developed in the 1990s is that a specific multidimensional framework is used to measure the functioning of factors. Thus, each factor can be defined and measured by using five dimensions: *frequency, focus, stage, quality, and differentiation*. Specifically, frequency is a quantitative way to measure the functioning of each factor, whereas the other four dimensions examine qualitative characteristics of the functioning of each factor. The dimensions are not only important from a measurement perspective but also from a theoretical point of view. Actions of teachers associated with each factor can be understood from different perspectives and not only by giving emphasis to the number of cases or to the time duration the actions occur in teaching (i.e., frequency dimension). In addition, the use of these dimensions may help us develop strategies for improving teaching since the feedback provided to teachers could refer not only to quantitative but also to qualitative characteristics of their teaching practices.

Levels of teaching skills based on the Dynamic Model

The dynamic model is also based on the assumption that teacher factors are inter-related and the importance of grouping of factors has been demonstrated (Creemers & Kyriakides, 2008). Specifically, a longitudinal study revealed that the eight teacher factors and their measuring dimensions can be grouped into five levels, situated in a developmental order. These levels were found to be associated with student outcomes, thus, teachers who demonstrate competencies in relation to higher levels were found to be more effective than those situated at the lower levels. This association was found for achievement in different subjects and for both cognitive and affective outcomes (see Kyriakides et al., 2009).

The above finding is in line with the theories related with the stage models of professional development. Over the past three decades, cognitive psychology has produced a range of models of how teachers and other professionals develop expert skill (e.g., Berliner, 1994; Dreyfus & Dreyfus, 1986; Sternberg et al., 2000). Although these models vary with respect to both the number of stages that must be passed through and the nature of each stage, all have fixed sequences of stages representing successively higher level of knowledge and skills acquisition. For instance, Dreyfus and Dreyfus (1986) argued that acquisition in each new area typically proceeds through five skill stages: novice, advanced beginner, competent, proficient and expert. What seems to be the principle advancement of the five levels proposed by Kyriakides et al., (2009), compared with the previous stage models, is that the content of each level is now specifically determined (in terms of specific teaching skills), whereas previous stage models suffered from vagueness on what could actually constitute each developmental level (Dall'Alba & Sandberg, 2006). The teacher factors of the dynamic model included in each level are presented in Figure 1.

Figure 1. Developmental levels of teaching skills based on the Dynamic Model of EER



*The first letter describing each skill refers to the measuring dimension as follows: F: Frequency, S: Stage, Fo: Focus, Q: Quality, D: Differentiation

As we can observe from Figure 1, the five levels are described in a distinctive way. The first three levels are mainly related with the direct and active teaching approach by moving from the basic requirements concerning quantitative characteristics of teaching routines (e.g., management of time, providing structuring and application tasks) to the more advanced requirements concerning the appropriate use of these skills as these are measured by the qualitative characteristics of these factors (e.g., asking process and product questions, providing appropriate feedback). These skills gradually also move from the use of teacher-centred approaches to the active involvement of students in teaching and learning. The last two levels are more demanding since teachers are expected to differentiate their instruction (level 4) and demonstrate their ability to use the new teaching approach by engaging students to orientation and modelling tasks (level 5). Based on the above findings, the DA to teacher professional development has been developed aiming to facilitate the utilisation of the knowledge base of EER for improvement purposes. The main assumptions, features and implementation phases of DA are presented in the following sections.

The Dynamic Approach: Assumptions and Main Features

The first essential characteristic of the DA has to do with the fact that teacher factors concerned with teacher behaviour in the classroom were found to be related to each other, as discussed in the previous section. The grouping of factors highlights the need for establishing an integrated approach to teacher professional development, which could be situated between the competency-based approach (Brooks, 2002) and the holistic approach (Feiman-Nemser, 1990). Therefore, the DA is based on the assumption that improvement of teacher effectiveness can be focused neither on the acquisition of isolated skills/competencies (Gilberts & Lignugaris-Kraft, 1997) nor in reflection across the whole process of teaching in order to help teachers get "greater fulfilment as a practitioner of the art" (Clarke & Hollingsworth, 2002, p. 948) without considering the professional needs and developmental priorities of the teachers.

Second, the DA takes into account the importance of identifying specific needs and priorities for improvement of each teacher / group of teachers. This implies that, unlike most professional development approaches with a "one size fits all" orientation, the content of the training program should vary accordingly, since teachers with the same profile (i.e., teaching experience, initial training qualifications) may have different

priorities for improvement. In order to identify these priorities, multiple evaluation data related with teacher behaviour in the classroom should be collected and factors that need to be addressed and further developed should be identified.

Thirdly, it is acknowledged that teachers should be actively involved in their professional development in order to better understand how and why the factors / teaching skills addressed have an impact on student learning. For example, in training courses on improving factors concerned with classroom management, teachers need to discuss and reflect in order to gain a better understanding of how the factors addressed are related with the effective use of teaching time which is always limited. This implies that we should use the knowledge-base of EER in order to design professional development programs which aim not only to help teachers understand the importance of teacher factors, but also to develop their skills associated with these factors and implement those skills in their classrooms. In this context, the approach promotes the establishment of strategies for teacher professional development which give emphasis on the evidence stemming from theory and research. Thus, the value of a theory-driven approach to teacher training and professional development is stressed. Taken together with the need to collect multiple evaluation data about the skills of teachers to identify their improvement priorities mentioned above, it is argued that a theory-driven and evidence-based approach to teacher training and professional development should be established.

Fourth, teachers should become aware of both the empirical support available related to the factors involved in their developmental program and the way these factors operate within a conceptual framework (Sammons, 2009). Through this approach, teachers are offered the opportunity to utilise in a flexible manner the existing knowledge-base on effective teaching, adapt it to their specific needs, and develop their own strategies and action plans for improvement. Thus, the DA is neither based on improvement prescriptions or predetermined requirements for teachers to follow in order to improve their skills, nor on relying solely on teachers to identify by themselves what can be done and how in order to improve the quality of their teaching. The DA provides teachers the opportunity to identify their improvement needs and make use of the available knowledge-base in order to develop their action plans and critically reflect on their efforts in order to improve their teaching skills.

Fifth, building upon the previous point, the DA supports that the Advisory and Research Team (ARTeam), responsible for the coordination and the general provision of

the developmental program, has an important role in facilitating and supporting teachers in their efforts to develop and implement their action plans in their classrooms. Although each teacher is treated as a professional responsible for designing his/her own action plan and implementing his/her own improvement strategies, teachers are not left alone to design and implement their strategies and actions, but are encouraged to make use of the expertise and knowledge of the ARTeam and any other available resource within and/or outside the school. In such an integrated approach, teachers are the ones to take decisions relating to the improvement actions and tasks to be designed and implemented. By doing so, not only is ownership of the improvement effort established, but the teachers` experiences and the context of the school and classroom are also taken into account (Muijs, 2008). At the same time, the ARTeam has an important role to play in designing teachers' improvement strategies. The ARTeam is expected to share its expertise and knowledge with practitioners and help them develop strategies and action plans that are in line with the relevant knowledge base of effective teaching. To foster such discussions, the ARTeam must help teachers to establish trust, develop communication norms that enable critical DAlogue, and maintain a balance between respecting individual community members and critically analysing issues in their teaching.

Sixth, monitoring the implementation of teacher action plans in classroom settings is an essential part of the DA. During this procedure, teachers are expected to continuously develop and improve their action plans based on the information collected through formative evaluation. Critical reflection on the implementation of the action plans is also an important aspect of formative evaluation (Admiraal & Wubbels, 2005). It is important to stress that critical reflection and collaboration with peers are essential elements in all aspects of learning and throughout the improvement process. Thus, the DA seeks to initiate changes in educational practices, by encouraging teachers to systematically reflect on, and working with other teachers throughout the whole curriculum, in order to improve the effectiveness of existing practices and assisting on the development of new, based on the grouping of factors included in the dynamic model of EER and their particular priorities for improvement. For example, teachers could be encouraged to keep their own reflective DAries in order to identify ways to improve their action plans. At the same time, the ARTeam should help teachers collect additional data from other sources and test the internal validity of their evaluation mechanism by comparing data collected from different sources.

Finally, the DA refers to the importance of conducting summative evaluation in order to identify the impact of the developmental program on the teaching skills of the participating teachers and on the learning outcomes of their students. Despite the number of studies on teacher professional development, the majority of these do not measure the impact of different approaches and programmes on student learning outcomes (Cochran-Smith & Zeichner, 2005; Borko, 2004). Measuring the short- and the long- term impact of the proposed approach is important since it could help us to investigate the added-value of using the DA. The results of summative evaluation are also important for taking decisions on whether some groups of teachers have developed their practices successfully and, thus, need to design new action plans in order to address new priorities for improvement. This implies that teachers should be continuously involved in improvement efforts in order to move from their current level to more demanding levels of effective teaching.

Main Implementation Steps

In the context of the framework described above, this section describes the basic implementation steps and procedures of the DA. As demonstrated on Figure 2, the DA is based on a sequence of five basic implementation steps which are elaborated below.

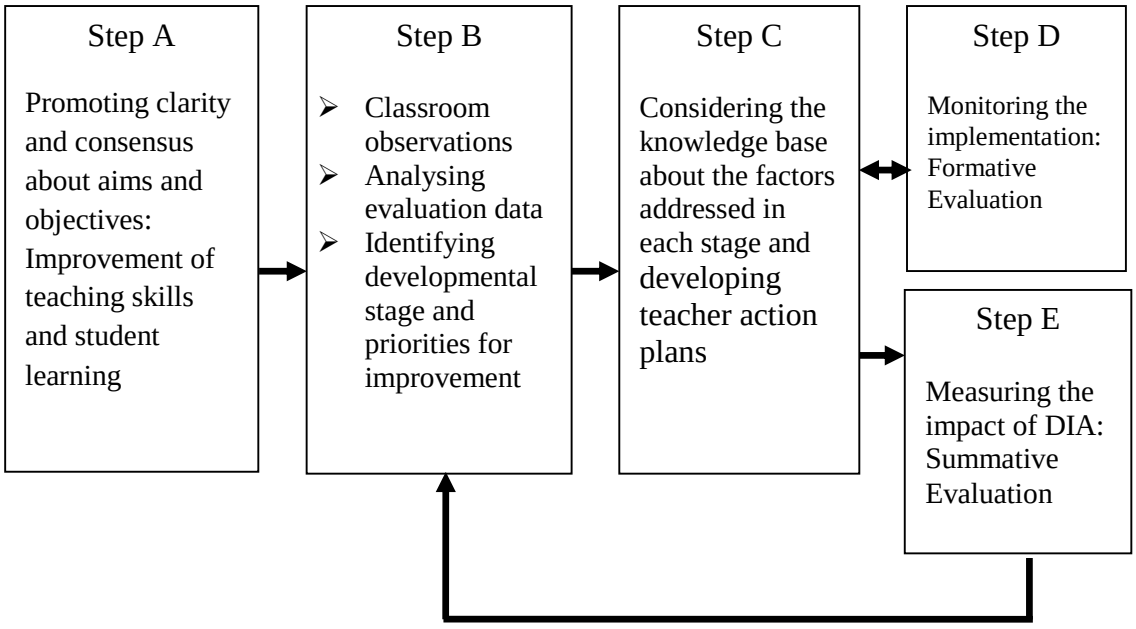


Figure 2. The major implementation steps of the Dynamic Integrated Approach

A) Establishing clarity and consensus about the general aims and the objectives of the teacher professional development program.

At this very first step of the DA, it is emphasised that the ultimate aim of the improvement effort is to enhance student learning. To achieve this, professional development is expected to help teachers improve their teaching skills and classroom practices, thus, the importance of the classroom level, as the central point for improvement is acknowledged (Reynolds et al., 1993). As Scott and Dinham (2002, p. 112) argue, '…quality of teaching becoming a major focus in the educational systems of many countries responding to teacher demands for professional development that matters in their everyday tasks and activities.' This step is based on the assumption that it is important to start with a clear understanding of the destination and how improvement of quality in education will be achieved. This could also be considered as "a purposeful task analysis" (Wiggins & McTighe, 1998: 8), suggesting a planning sequence. Commitment to the implementation of the training program by both the participating teachers and the research and advisory team should be established. The importance of developing a theory driven, but at the same time, evidence based program to address the specific needs and priorities for improvement of the participating teachers is elaborated. Thus, at the next step data should be collected in relation to teaching skills in order to identify the professional needs and the priorities for improvement for each teacher.

B) Identify needs and priorities for improvement through empirical investigation.

The use of a validated framework, such as the dynamic model of EER, on the basis of which the content of the training programme is to be selected and formulated, cannot in itself ensure that the programme will be effective for all participating teachers. The DA supports that not only should a theory-driven approach be followed to improve quality of teaching, but emphasis should also be placed on collecting data in order to identify the teaching needs and priorities for improvement for different groups of participants, thereby facilitating the design of relevant improvement efforts with differentiated content and focus. This is important, since teachers seem to consider new initiatives on their individual merits, particularly in relation to how they will benefit classroom teaching (Corkindale & Trorey, 2002). Teachers have turned away from various professional development approaches, which are not seen to have ready relevance

to and application in, the classroom and are not geared to teachers' needs (Dinham et al., 2000).

From this perspective, the second step of the proposed approach is based on the assumption that in any effort to train teachers, an initial evaluation of their teaching skills should be conducted to investigate the extent to which they possess certain teaching skills whilst identifying their needs and priorities for improvement. The teaching skills of the participants can be evaluated by the ARTeam, by utilising the instruments applied in studies testing the validity of the dynamic model at the teacher level (see Kyriakides & Creemers, 2008; Antoniou & Kyriakides, 2011). Based on the evaluation results, teachers are allocated into different groups based on their professional needs (i.e., level of teaching skills). The results of the initial evaluation provide suggestions for the content of training for different groups of teachers. This is important, since the content and development of educational material for the training programmes should correspond to the professional needs and *proximal development* of each group of teachers. According to Berliner (1994), it would be beneficial to assist those willing to progress by providing training and feedback appropriate to their level of development. For example, teachers must master simple but necessary routines such as teaching skills related to the "direct teaching approach" in order to move to higher levels involving the use of "new teaching approaches" and differentiation. Furthermore, the DA supports that the effort to identify teachers' needs and priorities for improvement should be guided by the knowledge base of EER as it is described in the dynamic model. This is an important issue that needs to be taken into account in conducting the initial evaluation especially since the dynamic model refers to teaching skills found to be related to student achievement.

C) Provide guidelines for improvement and reflection opportunities.

Having identified teachers' needs and priorities for improvement, teachers in each group should then engage in developmental activities towards improving their teaching skills. Thus, the third step of this approach relates to the provision of appropriate material and guidelines to teachers for designing their action plans for improvement. The ARTeam also provides the teachers of each group with supporting literature, research findings and activities related to the teaching skills in their developmental level. For example, the teachers in the first level of teaching skills should focus and receive material and guidance on the distribution of teaching time and ways of dealing with time management

effectively. Case studies could be administered to the teachers in this group to discuss the importance of the quantity of teaching time as an effectiveness factor associated with student learning. In addition, material from the literature could be provided regarding the management of the classroom as an efficient learning environment, in order to maximise engagement rates (Creemers & Reezigt, 1996). Through discussion, it is expected that teachers will realise that learning takes place in restricted time limits in which many important activities should take place. Extra-curricular administrative activities such as announcements, dealing with discipline problems and commenting on irrelevant issues could further reduce the time available for learning. Thus, the teachers are expected to understand that actions should be taken in order to improve their skills in management of time and reflect on how to allocate time in each learning activity sufficiently. In addition, examples for teaching specific subjects from the school curriculum could be discussed with teachers. In this way, teachers are encouraged both to reflect on these aspects of their teaching practice and provide their own examples. Moreover, teachers are provided with opportunities for collective reflection and critical learning, features closely related to active learning (Elliot & Calderhead, 1995). Besides individual reflection, collective reflection can be a fruitful tool for enriching and widening a person's thinking, especially since teachers' work conditions are often claimed to support individualism and privacy. The underlying assumption is that the group-based management structure could utilise the accumulated experience and knowledge of the team to facilitate improvement. As Desimone (2009) argues, 'Such arrangements set up potential interaction and discourse, which can be a powerful form of teacher learning' (p. 184).

Subsequently, with the support of the ARTeam teachers should develop their own action plans for improvement. It is also emphasized that no single strategy will always work in every school, for every teacher, all of the time. Local customisation is necessary for the success of programmes of teacher learning or professional development (Fishman, Marx, Best, & Tal, 2003). The basic elements of a general plan of action should also be discussed. Teachers should use various techniques and methods for gathering evidence on the effectiveness of their action plans. For this reason, teachers are encouraged to keep a reflective DAry. This DAry could contain personal accounts of observations, feelings, reactions, interpretations, reflections, hunches, hypotheses and explanations. Teachers could also ask their pupils to keep DAries, which could enable the teacher to compare their experiences of the situation with those of the pupils'. Moreover, other teachers at the

school could observe their teaching (e.g., acting as "critical friends"), following a peer-coaching approach (Joyce & Showers, 1995).

D) Establish a formative evaluation mechanism.

The next implementation step of the DA refers to the establishment of formative evaluation procedures. Formative evaluation is the method of ongoing and concurrent evaluation which aims to improve the programme (Popham, 2006). The formative evaluation procedures should be carried out on a regular basis (e.g., in monthly sessions) to provide information and feedback for improving: a) the quality of teachers' learning, b) the extent to which they implement the teaching skills in their classrooms and finally, c) the quality of the programme itself. Such formative evaluation procedures should involve: the identification of the learning goals, intentions or outcomes, and criteria for achieving them; the provision of effective, timely feedback to enable teachers advance their learning; the active involvement of teachers in their own learning, and finally teachers responding to identified learning needs and priorities by improving their teaching skills. Furthermore, the monthly sessions could provide teachers with the opportunity to revise and develop further their action plans on a systematic basis, based on their own and others' experiences and also based on the literature on effectiveness factors which correspond to their level. This can be achieved with the assistance and guidance of the ARTeam. For example, through formative evaluation in each monthly session, teachers could be provided the opportunity to: a) report teaching practices and comment on them, b) identify effective and non-effective teaching practices, c) understand the significance of the teacher factors which correspond to their competency level, and d) understand how these factors could be linked with effective teaching and learning. At the same time, the teachers at each level should receive systematic feedback and suggestions from the ARTeam. During the program, members of the ARTeam should visit teachers at their schools to discuss emerging issues related to the implementation of their action plans and provide support and feedback.

E) Establish a summative evaluation system.

The final step of the proposed DA is concerned with establishing a summative evaluation system. A value-added approach should be adopted (Antoniou & Kyriakides, 2011). This implies that at the beginning and at the end of the school year teaching skills

and students' outcomes should be measured, so as to identify the net effect of the professional development program. Specifically, the teaching skills of the participating teachers should again be evaluated by focusing on the eight factors of the dynamic model concerning teacher behaviour in the classroom. Data on student achievement should also be collected, in order to measure the effectiveness of the DA in terms of student achievement gains. The emphasis of the summative evaluation should not be on comparing teachers with each other, but on identifying the overall impact of the programme on the development of teachers' skills and its indirect effect on student learning. The results of such an evaluation system could assist in measuring the effectiveness of the DA and allow subsequent decisions to be made regarding the continuity of the programme.

Conclusions and suggestions for further research

This paper advocates the use of the DA, an evidence-based and theory-driven approach towards teacher training and professional development and particularly towards the improvement of teaching skills and student outcomes. The proposed approach integrates research findings from teacher effectiveness, such as the grouping of teaching skills included in the Dynamic Model of EER, with research findings from teacher training and professional development, such as the utilisation of critical reflection, development of action plans, mentoring by the Research and Advisory Team and peer coaching. The findings of the studies utilising the DA, briefly presented in this paper, reveal the added value of using this approach to improve teaching skills and student achievement.

The DA and the results of the studies conducted so far to investigate the validity of this approach (e.g., Antoniou & Kyriakides, 2011; Antoniou & Kyriakides, in press, Christoforidou, Kyriakides & Antoniou, 2012), have important implications for organising teacher professional development courses. Such implications are related with the need to develop and provide developmental programmes which address the participants' professional needs and immeDAte priorities for improvement. This also implies that we should move away from professional development approaches with a "one size fits all" orientation and acknowledge in practice the need to differentiate the content of the various courses according to the participants' needs. Like Combs et al., (1974) argue, "in the first place, it is a fallacy to assume that the methods of the experts either can or should be taught directly to beginners" (p.4). Moreover, according to

Berliner (1994), we probably need to think through the scope and sequence of teacher education experiences in the same way and with the same care that we develop scope and sequence guides for students from kindergarten to twelfth grade. Decision making, priority setting, and other aspects demonstrating personal control over the environment are characteristic of the developmental stage of competence teacher, rather than that of a novice.

This of course yields an additional implication, related with the need to measure and evaluate the teaching skills of the participating teachers. Based on the evaluation findings, teachers should be classified into groups according to the level at which they were found to belong. Thus, it is important to acknowledge that more resources may be needed in order to organise training courses based on the DA, in comparison with other teacher professional development approaches. Such resources are related with the extra amount of time that tutors would need in order to carry out classroom observations and collect evaluation data on the teaching skills of the participating teachers. However, this is a crucial element of the DA, since unless the teaching skills of the participating teachers are measured, improvement priorities cannot be identified and action plans addressing those needs cannot be developed. The studies, briefly reported in this paper, demonstrate that although the effective implementation of the DA needs more resources, the approach could be considered as cost-effective since a significant impact on the quality of teaching and student learning has been identified.

Moreover, the results of the studies employing the DA, provide support to the argument that it is time to stop assuming that all teachers are in possession of effective teaching skills that develop naturally and without the need for training and reflection addressing specific needs. As with all skill learning, regardless of whether it involves performance skills or cognitive skills, there is a need for programs that train for the desired skills (Cornford, 1996). This attempt is supported by Desimone et al. (2002), arguing that focusing on specific teaching practices in professional development, increase teachers' use of those practices in the classroom and thus students' learning. That is not to deny in any way that reflective thinking and critical analysis are important and, for this reason these two elements have been utilized in the development of the DA.

Particularly, according to the DA, reflection for understanding and critical thinking on teaching skills and classroom practices, are important elements in all aspects of learning and performance. Through reflection teachers participate consciously and

creatively in their own growth and development (Zeichner & Liston, 1996). Reflection enables practitioners to analyse, discuss, evaluate and change their own practice, adopting an analytical approach. From this perspective, the DA supports that at the same time there must be appropriate content or a coherent body of knowledge, supported by empirical data and validated theoretical frameworks, to guide the reflection process and facilitate teacher improvement.

References

Admiraal, W., & Wubbels, T. (2005). Multiple voices, multiple realities, what truth? Student teachers' learning to reflect in different paradigms. *Teachers and Teaching: theory and practice, 11*(3), 315–329.

Antoniou, P., & Kyriakides, L. (in press). A Dynamic Approach to Teacher Professional Development: Impact and Sustainability of the Effects on Improving Teacher Behavior and Student Outcomes. *Teaching and Teacher Education*

Antoniou, P., & Kyriakides, L. (2011). The impact of a dynamic approach to professional development on teacher instruction and student learning: results from an experimental study. *School Effectiveness and School Improvement, 22*(3), 291-311.

Avalos, B. (2011). Teacher professional development in teaching and teacher education over ten years. *Teaching and Teacher Education, 27*(1), 10-20.

Ball, D.L., & Cohen, D.K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In G. Sykes & L. Darling-Hammond (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3-32). San Francisco: Jossey Bass.

Ball, D.L., & Forzani, F. M. (2011). Building a common core for learning to teach, and connecting professional learning to practice. *American Educator, 35*(2), 17-21, 38-39.

Berliner, D. (1994). Expertise: The wonder of exemplary performances. In J. Mangieri & C. Block (Eds.), *Creating powerful thinking in teachers and students: Diverse perspectives,* (pp. 161–186). Fort Worth, TX: Harcourt Brace College.

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

Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development: Processes and content. In P. Peterson, E. Baker, & B.

McGaw (Eds.), *International encyclopeDA of education*, Vol. 7 (pp. 548–556). Oxford: Elsevier.

Brooks, R. (2002). The individual and the institutional: Balancing professional development needs within further education. In G. Trorey & C. Cullingford (Eds), *Professional Development and Institutional needs* (pp. 35-50). Hampshire, England: Ashgate Publishing.

Christoforidou, M., Kyriakides, L. & Antoniou, P. (2012). The Impact of the Dynamic Approach to Teacher Professional Development upon Teachers' Skills in Assessment. Paper presented at the American Educational Research Association (AERA) Annual Meeting, Vancouver, British Columbia, Canada.

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*(8), 947–967.

Cochran-Smith, M., & Zeichner, M.K. (2005). *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education.* AERA Panel on Research and Teacher Education. Routledge.

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

Combs, A.W., Blume, R.A., Newman, A.J., & Wass, H.L. (1974). *The professional education of teachers: A humanistic approach to teacher preparation.* Boston: Allyn & Bacon.

Corkindale, J., & Trorey, G. (2002). Career Dynamics in Further and Higher Education. In G. Trorey, & C. Cullingford (Eds), *Professional Development and Institutional needs* (pp.79-101). Hampshire, England: Ashgate Publishing.

Cornford, I.R. (1996). Experienced teachers' views of competency-based training in NSW TAFE. In Learning & work: The challenges: Con, fence papers, Vol. 4 (pp. 105-115). Brisbane: Griffith University, Centre for Learning and Work Research.

Creemers, B.P.M., Kyriakides, L., & Antoniou, P. (2012).*Teacher Professional Development for Improving Quality of Teaching*: New York, USA.

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

Creemers, B.P.M., & Reezigt, G.J. (1996). School level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement, 7*(3), 197–228.

Dall'Alba, G. & Sandberg, J. *(*2006*).* Unveiling professional development: a critical review of stage models. *Review of Educational Research, 76,* 383–412.

Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. Phi Delta Kappan, 76(8), 597 - 604.

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., Porter, A., Garet, M., Yoon, K. S., & Birman, B. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis, 24*(81), 81–112.

Dinham, S., Brennan, K., Collier, J., Deece, A., & Mulford, D. (2000). The Secondary Head of Department: Key Link in the Quality Teaching and Learning Chain. *Quality Teaching Series, 2,* 1-35.

Dreyfus, H.L., & Dreyfus, S.E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer.* New York: Free Press.

Elliot, B., & Calderhead, J. (1995). Mentoring for teacher development: possibilities and caveats. In T. Kerry & A. Shelton-Mayes (Eds), *Issues in Mentoring* (pp.35-55). London: Routledge in association with the Open University.

Feiman-Nemser, S. (1990). Teacher Preparation: structural and conceptual alternatives. In W. Houston (Ed.), *Handbook of Research on Teacher Education* (pp. 212-233). New York: Macmillan.

Fishman, B., Marx, R., Best, S., & Tal, R. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education, 19*(6), 643-658.

Gage, N.L. (1978). *The scientific basis for the art of teaching.* New York: Teachers College Press.

Garet, M.S., Porter, A.C., Desimone, L., Birman, P.F., & Yoon, K.S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915-945.

Gilberts, G.H., & Lignugaris-Kraft, B. (1997). Classroom management and instruction competencies for preparing elementary and special education teachers. *Teaching and Teacher Education, 13*(6), 597–610.

Janosz, M., Archambault, I., & Kyriakides, L. (2011). The cross-cultural validity of the dynamic model of educational effectiveness: A CanaDAn study. *Paper presented at the 24th International Congress for School Effectiveness and Improvement (ICSEI) 2011*. Limassol, Cyprus, January 2011.

Joyce, B., & Showers, B. (1995). *Student achievement through staff development* (2nd ed.). White Plains, NY: Longman.

Joyce, B., Weil, M., & Calhoun, E. (2000). *Models of teaching*. Boston: Allyn & Bacon.

Katz, G.L., & Raths, D.J. (1984). *Advances in Teacher Education* (Vol. 1). New Jersey: Ablex Publishing Corporation.

King, P.M., & Kitchener, K.S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults.* San Francisco:Jossey-Bass.

Kyriakides, L., Creemers, B.P.M., & Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education, 25*(1), 12–23.

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.

Muijs, D. (2008). Widening opportunities? A case study of school-to-school collaboration in a rural district. *Improving Schools, 11*(1), 61-73.

Muijs, D., & Reynolds, D. (2001). *Effective Teaching: evidence and practice*. London: Sage.

Popham, W.J. (2006). Phony formative assessments: Buyer beware! *Educational Leadership, 64*(3), 86–87.

Reynolds, D., Hopkins, D., & Stoll, L. (1993). Linking school effectiveness knowledge and school improvement practice: Towards a synergy. *School Effectiveness and School Improvement*, *4*, 37–58.

Sammons, P. (2009). The dynamics of educational effectiveness: a contribution to policy, practice and theory in contemporary schools. *School Effectiveness and School Improvement, 20*(1), 123–129.

Scheerens, J., & Bosker, R.J. (1997). *The foundations of educational effectiveness.* Oxford: Pergamon.

Scott, C., & Dinham, S. (2002). The beatings will continue until quality improves: Carrots and sticks in the search for educational improvement. *Teacher Development, 6*(1), 15-31.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis research. *Review of Educational Research, 77,* 454-499.

Shulman, L. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review, 57*(1), 1–22.

Spillane, J. P. (1999). External reform initiatives and teachers' efforts to reconstruct practice: The meDAting role of teachers' zones of enactment. *Journal of Curriculum Studies, 31,* 143–175.

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., Snook, S. A., & Grigorenko, E. L. (2000). *Practical intelligence in everyday life*. New York: Cambridge University Press.

Wiggins, G., and J. McTighe. 1998. *Understanding by design*. Alexandria, VA: ASCD.

Zeichner, K., & Liston. D. (1996). *Reflective teaching: An introduction*. Mahwah, NJ: Lawrence Erlbaum Associates.

Promoting Formative Assessment in Mathematics: The Impact of a Teacher Professional Development Program Based on the Dynamic Approach

Leonidas Kyriakides, Evi Charalambous, & Andria Dimosthenous

Department of Education, University of Cyprus, Nicosia, Cyprus

Email: kyriakid@ucy.ac.cy, charalambous.e.evi@ucy.ac.cy, & dimosthenous.antria@ucy.ac.cy

Abstract: This paper presents the main results of the FORMAS project which aimed to contribute in improving secondary teachers' assessment skills by supporting them to conduct assessment for formative reasons and through this to promote their students' learning outcomes in Mathematics (cognitive and meta-cognitive). At the beginning of the school year 2019-2020, a sample of 206 secondary school teachers who teach Mathematics in lower secondary schools (i.e., Grades 7, 8 and 9) in four European countries (Belgium, Cyprus, Greece, and the Netherlands) was randomly allocated into the experimental and the control group. A questionnaire for measuring teachers' professional needs in assessment as well as a battery of tests for measuring students' learning outcomes in mathematics (cognitive and meta-cognitive) have been developed. A Teacher Professional Development (TPD) course on formative assessment based on the main assumptions of the Dynamic Approach (DA) has been offered to the teachers of the experimental group. In each participating country, teachers of the experimental group managed to improve their assessment skills, whereas no improvement was observed among the teachers of the control group. In addition, by using multilevel modeling techniques it was found that in each country, students of the experimental group made more progress in their cognitive and meta-cognitive learning outcomes in Mathematics than students of the control group. Implications for the development of a policy on formative assessment including TPD are briefly discussed.

Introduction

Over the past decades, classroom assessment has been a centrepiece of various educational improvement efforts, placing an emphasis on the role of formative assessment practice in supporting student learning (Black, 2016; Hattie, 2009; Hopfenbeck & Stobart, 2015). More specifically, a number of effectiveness studies have identified that teachers who use assessment for formative rather than summative purposes were found to be more effective in promoting student learning outcomes (Creemers & Kyriakides, 2008; Hattie & Temperley, 2007). However, even though that teachers appear to hold positive views towards assessment that aids learning, their everyday practice appears to be summative oriented (Earl & Katz, 2000; Kahn, 2000). This can partly be attributed to the fact that teachers do not receive sufficient training in classroom assessment both during their initial studies as well as during their in-service professional

development programs (DeLuca & Klinger, 2010). Moreover, it should be noted that Teacher Professional Development (TPD) programs intended to improve assessment practice are scarce and have so far resulted in mixed results regarding their impact on teacher skills and on student learning outcomes (e.g., Randel, Apthorp, Beesley, Clark, & Wang, 2016; Schneider & Meyer, 2012; Schneider & Randel, 2010).

In this context, the FORMAS project aims to contribute in improving professional standards of secondary teachers by supporting them to conduct assessment for formative reasons and become more effective in terms of promoting student learning outcomes (cognitive and meta-cognitive). To achieve this purpose, the research team of this project has developed a comprehensive framework for measuring teachers' assessment skills and established valid instruments to measure teachers' professional needs as well as student learning outcomes in Mathematics (cognitive and meta-cognitive). Also, a Teacher Professional Development (TPD) course on assessment based on the main assumptions of the Dynamic Approach (DA) (Creemers & Kyriakides, 2012) has been designed and used in this project. This DA recognises the importance of involving teachers in critical reflection upon their practice and supporting them to develop and implement their own improvement strategies and action plans (for more information on the use of DA for TPD purposes see Creemers, Kyriakides, & Antoniou, 2013).

Consequently, in this paper, the impact of this TPD course on improving teachers' assessment skills and through that on promoting student learning outcomes in Mathematics (cognitive and meta-cognitive) is examined. More specifically, in the first part of this paper we refer to main methods of the study and in the second part the main findings are presented. Lastly, implications for developing a policy on TPD to promote formative assessment are discussed in the final part of the paper.

Methods

1. Ethical Aspects

All necessary authorisations and permissions to conduct the intervention study were settled by the participating countries' corresponding authorities, which varied according to the structure of the educational system of each country. More specifically, a consent form was signed by the parents of the students who participated in the study. In addition, all data were gathered anonymously (both from students and teachers). Anonymity was

also applied at school and teacher level, since neither the names of the participating teachers and schools nor their region were made known to the public. Consequently, all data were entered in the data bank by using specific student, teacher and school codes.

2. Participants

At the beginning of school year 2019-20, each country team (Belgium, Cyprus, Greece, and the Netherlands) invited 60 secondary school teachers who teach Mathematics in Grades 7, 8 and 9 to participate in the study. From all the invited teachers, 206 agreed to participate. The number of female teachers participating in the study was 151 (73.3%) and the number of male teachers was 55 (26.7%). As regards to the years of experience of the teacher sample, the mean was 15.57 (with a standard deviation of 8) and their year experience varies from 1 to 39. These teachers were randomly split into two groups: the experimental (n=102) and the control group (n=104). Randomization was done at the school level to avoid any spillover effect. Students of Grades 7, 8 and 9 of the teacher sample participated in the study. More specifically, all students of two classrooms per teacher were randomly selected. Our student sample was 5447 students coming from 294 classrooms. Teachers of the experimental group were invited to participate in a TPD course with a focus on student assessment. Teachers of the control group did not attend any TPD course. However, they were provided the opportunity to receive the course's material during the next school year.

Data on teacher skills and student achievement were collected not only at the beginning but also at the end of the TPD course. In this way, we were able to compare the impact of the TPD course on both improving the assessment skills of teachers and on student achievement gains (cognitive and meta-cognitive) in Mathematics. It is, however, important to note that due to the COVID_19 pandemic, a significant number of teachers and students were not able to participate in the intervention from March 2020 till the end of the school year. Although 206 teachers from the four countries participated in the initial measurement, only 166 teachers participated in both measurement occasions. As regards to the student sample, 5447 students participated in the initial measurement of cognitive learning outcomes in Mathematics, whereas 4012 students participated in both measurements. Moreover, 5544 students participated in the initial measurement of meta-cognitive learning outcomes in Mathematics, whereas only 4170 students participated in both measurements.

To search for selection bias, inferential analysis was conducted to test for any differences between teachers and students who participated only in the first measurement occasion and those who participated in both measurement occasions. Comparison was also made between the control and the experimental group (see more information in the Results section).

3. Design of the TPD – Steps of the Intervention

Recognizing the role of TPD in improving teachers' teaching and assessment practices (Borko, 2004; Kennedy, 2016), this project used the DA to TPD (Creemers et al., 2013) for the design and delivery of the TPD course. The DA was considered as an appropriate approach for improving teachers' assessment skills since previous studies provided support for the effectiveness of the DA on the development of teaching and assessment skills and student learning outcomes (e.g. Antoniou & Kyriakides, 2011, 2013; Creemers et al., 2013). The DA is based on the assumption that teacher improvement efforts should aim at the development of teaching skills, which have been empirically associated to positive student learning outcomes (Garet et al., 2016; Darling-Hammond et al., 2019). This way, professional development can have an impact not only on teachers' skills but on student learning outcomes as well. The DA supports that teacher training and professional development should focus on how to address specific groupings of teacher factors in relation to student learning rather than to an isolated teaching factor or teaching skill without considering the professional needs of each group of teachers. Teachers differ widely in their levels of professional competence, and these differences are systematically associated with differences in their professional practice (Kunter, Kleickmann, Klusmann, & Richter, 2013). Therefore, an initial evaluation of teachers' skills to identify priorities for improvement is necessary to adjust the content of the training to teachers' professional needs. Moreover, the DA encourages participants to engage into systematic and guided critical reflection on their teaching practices. Reflection and critical analysis are therefore considered as essential elements of the process of improvement but at the same time knowledge, skills, as well as the ability to act upon critical thinking are considered prerequisites of effective reflective practice (Kennedy, 2016; Boud, Keogh, & Walker, 2013; Darling-Hammond et al., 2019).

The DA refers to four main steps that need to be considered when designing a TPD course. These steps were used to design and guide the implementation of the TPD program under this project.

Step 1: Initial evaluation of teachers' assessment skills and student learning outcomes

The first step is concerned with the identification of the professional development needs of each teacher using a teacher questionnaire for measuring assessment skills. In this project, this initial measurement took place during an introductory session (i.e. session 1) and helped us identify priorities for improvement for the participating teachers based on their needs. Teacher data were collected during this first step and used as the initial evaluation of teachers' assessment skills. In addition, data on student achievement were collected using external written forms of assessment designed to assess cognitive and meta-cognitive knowledge and skills in mathematics (see next section). Specifically, three stages of teacher assessment behaviour were identified (see Data Collection section). Similar stages are expected to be identified when the questionnaire is administered in other contexts (see Kyriakides, Creemers, Panayiotou, & Charalambous, 2021).

Step 2: Offering Training Sessions to Teachers of the Experimental Group

The three stages identified confirm the initial assumption of our project that teachers have differentiated professional needs when it comes to student assessment. This implies that a common training to all participating teachers, as is usually the case, is not an appropriate solution. The TPD course offered under this project was organized in a way that could accommodate these differentiated needs, by providing differentiated training to each group of teachers based on their initial evaluation results. Thus, this second step is concerned with the support that the research and advisory team (i.e. teacher educators) of each participating country has provided to the teachers to help them establish their own action plans. Specifically, teacher educators provided teachers of each group with training material, opportunities for application of new knowledge and supporting literature related to the assessment skills of their group, and with clear instructions about the area on which each group should concentrate for improvement. As a result, each teacher will be able to develop his/her own action plan.

More specifically, two handbooks have been developed for the purposes of the TPD course: (a) a teacher trainer handbook and (b) a teacher handbook. The *teacher trainer handbook* aims to support teacher trainers in the delivery of the TPD program in student assessment and includes the theoretical background and main assumptions of the DA, the rationale of the training, the general structure of the training and the role of the teacher educators. It also presents a detailed guide outlining the training material and the necessary information for the implementation of each training session for all groups of teachers. The *teacher handbook* is addressed to teachers interested in improving their skills in student assessment and aims to support teachers to be engaged in a self-study process focused on improving teachers' skills in assessment and through that on promoting student learning outcomes.

Acknowledging that the duration of a TPD both in terms of span of time over which the TPD is spread and the number of hours spent in the TPD (Desimone, 2009) affects the impact that the program can have on teacher knowledge and skills, five 3-hour sections (four face to face and one online due to the COVID_19 pandemic) training sessions were offered to each one the three groups of the participating teachers over the period of the school year 2019-20 (September-May). This allowed teachers to use the time-lapse in-between sessions to implement actions for improvement, get feedback on their efforts and adjust their actions accordingly. It also enabled the evaluation of the program (i.e., pre- and post-measurements) in order to identify its impact on the development of teachers' skills and its effect on student learning. The TPD training was offered by members of each research team of each participating country with a strong background in educational assessment and previous experience in the development and implementation of TPDs in student assessment. Interventions implemented by researchers have been argued to yield higher effects as the researchers appear to be more motivated to test an intervention while at the same time knowing best how the intervention should be implemented (de Boer, Donker, & van der Werf, 2014). To avoid possible trainer effect, it was decided that trainers will rotate between the three groups.

The first session of the TPD course was common for all teachers and in this session the initial measurement of teacher assessment skills was also carried out. For sessions 2 to 5, teachers were grouped based on their professional needs as these are identified by the initial measurement of their assessment skills. Table 1 below shows the

content covered in sessions 2-5 for each group. A detailed description of each session can be found in the teacher trainer handbook (see http://www.ucy.ac.cy/formas/en/resources).

Table 1. The content of sessions 2-5 for each group

	Session 2	***Session 3***	***Session 4***	***Session 5***
GROUP A	Creating a culture that can foster formative assessment	Designing representative and valid assessments	Developing different types of assessment items to achieve quality in assessment	Assessing homework for formative purposes
GROUP B	Providing constructive feedback to students	Using different types of assessment techniques in an efficient and systematic way	Formulating assessment success criteria and involving students in the process of assessment	Using rubrics/checklists to record results from different assessment techniques
GROUP C	Recording results in ways that enable us to identify the needs of each student	Using assessment to assess individual/group work	Differentiation in assessment: facing the challenges	Differentiation in assessment: implications for using self-assessment and recording/reporting results

It is also important to mention that given the decision to focus the study to secondary school teachers that taught Mathematics during the school year 2019-20, the content of the TPD course (i.e., examples, application activities etc.) was adjusted to address the subject of Mathematics. However, this was not a training on how to teach Mathematics but on how to assess Mathematics. So, the essence was not so much the mathematical content but the assessment skills necessary to assess this content.

Step 3: Formative evaluation procedures during the TPD course

The third step of the DA comprises the establishment of formative evaluation procedures throughout the sessions. This means that teacher educators worked closely with participating teachers to help them identify their learning goals and choose actions that can aid their achievement. Most importantly, they provided constructive feedback during and through the sessions to support teachers' improvement efforts. More specifically, with the support of the research and advisory team, teachers of each group were asked to reflect on their experiences and identify effective or non-effective practices, share comments on the activities implemented and receive and provide constructive feedback. Furthermore, teachers were asked to complete application activities related to their focus area. The purpose of these application activities was to provide teachers with opportunities to practice the skills under focus as well as to encourage collaboration within the team. Teachers were then expected to collaborate to develop appropriate record templates for given assessment activities that allowed the use of data for formative purposes. Teachers were also encouraged to revise their action plans, based on their own and others' experiences and on the material provided; this done always under the support and guidance of the research team.

Step 4: Final evaluation of teachers' assessment skills and student outcomes

The fourth and final step of the TPD course aims to identify its impact on the development of teachers' assessment skills and its indirect effect on student learning. Therefore, summative evaluation was carried out by the research team of the project after the end of the TPD course. Teachers' assessment skills and student learning outcomes (cognitive and meta-cognitive) in Mathematics were measured by using the same procedures and instruments as in step 1. Further information about the initial and final data collection phases are provided in the data collection section.

4. Data Collection

To examine the impact of the TPD course based on the DA, data concerning teachers' assessment skills, as well as student performance in Mathematics (cognitive and meta-cognitive) were collected. The instruments used were: a) a teacher questionnaire, b) a

battery of curriculum-based written tests in mathematics (measuring cognitive skills), and c) a battery of tests measuring meta-cognitive skills in mathematics.

a) Teacher questionnaire

A questionnaire was used to measure teachers' skills in assessment. A validation study of the teacher questionnaire took place in the four participating countries in June 2019. During the validation study, data from a total of 574 teachers from the four participating countries were gathered and both across and within-country analyses were conducted by using the Extended Logistic Model of Rasch (Andrich, 1988). In each country, the procedure for detecting pattern clustering in measurement designs developed by Marcoulides and Drezner (1999) and then the Saltus model were used to classify the items into *stages of assessment skills*. A *three-cluster solution* was found to be the most appropriate and comparable solution across the four countries. Based on this solution, the same stages in each participating country were established. More specifically, according to the results of the within-country analysis, the three-cluster solution was able to explain at least 55% of the total variance in each country. By conducting an across-country analysis, the three-cluster solution was in a position to explain 69% of the total variance. These analyses provided empirical support to the validity of the teacher questionnaire. It is also important to note that we had to remove some items which were considered as problematic in specific countries. Translation of the final version of the questionnaire from English to Greek and Dutch were also carried out by the country teams and back translation was followed to ensure translation into the two languages was appropriate.

The final version of the teacher questionnaire was administered to all teachers of the experimental and control group both at the beginning and at the end of the intervention. It is acknowledged that the choice of a self-reported questionnaire to collect data on teachers' assessment skills raises questions concerning the validity of the data gathered. Given that assessment is not a one-instance process but an integral part of the teaching process, the use of a questionnaire was considered more appropriate for measuring a wide range of assessment skills situated at different phases of teachers' practice. For example, skills related to the planning and construction phase cannot be measured during a class observation, since teachers usually construct their assessment instruments outside the classroom, perhaps even at home. In addition, whereas assessment administration takes place during classroom instruction, classroom

observation would have given us just a part of the picture; for example, a teacher may use performance assessment to assess his/her students, but he/she may not have used it on that day. Furthermore, using classroom observation would have not allowed us to measure skills related to the recording or the reporting of data, since once again these phases usually take place outside the classroom. The fact that the participants were teachers who showed a special interest in improving their assessment skills and that they were informed that the questionnaire data will define the content of the TPD program to follow, increases the possibility that teachers were sincere in their responses. However, the limitation of collecting data through teacher self-reports is acknowledged. Nevertheless, the analysis of data using the Rasch and the Saltus models generated empirical support to the construct validity of the questionnaire. In addition, the predictive validity of the instrument was found to be satisfactory since data from the initial measurement (i.e., beginning of the intervention) per item were found to be strongly correlated with those emerged from the final measurement (i.e., end of the intervention). One could also argue that even if under or over rating occurred, this may have happened in both measurement occasions, and not necessarily in only one group. Therefore, by conducting a group randomization study and collecting data both at the beginning and at the end of the intervention enabled us to compare the progress that each group made irrespective of the limitations that arise from our decision to measure assessment skills through a self- report questionnaire.

b) Written tests in mathematics measuring cognitive learning outcomes

A battery of mathematics tests was used to assess students' cognitive learning outcomes at the beginning and at the end of the intervention. This battery of written tests was developed by a group of expert teachers and teaching mathematics academics in each participating country. More specifically, the following four Mathematics tests were developed for measuring cognitive skills: a) *Grade 6 test:* this test was based on the curriculum of Grade 6 and was used to measure achievement of students who were at the beginning of Grade 7, b) *Grade 7 test:* this test was based on the curriculum of Grade 7. It was used to measure achievement of students who were at the beginning of Grade 8 and of students who were at the end of Grade 7, c) *Grade 8 test:* this test was based on the curriculum of Grade 8 and was administered to measure achievement of students who were at the end of Grade 8 and of students who were at the beginning of Grade 9, and lastly, d) *Grade 9 test:* this test was based on the curriculum of Grade 9. It was used for

measuring student achievement at the end of Grade 9. Thus, the initial measurement tests were based on the curriculum of each previous grade whereas the final measurement tests were concerned with the curriculum of the current grade. It is, finally, important to note that at least 15% of common items were included in each test for equating purposes.

A validation study took place in the four participating countries during school year 2018-2019. All country teams managed to collect data from at least 120 students per grade. Given the length of each test and the available data, we conducted across-country analyses to test the validity of each test. More specifically, the Extended Logistic Model of Rasch was initially used and four different analyses (one for each test) were conducted. The fit indices that occurred showed that the data emerged from each test fit well to the model. Although the results of the across-country analyses provided empirical support to the validity of each test, it was found that in the case of Grade 8 test the separability index was low. Therefore, we decided to run descriptive analysis per country and search for items with extremely low or extremely high percentage of success. Since data collection for the validation study per country was done in a very small number of schools, we focused our attention to those items who had extremely low rates of success in more than one country. It was found out that almost all items had good facility values in each country. However, three items included in the Grade 8 test were found to have very low success rates both in Belgium and Greece and therefore it was decided to replace those items with new ones. The final version of the tests was then developed and translated into Greek and Dutch.

The test administered to Grade 9 students when they were at the end of the school year was purposefully more difficult than the one administered to Grade 7 students when they were at the beginning of the school year, so as to correspond to their age skills, maturity stage, and level of mathematics knowledge. As a consequence, Item Response Theory (IRT) was used for equating the tests (Hambleton & Swaminathan, 1985). Specifically, the scores were transformed into the same scale on the basis of the characteristics of IRT models, with students' latent level of ability (y) and difficulty level of an item (b) being identical when certain preconditions were fulfilled (Bond & Fox, 2001). The latent ability level for each student could be determined in every version as long as there were so-called 'anchoring items' connecting the versions. As it was mentioned above, sufficient common items (i.e., approximately 15 per cent of anchoring items across all tests) with representative content to be measured (Kolen & Brennan,

1995) were used. Thus, estimation was made using the Extended Logistic Model of Rasch (Andrich, 1988) and separate within-country analyses were conducted. The within-country analyses revealed that each scale had satisfactory psychometric properties in each country. Specifically, for each scale the indices of cases (i.e., students) and item separation were higher than 0.82, indicating that the separability of each scale was satisfactory (Wright, 1985). Moreover, the infit mean squares and the outfit mean squares of each scale were near one and the values of the infit t scores and the outfit t scores were approximately zero. Furthermore, each analysis revealed that all items had item infit with the range 0.84 to 1.19. Therefore, for each assessment period, achievement in mathematics was estimated by calculating the Rasch person estimates. It is finally important to note that none of the respondents achieved full score, and none showed full zero performance. Based on the range of the results, the ceiling and floor effects in the attainment data were not observed.

c) Written tests in mathematics measuring meta-cognitive learning outcomes

A battery of tests was used to assess students' meta-cognitive learning outcomes in mathematics at the beginning and at the end of the intervention. As in the case of the cognitive tests, these were developed by a group of expert teachers and teaching mathematics academics in each participating country. It is important to mention that these tests were based on an adaptation of the "Meta-cognitive Skills and Knowledge Assessment - MSA" tool (Desoete, Roeyers, & Buysse, 2001) which takes into account the theoretical framework of Brown (1978) and aims to measure two meta-cognitive components: 1) three types of knowledge of cognition (i.e., declarative, conditional, procedural) and 2) four types of regulation of cognition (i.e., planning, monitoring, evaluation, information management skills). More specifically, in this study, four Mathematics tests for measuring students' meta-cognitive skills were developed by following the same procedures as for the development of the Mathematics tests for measuring students' cognitive skills (see previous section).

A validation study took place in the four participating countries between during the school year 2018-2019. All country teams managed to collect data from at least 120 students per grade. Given the length of the meta-cognitive mathematics test and the data per scale that are generated, it was decided to search for the validity of four scales. Specifically, we treated the items measuring *declarative and procedural knowledge* as

belonging to one scale especially since all these items belong to one of the two overarching factors of meta-cognition according to Brown's theory (i.e., knowledge of cognition). This decision was supported by the results of all four exploratory factors analyses (one per test) which revealed that the first eigenvalue was extremely high. In each analysis, the one-factor model was able to explain more than 55% of the total variance. For this reason, we decided to treat the items of these two aspects of knowledge of cognition as belonging to a single scale. Therefore, we run four different across-country analyses (per test), to find out if the data that emerged from the study could help us generate the following four scales: *(1) Knowledge of Cognition, (2) Prediction, (3) Planning, and (4) Evaluation.*

For each test, the Rasch model was applied on the whole sample of students four times to test the validity of each of the above four scales using the computer program Quest. For each test, the fit indices of all scales but evaluation were found to be appropriate. In regard to the scale measuring students' evaluation skills, it was found out that only the data that emerged from all the items of Grades 6 and 9 meta-cognitive test fit to the Rasch model. By looking at the indices of each item in the Grade 7 and in the Grade 8 meta-cognitive tests, it was found out that one item of the Grade 7 test and two items of Grade 8 test could have been removed. For this reason, we conducted the analysis of all evaluation items of Grade 7 meta-cognitive test excluding the specific item and found out that the fit indices of the scale were substantially improved. In regard to the Grade 8 meta-cognitive test, we were in a position to generate a valid and reliable scale when one of these two items were removed. The final version of the meta-cognitive tests was then developed and translated into Greek and Dutch. Rasch analyses of data emerged from each measurement occasion provided further support to the construct validity of the meta-cognitive tests. It is, however, important to mention that the predictive validity of the scale measuring procedural and declarative knowledge (i.e., knowledge of cognition) was not found to be satisfactory. Therefore, for each measurement occasion, only three scores measuring regulation of cognition (per student) were used in our attempt to measure the impact of the TPD on promoting students' meta-cognitive skills.

Results

The main study was conducted during the school year 2019-2020. During the second term of the school year, the pandemic of COVID-19 had a negative impact in the operation of

all schools in the four participating countries. Due to the circumstances of the pandemic, a significant number of teachers and students who participated in the project and from whom data from the pre-measurement phase (i.e., September 2019) were collected were not able to participate during the final measurement occasion which took place in May 2020 (i.e., a period where lockdown measures were taken in most areas of the participating countries). Therefore, we had to check for any selection bias that may have occurred in our sample due to this unexpected situation. Specifically, it was important to investigate whether the teachers and students who participated only at the initial measurement (i.e., beginning of the school year 2019-2020) had different background characteristics from those who participated in both measurement occasions. In addition, we examined whether there was any difference in terms of any background characteristic (including prior achievement) between the experimental and the control group. In the following sections, the main findings of the study concerning the impact of the intervention on improving teachers' assessment skills and on student achievement (cognitive and meta-cognitive learning outcomes) are also presented.

1. Impact on Teachers' Assessment Skills

The first step was to search for any selection bias in our teacher sample. Specifically, 206 teachers participated in the pre-measure whereas only 166 participated in both measurement occasions. As a consequence, our first step was to compare the characteristics of those teachers who participated in both measurement occasions (n=166) and those who were not in a position to participate at the end of the intervention (n=60). The chi-square test did not reveal any difference between the two groups in terms of gender. Moreover, the t-test did not reveal any statistically difference at .05 level between these two groups in terms of years of experience and in terms of their prior skills in assessment. These results reveal that the teachers who did not participate in the final measurement occasion had similar characteristics with those that participated in both measurement occasions. However, this comparison was only possible to be made in relation to those variables that we had the chance to collect data. It is therefore not possible to find out whether the two groups differ in terms of any other characteristic that was not considered in this study and which might be related to their ability to improve their assessment skills. Our second step to control for any selection bias was to compare the teachers of the experimental and control groups who managed to participate in both

occasion measurements in terms of all the independent variables of this study (i.e., gender, years of experience and their assessment skills at the beginning of the intervention). The chi-square test did not reveal any statistically significant difference between the experimental and control group in terms of teacher gender in each participating country (Cyprus: $X^2 = 0.78$, df=1 p = 0.38; Belgium: $X^2 = 0.03$, df=1, p = 0.87; Greece: $X^2 = 0.53$, df=1, p = 0.47; the Netherlands: $X^2 = 0.02$, df=1, p = 0.89). Also, the t-test did not reveal any statistically significant difference between the two groups in terms of years of experience (Cyprus: t = 1.32, df = 62, p = 0.19; Belgium: t = 0.45, df = 12, p = 0.66; Greece: t = 0.49, df = 44, p = 0.63; the Netherlands: t = -1.97, df = 40, p = 0.06). In regard to the assessment of skills of teachers, it is important to note here that the Extended Logistic Model of Rasch (Andrich, 1988) was used to identify the extent to which the assessment skills measured by the teacher questionnaire could be reducible to a common unidimensional scale. It was, therefore, decided to treat the Rasch person estimates as measures of teachers' skills in assessment. Table 2 presents the means and standard deviations of teacher scores which emerged by measuring assessment skills at the beginning and at the end of the intervention per group. In addition, the results of the t-test, comparing the control with the experimental group in terms of their pre and final measure of their assessment skills are presented in Table 2.

Table 2. Means and SD of pre and final measures of teacher assessment skills per group across countries, and the t-values

	Control		**Experimental**		**t-test**		
	x	**SD**	**X**	**SD**	**t**	**df**	**P**
Pre-measure	-0.05	0.62	-0.06	0.48	0.12	164	0.91
Post-measure	-0.02	0.57	0.37	0.60	-4.25	164	0.001

The following observations arise from Table 2. First, it can be observed that the initial mean scores of the assessment skills of the two groups were almost the same. The t-test revealed that there wasn't any statistically significant difference at .05 level in terms of the initial measurement between the two groups. Second, one can see that the experimental group had a higher mean score at the end of the intervention from the control

group, and a statistically significant difference between the two groups can be observed (t=-4.25, df=164, p<.001). In addition, the t-test paired revealed that the mean scores of the teachers' assessment skills were higher at the end of the intervention compared to their scores at the beginning of the intervention. This difference was statistically significant at .001 level (t=13.17, df=86, p<0.001). This finding reveals that teachers employing the DA, managed to improve their assessment skills. On the other hand, the t-test paired did not reveal any statistically significant improvement in the skills of the control group (t=0.79, df=78, p= 0.43).

Multiple regression analysis was then used to test if any of the background characteristics of teachers (i.e., gender and years of experience), the initial performance of teachers in assessment and the use of the DA are associated with the final score of teachers' skills in assessment. Below, you can find the equation of the multiple regression analysis. This equation reveals that both the initial measurement and the use of DA are associated with the assessment skills of teachers at the end of the intervention.

*Post score = 0,028 + 0,865 * Prior Score + 0,397 * Group + residual*

As it was mentioned in the methods section, the Saltus model was used to find the stage of each teacher at the beginning and at the end of the intervention. Then, the Mann Whitney analysis was used to search for any differences between the control and experimental group in terms of teachers' stages at the beginning and at the end of the intervention. The Mann Whitney test did not reveal any statistically significant difference between the control and experimental group in terms of the stage that each teacher was found to be situated at the beginning of the intervention (Mann-Whitney, U=3357, z=-0.32, p=0.75). On the contrary, a statistically significant difference at the end of the intervention (Mann-Whitney, U=1869.5, z=-3.94, p<0.001) was found.

Finally, we compared the stage that each teacher was found to be situated at the beginning and at the end of the intervention. It was observed that none of the teachers of the control group managed to move from the stage he/she was found to be situated at the beginning of the intervention to a more demanding stage. On the other hand, a stepwise progression was observed in the experimental group since more than two out of five teachers of the experimental group (i.e., 42.53%) managed to move at the next more demanding stage. Specifically, 16 teachers managed to move from stage 1 to stage 2 and

17 teachers moved from stage 2 to stage 3. It is important to note that four teachers managed to jump from stage 1 to stage 3. One can therefore claim that three different analyses reveal that the intervention had a significant impact on improving teacher assessment skills. It is also important to note that these three analyses are concerned not only with the Rasch score measuring teacher assessment skills but also with the stage at which teachers of each group were found to be situated at the beginning and at the end of the intervention.

2. Impact on Students' Cognitive Learning Outcomes

This section presents the results concerning the impact of the intervention (TPD course) on students' cognitive learning outcomes in Mathematics.

a) Searching for selection bias

The first step was to search for any selection bias in our sample. Specifically, 5447 students participated in the premeasure out of which only 4012 participated in both measurement occasions. This implies that the percentage of the missing cases at the student level is relatively high (i.e. 26.34%). However, the t-test did not reveal any statistically significant difference at .05 level in terms of prior achievement between students who did not participate in both measurement occasions with those who were considered in the final analysis (t=1.59, df=5445, p= 0.11). Furthermore, the chi-square test revealed no differences between those who participated and those who did not participate in both measurement occasions in terms of gender (X^2= 0.98, df=1, p= 0.32).

Our next step was to compare the background characteristics of the experimental group with those of the control group of students who participated in both measurement occasions. The t-test revealed no statistically significant difference at .05 level in terms of student achievement in mathematics between the two groups. However, the chi-square test, revealed a statistically significant difference between the control and experimental group (X^2=10.9, df=1, p=.001) in terms of gender. Specifically, the percentage of boys of the experimental group was higher (47.2%) when compared with the percentage of boys of the control group (42.8%). This implies that we need to control for the observed gender difference between the two groups in our attempt to search for the impact of DA on students cognitive learning outcomes.

b) *Impact on students' cognitive learning outcomes*

At the next step, multilevel regression analysis was conducted to find out whether teachers employing the DA were more effective than the teachers of the control group in terms of promoting their students' cognitive learning outcomes in Mathematics. Table 3 presents the results of the multilevel analysis (students within teachers) of student achievement in mathematics at the end of the intervention. The following observations arise from Table 3.

Table 3. Parameter estimates and standard errors for the analysis of *Mathematics* achievement across countries (students within teachers)

Factors	**Model 0**	**Model 1**	**Model 2**
Fixed part			
Intercept	0.15 (.04)	-0.03 (.06)	-0.08 (.05)
Student level			
Prior achievement		0.59 (.02)	0.59 (.01)
Gender (0=boy, 1=girl)		0.13 (.03)	0.13 (.03)
Cyprus		0.19 (.07)	0.15 (.05)
Belgium		0.26 (.12)	0.23 (.11)
Greece		0.10 (.08)*	
DA (0=control, 1=experimental)			0.20 (.05)
Variance components			
Teacher	0.15 (.02)	0.08 (.01)	0.06 (.01)
Student	0.85 (.02)	0.61 (.01)	0.60 (.01)
Significance test			
X^2	11595.62	9523.30	9507.11
Reduction		2072.32	16.19
Degrees of freedom		4	1
p-value		.001	.001

**Not statistically significant effect at .05 level*

First, three different empty models were developed and was found out that the "students within teachers" model fit better to the data than any other model (i.e., students within classrooms or students within classrooms within teachers). Then, the following explanatory variables were added to the empty model: *prior achievement* as Z-score with a mean of 0 and a standard deviation of 1 {this is a way of cantering around the grand mean (Bryk & Raudenbush, 1992) and yields effects that are comparable}, *gender* as a dummy variable (0=boy, 1=girl) and *three* dummy variables to control for the country effect. Since the number of countries involved in this project was small, it was decided to control for any country effect by adding three dummy variables to the empty model and treating the Netherlands as the reference group. The following observations arise from the figures of the third column of Table 3 (i.e., Model 1). First, model 1 was able to explain 32.67% of the total variance most of which was attributed to the student level. Second, prior achievement and gender had statistically significant effects at .05 level on student achievement in Mathematics at the end of the intervention. With regard to gender effects, since we treated boys as the reference group, the positive parameter reveals that the girls had better results than the boys. Third, by considering the Netherlands as a reference group, it can be observed that students from Cyprus and Belgium had better results in Mathematics when compared to students from the Netherlands. It is important to note here that our intention was not to compare the results among the four countries but rather to control for any possible country effect due to the multilevel structure of our data. The results indicating different level of performance across countries cannot be seen as revealing differences in the effectiveness status of the four countries, since our sample is not representative at the country level.

In model 2, we searched for the extent to which the use of DA had a statistically significant effect on student achievement in mathematics at the end of the intervention. For this reason, the dummy variable DA (0=control, 1=experimental) was added to Model 1. As can be seen from the figures of the fourth column of Table 3 (i.e., Model 2), the DA was found to have a statistically significant effect on student achievement in Mathematics at .001 level and model 2 was found to fit better to the data than model 1. It is important to note here that the models presented in Table 3 were estimated without the variables that did not have a statistically significant effect at 0.05 level.

3. Impact on Students' Meta-Cognitive Learning Outcomes

This section is concerned with the impact of the intervention (i.e., the TPD course based on DA) on students' meta-cognitive learning outcomes in Mathematics. As was mentioned in the methods section, we were able to establish four different scales measuring: a) knowledge of cognition and b) three different types of regulation of cognition (i.e., *Prediction, Planning* and *Evaluation).* In the case of the scale measuring *knowledge of cognition,* the correlation between the initial and the final measurement was not statistically significant at .05 level. This implies that the predictive validity of this scale is not satisfactory. Therefore, we only searched for effects of the intervention on each of the three scales measuring the three types of regulation of cognition (i.e. *Prediction, Planning* and *Evaluation*).

a) Searching for selection bias

The first step of the analysis was to search for any selection bias in our student sample. More specifically, 5544 students participated at the pre-measure, of which only 3870 participated in both measurement occasions for each of the three scales (Prediction, Planning, Evaluation). However, using the t-test, we found no statistically significant difference at .05 level in prior achievement in each scale between students who did not participate in both measurement occasions and those who participated in both occasions. Moreover, the chi-square test did not reveal any statistically significant difference at .05 level in terms of gender. We also compared the control and the experimental group in terms of their background characteristics including their prior achievement in each type of regulation of cognition. The chi-square test did not reveal any statistically significant difference between the control and experimental group in terms of gender (X^2= 0.08, df=1, p= 0.78). In addition, the t-test revealed no statistically significant difference at .05 level between these two groups in terms of each of the three scales measuring students' meta-cognitive skills at the beginning of the intervention.

b) Students' meta-cognitive learning outcomes

In order to search for the impact of the intervention on improving students' meta-cognitive learning outcomes, separate multilevel regression analyses for each scale

measuring regulation of cognition (i.e. Prediction, Planning, Evaluation) were conducted. Tables 4, 5 and 6 present the results of these three multilevel regression analyses. The following observations arise from these tables. First, empty models for each scale were developed. The empty models in each of the three analyses revealed that the student within teacher model had better fit to the data than any other model. Then, the following background variables were added to the empty model: *prior achievement*, *gender* as a dummy variable (0=boy, 1=girl) and *country* as a dummy variable. Since the number of countries involved in this project was small (n=4), it was decided to model the country effects by adding three dummy variables to the empty model (i.e., treating the Netherlands as the reference group) and not to consider the country as an extra level of data. It can be observed that the effect of prior achievement was statistically significant in all three analyses at .05 level. In regard to the gender effect, girls had better results in prediction (as shown from the positive parameter for gender in model 1 of table 4), whereas in regard to Planning and Evaluation no gender differences were identified (see tables 5 and 6). In addition, in the case of prediction, it was found that only students from Belgium had better results than those of Netherlands (shown from the positive parameter for Belgium in Model 1 of Table 4), whereas students from Cyprus and Greece did not differ in their performance from those of the Netherlands. In regard to planning, there were no differences in students' achievement among the four countries. Finally, in regard to the evaluation scale, it was found that only students from Cyprus had better results compared with the Netherlands (see the positive parameter for Cyprus, in Model 1 of Table 6). It is finally important to note here that since our sample was not nationally representative, country differences were only examined due to the multilevel structure of our data and the need to control for the effect of this level in our attempt to search for the effect of the TPD course based on DA on students' metacognitive skills at the end of the intervention.

In model 2, we attempted to find out whether the TPD course on assessment based on DA had a statistically significant effect on students' meta-cognitive learning outcomes. For this reason, the dummy variable DA (0=control, 1=experimental) was added to Model 1. The results of all three analyses revealed that the DA had a statistically significant effect on students' meta-cognitive achievement at .001 level (see figures of the fourth column of Tables 4, 5 and 6). It is important to note here that the models presented in Tables 4, 5 and 6 were estimated without the variables that did not have a statistically significant effect at 0.05 level.

Table 4. Parameter estimates and standard errors for the analysis of *Prediction* across countries (students within teachers)

Factors	**Model 0**	**Model 1**	**Model 2**
Fixed part			
Intercept	0.11 (.03)	-0.00 (.05)	-0.06 (.04)
Student level			
Prior achievement		0.26 (.02)	0.26 (0.02)
Gender (0=boy, 1=girl)		0.13 (.03)	0.12 (.03)
Cyprus		0.10 (.06)*	
Belgium		0.29 (.14)	0.21 (.12)
Greece		0.03 (.07)*	
DA (0=control, 1=experimental)			0.23 (.05)
Variance components			
Teacher	0.07 (.01)	0.07 (.01)	0.05 (.01)
Student	0.85 (.02)	0.75 (.02)	0.74 (.02)
Significance test			
X^2	12040.63	10907.09	10885.40
Reduction		1133.54	21.69
Degrees of freedom*		3	1
p-value		.001	.001

**Not statistically significant effect at .05 level*

Table 5. Parameter estimates and standard errors for the analysis of *Planning* across countries (students within teachers)

Factors	**Model 0**	**Model 1**	**Model 2**
Fixed part			
Intercept	0.13 (.03)	0.09 (.05)	-0.01 (.03)
Student level			
Prior achievement		0.11 (.02)	0.12 (.02)
Gender (0=boy, 1=girl)		0.01 (.03)*	
Cyprus		0.08 (.06)*	
Belgium		0.10 (.13)*	
Greece		0.01 (.07)*	
DA (0=control, 1=experimental)			0.27 (.04)
Variance components			
Teacher	0.06 (.01)	0.06 (.01)	0.03 (.01)
Student	0.92 (.02)	0.81 (.02)	0.78 (.02)
Significance test			
X^2	12082.05	11494.12	11461.65
Reduction		587.93	32.47
Degrees of freedom		1	1
p-value		.001	.001

**Not statistically significant effect at .05 level*

Table 6. Parameter estimates and standard errors for the analysis of *Evaluation* across countries (students within teachers)

Factors	**Model 0**	**Model 1**	**Model 2**
Fixed part			
Intercept	0.13 (.02)	0.06 (.05)	-0.04 (.04)
Student level			
Prior achievement		0.26 (.02)	0.26 (.02)
Gender (0=boy, 1=girl)		0.03 (.02)*	
Cyprus		0.12 (.06)	0.12 (.04)
Belgium		0.23 (.13)*	
Greece		-0.00 (.07)*	
DA (0=control, 1=experimental)			0.26 (.04)
Variance components			
Teacher	0.05 (.01)	0.05 (.01)	0.03 (.01)
Student	0.90 (.02)	0.82 (.02)	0.78 (.02)
Significance test			
X^2	11418.07	10568.15	10534.79
Reduction		849.92	33.36
Degrees of freedom		2	1
p-value		.001	.001

**Not statistically significant effect at .05 level*

Concluding Remarks and Implications for Research, Policy and Practice

The aim of our study was to develop a TPD course that can have a positive impact on both assessment skills of secondary teachers and student learning outcomes (cognitive and meta-cognitive) in Mathematics, by using the DA. Unfortunately, the COVID-19 pandemic caused problems to the implementation of the main activities of the project during the last three months of the intervention. Specifically, based on the new laws and regulations that the governments of the four participating countries have taken to face COVID-19 virus, all public and private schools of all levels of education closed in March

2020. Consequently, the last (i.e., 5th) training session of the TPD course and the final measurement from teachers and students were rescheduled at a later point. It is acknowledged, therefore, that the pandemic has resulted in losing a relatively high percentage of our teacher and student sample, as indicated in the sections above. However, our attempt to search for any selection bias revealed that there was no difference between the teacher and student sample who participated in both measurement occasions and those who participate only at the initial measurement phase. In addition, almost no differences between the experimental and the control group were identified. It is finally important to note that the TPD course on assessment was found to have an effect on improving teachers' assessment skills and on promoting their students learning outcomes (cognitive and meta-cognitive). One could therefore claim that despite the difficulties that all country teams faced, it was possible to collect data from a relatively large sample and identify statistically significant effects of the TPD course not only on improving the assessment skills of participating teachers, but also on promoting their students' learning outcomes in mathematics (cognitive and meta-cognitive). As a result, implications of the findings for establishing a policy on TPD in assessment can be drawn.

The positive results of the evaluation of this intervention reveal that countries can make use of the DA to develop TPD courses on promoting formative assessment and through that achieve better student learning outcomes. More specifically, policy-makers can make use of the TPD course on assessment that was designed based on the main assumptions of the DA and which was implemented in this project and invite the respective stakeholders (i.e., educators/trainers) to make use of the Teacher Trainer Handbook (see http://www.ucy.ac.cy/formas/en/resources) to support teachers in improving their assessment skills and promoting the learning outcomes of their students. In addition, teachers who are interested in improving their skills in student assessment can make use of the Teacher Handbook (see http://www.ucy.ac.cy/formas/en/resources) that aims to support them to engage in a self-study process focused on improving their skills in assessment and through that on promoting student learning outcomes. We should, however, need to acknowledge that studies investigating the sustainability of this intervention are also needed, as well as research studies for exploring possibilities for scaling-up the specific project. It is finally pointed out here that readers who like to know more about this project can find more information and material in the official website of

the project, www.ucy.ac.cy/formas, for a comprehensive view of the outcomes of this project.

References

Andrich, D. (1988). A general form of Rasch's extended logistic model for partial credit scoring. *Applied Measurement in Education, 1*(4), 363–378.

Antoniou, P., & Kyriakides, L. (2011). The impact of a dynamic approach to professional development on teacher instruction and student learning: Results from an experimental study. *School Effectiveness and School Improvement, 22*(3), 291–311.

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.

Black, P. (2016). The Role of Assessment in Pedagogy – and Why Validity Matters. In Wyse, D. Hayward, L. & J. Pandya (eds.), *The Sage Handbook of Curriculum and Assessment*, (pp. 725 – 755). London: Sage Publications, Ltd.

Bond, T.G., & Fox, C.M. (2001). *Applying the Rasch model: Fundamental measurement in the human Sciences*. Lawrence Erlbaum Associates.

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

Boud, D., Keogh, R., & Walker, D. (Eds.). (2013). *Reflection: Turning experience into learning*. Routledge.

Brown, A.L. (1978). Knowing when, where, and how to remember: A problem of metacognition. *Advances in Instructional Psychology, 1*, 77-165.

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Creemers, B.P.M., & 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. (2012). *Improving Quality in Education: Dynamic Approaches to School Improvement.* London and New York: Routledge.

Creemers, B.P.M., Kyriakides, L., & Antoniou, P. (2013). *Teacher professional development for improving quality of teaching*. Dordrecht, The Netherlands: Springer.

Darling-Hammond, L., Oakes, J., Wojcikiewicz, S., Hyler, M. E., Guha, R., Podolsky, A., & Harrell, A. (2019). *Preparing teachers for deeper learning*. Cambridge, MA: Harvard Education Press.

de Boer, H., Donker, A. S., & van der Werf, M. P. (2014). Effects of the attributes of educational interventions on students' academic performance: A meta-analysis. *Review of Educational Research, 84*(4), 509-545.

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.

Desimone, L. (2009). Improving Impact Studies of Teachers' Professional Development: Toward Better Conceptualizations and Measures. *Educational Researcher. 38*(3), 181-199.

Desoete, A., Roeyers, H., & Buysse, A. (2001). Metacognition and mathematical problem solving in Grade 3. *Journal Learning Disabilities, 34*(5), 435-447.

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

Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., ... & Borman, G. D. (2016). Focusing on Mathematical Knowledge: The Impact of Content-Intensive Teacher Professional Development. NCEE 2016-4010. *National Center for Education Evaluation and Regional Assistance.*

Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer.

Hattie, J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

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

Hopfenbeck, T.N., & Stobart, G. (2015). Large-scale implementation of assessment for learning. *Assessment in Education, Principles, Policy and Practice, 22*(1), 1–2.

Khan, E. A. (2000). A case study of assessment in a grade 10 English course. *The Journal*

of Educational Research, 93(5), 276-286.

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

Kolen, M.J., & Brennan, R.L. (1995). *Test equating: Methods and practices.* Dordrecht, the Netherlands: Springer.

Kunter, M., Kleickmann, T., Klusmann, U., & Richter, D. (2013). The development of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 63-77). New York: Springer.

Kyriakides, L., Creemers, B.P.M., Panayiotou, A., & Charalambous, E. (2021). *Quality and Equity in Education: Revisiting Theory and Research on Educational Effectiveness and Improvement.* London and New York: Routledge.

Marcoulides, G.A., & Drezner, Z. (1999). A procedure for detecting pattern clustering in measurement designs. In M. Wilson, & G. Engelhard, Jr. (Eds.), *Objective measurement: Theory into practice* (Vol. 5). Ablex Publishing Corporation.

Randel, B., Apthorp, H., Beesley, A. D., Clark, T. F., & Wang, X. (2016). Impacts of professional development in classroom assessment on teacher and student outcomes. *The Journal of Educational Research, 109*(5), 491-502.

Schneider, M. C., & Meyer, J. P. (2012). Investigating the efficacy of a professional development program in formative classroom assessment in middle school English language arts and mathematics. *Journal of MultiDisciplinary Evaluation, 8*(17), 1–24.

Schneider, M. C., & Randel, B. (2010). Research on characteristics of effective professional development programs for enhancing educators' skills in formative assessment. In H. L. Andrade & G.J. Cizek (Eds.), *Handbook of formative assessment* (pp. 251–276). Abingdon: Routledge.

Wright, B.D. (1985). Additivity in psychological measurement. In E. E. Roskam (Ed.), *Measurement and personality assessment,* (pp. 101-112). Elsevier Science Publishers BV.

Formative Evaluation in Mathematics: An Exploration of Teachers' Attitudes, Self-efficacy and Experiences

Tine Mombaers & Peter Van Petegem

Department of Training and Education Sciences, University of Antwerp, Antwerp, Belgium

Email: tine.mombaers@uantwerpen.be & peter.vanpetegem@uantwerpen.be

Abstract: This qualitative study complements the broader (quantitative) FORMAS research project by elaborating on the attitudes, experiences, self-efficacy of Mathematics' teachers regarding formative assessment and what support measures they need to develop formative assessment further. Results show that teachers had little prior knowledge about formative assessment, and they see the added value of implementing formative assessment in Mathematics. Teachers' self-efficacy increased thanks to the workshops, but they still struggle with coming up with actions for some phases. This study also shows that personal teaching style and the type of class group affect the choice of activities that teachers will implement and that teachers could really use more support to facilitate and to further develop effective formative assessment in their Mathematics lessons.

Study context and research goals

The FORMAS research project studies the effect of supporting formative evaluation in Mathematics in secondary education. It is funded by the European Union and is conducted in four EU countries: Greece, Cyprus, The Netherlands and Belgium (Flanders). This qualitative research was carried out in Flanders, Belgium. This qualitative study complements the quantitative results of the teacher questionnaires.

The goal of this research was to get more insight in the prior knowledge, attitudes, self-efficacy in terms of formative evaluation in Mathematics and to find out how teachers experienced the workshops. An additional goal was to investigate which support measures they consider important to (further) develop formative evaluation. These aims result in the following research questions:

(1) What prior knowledge and attitude did teachers have about formative evaluation in Mathematics?

(2) How did teachers experience the workshops and what did their action plan consist of?

(3) How did teachers' attitude and self-efficacy regarding formative evaluation in Mathematics evolve throughout the FORMAS trajectory?
(4) What support measures do teachers find necessary to (further) develop formative evaluation (in Mathematics)?

Method

All (Flemish) teachers who had already participated in the overall study were asked to volunteer to participate in this qualitative study. This convenience sampling method resulted in a sample of six secondary school Mathematics teachers from five schools, five women and one man. Their teaching experience ranged from 5 years to 30 years. Four of the teachers had more than 20 years of teaching experience. They all took part in the FORMAS trajectory, more specifically they engaged in five workshops throughout the school year and they filled out teacher questionnaires.

Semi-structured interviews were carried out. Some of the main questions were derived from relevant literature. The core of the interview guide was formed by questions about the action plan teacher developed, attitude, self-efficacy and desired support.

The semi-structured interviews were carried out from mid-September until mid-October 2020. As the Covid19-pandemic did not allow us to interview participants in person, Skype was used as a platform. All interviews were audio taped and transcribed verbatim. The average duration of the interviews was 30 minutes.

Afterwards, the following themes were highlighted in the transcripts of all interviews: prior knowledge, attitude, action plan, self-efficacy and support measures. Given the limited number of interviews, they were not coded, but summarised. A summary was made per theme, creating a thematical overview. This overview facilitated vertical (per respondent) and horizontal analysis (across respondents) per theme.

Results

Prior knowledge

All participants thought that they had no or little prior knowledge on formative evaluation before they started the FORMAS trajectory. They did not really know what formative evaluation was exactly about. But in retrospect, most of the questioned teachers recognised a lot of techniques that they already used, without realising that they were applying formative evaluation in their math classes.

"I did not have that much prior knowledge on formative evaluation. But I heard some things in the workshops that I thought: Oh, I do this and I do that. So before the workshops not really, but by hearing all that it seemed that I knew more than I originally thought and I applied it more than I thought in beforehand." (R1)

Activities that several participants mentioned that they already made use of were assessing the initial situation, asking specific thinking questions to check students' understanding and letting students work in groups according to their math level.

Workshops

The questioned teachers reported that the cycle of formative evaluation provided more structure and clarity on what formative evaluation comprises and what the different steps are. Most participants added that the cycle itself did not provide guidelines to come up with actions. Most of them reported that networking, exchanging ideas and tips and tricks with their fellow teachers helped them to think of actions.

"I was under the impression that during the sessions there was expected a lot from us, that we were asked a lot. And that there were a lot of opportunities to reflect. And that we primarily shared experiences and that the lecturer did not tell us much or that there was not a theoretical framework that was developed and that was further discussed in groups. But we were immediately asked: What is your experience and what do you already do? The focus was primarily on learning from each other." (R2)

Action plan

During the FORMAS trajectory, participating teachers were asked to set up an action plan. Most teachers implemented actions for all five phases of formative evaluation. Three participants mention that they formulate goals in phase one. They explain their students these goals in their lessons, but one of the teachers also writes these goals on a test paper. One teacher explicitly points out that she does not add criteria to these goals, because it would take too much time to do so. Another teacher says that she works with criteria, but indicates that she has not yet worked out these criteria into rubrics, because it takes some time to do so. Another teacher does not see the surplus value in clarifying expectations.

> *"I have also tried the suggestion to tell our students what was expected from them before I started my lesson. Of course that was no problem for me, but I couldn't see its added value. I experienced that my students, regardless of whether I said what I expected from them or not, during the regular classroom activities, I also found that it was clear to my students that they knew what was expected from them."* (R2)

All questioned teachers carry out actions for phases two and three. Regarding phase two, teachers state that they elicit students' reactions using erase boards, colour cards and digital tools. Digital tools are used by all questioned teachers and they include Wooclap, BookWidgets, Kahoot and Socrative. Participants feel that these digital and non-digital tools facilitate getting more insight in what students understand and are capable of, which helps them in phase three. They point out that because of this important insight, it is easier to guide students and implement remedial actions. Yet, there is one teacher who is not convinced of the added value of using Wooclap in his class group because he only has a class group of four students. Just asking them and letting them put up their hands would give him the same information, as far as he is concerned.
Regarding phase four, four of the questioned teachers indicate that it is hard to communicate individually with students about their progress, primarily because it is very time consuming. Thus, they have not found an effective way to do this yet. There is one teacher who says that she provides students with more feedback on tests.

Some questioned teachers point out that the choice of specific actions depends on their teaching style, but also on the class group. They indicate that some actions do not fit their personal teaching style. There is for example one teacher who likes to improvise in her lessons and she does not feel comfortable using exit tickets and erase boards, but she likes working with colour cards. A second teacher indicates that exit tickets are not her cup of tea because she is too chaotic and mostly does not have any time left at the end of her lessons. In contrast, there is a teacher who points out that she likes working with exit tickets, because they show what is left of the lesson she taught and this way, she can easily see who needs extra explanation or who does not.
Not all actions are appropriate for all class groups according to the questioned teachers. A teacher mentions that while using placemats in a class, characters of certain students collided and this resulted in refusing to help each other. Another teacher says that in one

class group using erase boards does not work, because the students do not use them seriously.

> *"Yes, and it can differ across class groups. I have class groups, that if they get erase boards, they only write silly things on them and then pens will fly through the air. And then you do this in another way with them."* (R4)

Attitude

Before starting the FORMAS trajectory, four out of six participants report seeing the added value of formative evaluation and they were in favour of using it in Mathematics. They think it is important not to wait for a summative test to find out what their pupils have learned. One teacher points out that through formative evaluation, he can spot certain signals in time. This allows him to support these students and undertake remedial actions if necessary. Another teacher adds that she finds that formative evaluation helps to motivate students to take control of their own learning process and to gain insight in where they stand.

One participant says that she did not have a particular opinion about it, because she was not aware of what formative evaluation entailed. Another questioned teacher indicated that she did not see the advantages of formative evaluation in Mathematics. She related this to the wrong image she had about formative evaluation.

> *"I thought: Mathematics without grades, that's not possible."* (R6)

Having completed the FORMAS trajectory, all participants state that they are all in favour of formative evaluation. This includes, for example, seeing the development of students and paying more attention to subjects that students do not master yet. Strikingly, the questioned teachers have different views on how formative evaluation should be implemented. Two teachers want formative and summative evaluation to coexist and complement each other, while two other teachers would like to cut summative evaluation all together. But one of these teachers adds that this approach requires the whole school taking part and not only the team of Math teachers. Remarkably, the two teachers from the same school differ in this attitude.

> *"The FORMAS trajectory has given me insight in: I find formative evaluation something really good, but not when it is not generally applied by all teachers. That way, it is very hard to achieve because you process things in different ways. I would prefer abolishing all grades and only evaluate formatively. But*

unfortunately enough, when you do something and you let students do exercises, you very quickly get the question: Oh, but Miss, do we receive a grade for this? And that they will only put in effort in that case. Eventually, when you integrate formative evaluation and dedicate enough time to it, it can be useful, but the students have to be open to it." (R3)

Though seeing the advantages of formative evaluation, two participating teachers point out that there are limits in terms of time allocation. Implementing formative evaluation in the Math lessons should remain achievable and time should primarily be spent on preparing good exercises and not on administration.

For most questioned teachers, there is no difference in their attitude towards the various phases. They find all phases evenly important. When one phase is neglected, the whole system collapses in their view.

Finally, most teachers find it pleasant to get into formative evaluation, learning more about it. They also point out that they enjoy coming up with actions.

Self-efficacy

Prior to the FORMAS trajectory, most questioned teachers indicated that they did not feel very competent in applying formative evaluation. Half of the teachers report that they did not feel competent enough to tackle practical issues and try out actions with their students. There is one teacher however, who stated that this did not impede her from undertaking several actions. Setting goals was a strength that two participants mentioned that they already had before starting the FORMAS trajectory.

Thanks to the workshops, all participants state that they feel more competent in formative evaluation. They feel more confident, though they are also aware that they have more to learn. The questioned teachers mention that they think they possess the necessary skills to further develop formative evaluation in their Math lessons. Five teachers point out that they feel capable and have the confidence to try out several actions through trial and error, in search of higher efficiency and a better organisation.

Confidence in their capacities for implementing the different phases of formative evaluation differs across participants. All teachers mention that they struggle with implementing phase four. They point out that they find it hard to communicate efficiently with their students about their development in a structural way. These teachers also report

that they find it hard to find a manner that pays off for students, but is not too time consuming for them as a teacher.

> *"What I am struggling with, is students seeing their growth for themselves and I have not managed that piece just yet. You can keep track of it all, but it is only useful when students themselves can see in what way they have grown and how they can keep track of this in a way that is not too much work for me and that the student will see what works and what does not."* (R5)

Another phase in which participants lack confidence, is phase one. One teacher says that she is thinking hard about a good way to visualise these goals for students, while another teacher says that she is having difficulty with this phase due to her chaotic nature. There are two teachers who report that they lack competence in phase five. They believe that they should get more out of that phase than they do at this moment.

Overall, most teachers feel that they possess the necessary competencies to further develop formative evaluation in their Math lessons. They all state that they feel confident enough to grow further, but there is one teacher who is not sure about further unwinding formative evaluation in his lessons.

Support measures

There are three types of support measures that the questioned teachers would want to see to effectively implement formative evaluation further: organisational and material support and professionalisation.

Table 1. *Required support measures for formative evaluation (FE)*

organisational support	**professionalisation**	**material support**
possibility to intern with colleagues	getting help from other schools / organisations that have successfully applied FE	digital support
lesson blocks (of 2 teaching periods)	educating school management on FE	larger classrooms

parallel teaching hours with colleagues to co-teach to get smaller groups	training or course on activities that have proved to be effective in FE
freeing up time to collaborate	
report card with colours	

First, **organisational support** was highlighted by half of the teachers. Each teacher mentioned another example of this type of support. The first example that was reported is giving teachers the possibility to intern with colleagues. That way, teachers can learn more from one another. In addition, a teacher says that lesson blocks of two teaching periods would enable spending more time on activities regarding formative evaluation. She also states that parallel teaching hours would facilitate co-teaching and as a result, working with smaller groups.

> *"The fact that you are parallel with two colleagues in a not too large class group, then you can co-teach. I think that in that case it would be easier to use formative evaluation, because you can work with smaller groups. A group that you know of they are busy, you can put elsewhere... So I would primarily focus on that, because that's what I miss sometimes."* (R3)

Another teacher states that freeing up some time to work together on implementing formative evaluation is desirable. Teachers should have time, she says, to consult each other on a regular basis. She indicates that at her school, there are more teachers involved in formative evaluation, but they are working alongside each other instead of with each other.

A second example of organisational support is working with a different report card system. This teacher states that she would like to have a report card system in which formative evaluation can be included, next to summative evaluation. Colours can be used to indicate students' progress on partial aspects of Mathematics. She feels that this system could help indicating per criterium what the student masters or not.

Second, **professionalisation** was reported by half of the questioned teachers as one way to develop formative evaluation in Mathematics. On the one hand, professionalisation of the teachers was reported. Getting help from a school or an organisation that has already successfully implemented formative evaluation in a school,

is considered very helpful by one of these teachers. This would mean that they are not required to invent everything themselves, but can rely on good practices in other schools. Another teacher indicates that she would like to have extra training on activities regarding formative evaluation that are effective, based on previous research. On the other hand, one teacher points out that her school management must learn in the first place what formative evaluation is in order to be able to support their teachers.

> *"I have to say that my school management also has to learn what formative evaluation is. One team of school managers already knows what it is about, but the other isn't. We have seven members of school management at our school. It depends on in which department I am teaching and the extent you are supported or not. There is one department in which I'm highly supported and another one in which I am thwarted. ... They have to be immersed in formative evaluation, as I have been."* (R6)

Third, there are two teachers that mentioned **material support** would help implementing and developing formative evaluation. Both indicate that digital support is indispensable in formative evaluation. They would like suitable ICT equipment for their students (computers or tablets) to use digital tools. Larger classrooms is another way of material support that is mentioned by one of these two teachers. She reports that larger classrooms would facilitate using other working methods.

Conclusions

This study shows that the questioned teachers had little or no prior knowledge on formative evaluation. Therefore, it is paramount to educate teachers on formative evaluation, so they know what it means. It was also clear that these questioned teachers already implemented different actions that had to do with formative evaluation. This shows that there are probably still a lot of teachers who do not know what formative evaluation exactly entails and they may not realise that they already apply formative evaluation to a certain extent. With this insight, they may be easily persuaded to use formative evaluation when they realise that this will not take too much effort.

Moreover, most participants saw the added value of formative evaluation in Mathematics. We do have to admit that the questioned teachers were probably highly motivated to learn more about formative evaluation and use it in their lessons, otherwise they would not have enrolled in the research project (workshops included). In this study

we distinguish two different attitudes towards formative evaluation in Mathematics. The first view is that formative evaluation should complement summative evaluation. They must go hand in hand. The second view is that formative evaluation should be the only form of evaluation, thus summative evaluation should not be organised. These results show that the attitude and view on formative evaluation is very personal. When there are different views in the same teaching group or school, teachers (and school management) should find a compromise in how they want to evaluate.

Participants reported that their self-efficacy increased thanks to the workshops. According to the teachers in this study, they received theoretical information in the workshops from the lecturer and practical tips and tricks from their peers. This shows that educating teachers through workshops can be highly beneficial to develop higher self-efficacy. Thus, enough effort must be made by policy makers, school management and pedagogical counselling in instructing teachers on what formative evaluation is about and how it can be applied.

Most teachers implemented actions in their action plan for all phases of formative evaluation, but actions in phase two (getting and collecting student reactions) and three (analysing and interpreting students' reactions) were reported the most. Carrying out phase four is, according to most questioned teachers, very difficult. They find it hard to come up with an effective way to map students' progress and visualise this for their students, without taking too much of their time. Further professionalisation on this particular phase, may help teachers tackle this problem.

This study also shows that personal teaching style and the type of class group affect the choice of activities that teachers will implement. Teachers are more likely to use certain activities if they match their teaching style. Moreover, the effectiveness of activities can rely on the type of class group. Therefore, it is up to teachers to find the activities that best reflect their teaching style and that are appropriate to use in their particular class groups. In formative evaluation, there is not 'one size that fits all'.

Finally, we can conclude based on these research results that teachers could really use more support to facilitate and to further develop effective formative evaluation in their Mathematics lessons. School management should increase their support in terms of organisation and materials, but also in terms of professionalisation opportunities.

Three Secondary teachers explore students' understanding of Mathematics through Formative Assessment

Eirini Kouletsi[a], Kalliopi Ardavani[b] and Eleni Tsoukala[c]

[a]*Teacher trainer (K-12 Education), MEd, Athens, Greece*

[b] *Teacher trainer (K-12 Education), Athens, Greece*

[c] *High School teacher (K-12 Education), Athens, Greece*

Email: kouleir@gmail.com

Abstract: This paper aims to describe three mathematics teachers' didactic proposals designed and conducted with secondary school students to study the integration of formative assessment and self-assessment tools in mathematics courses, through their participation in the FORMAS Professional Development Project. The outcome indicates that both students and teachers have benefited. In order to implement formative assessment techniques, new pedagogy had to be employed. In the light of the new working environment the students felt free to take responsibility of their learning process, whereas the teachers tackled with new learning and teaching issues that support their professional development. The teachers feel the need to urge for ongoing, professional learning regarding formative assessment.

Introduction

Formative assessment in general and for mathematics especially, in short, is defined as a process used during instruction to adjust teaching and learning. It can be seen as the glue that holds the dimensions of a mathematically powerful classroom together (Burkhardt et al, 2019).

At the same time the FORMAS Professional Development Project aims to contribute in improving professional standards of secondary teachers by supporting them to conduct assessment for formative reasons (including Student Self-Assessment) that can help students identify learning needs and take actions to address these needs.

We are three high school mathematics teachers who, through our participation in the FORMAS project, utilized various techniques for assessing the involvement of students with mathematics, with the aim of enhancing the learning and teaching process. Our involvement in the project included professional development consisting of four modules: participation in an educational community through lectures held at the University, implementation in practice, reflection on our practice and constructive feedback from our group leaders and the rest colleagues- participants in the project.

When promoting formative assessment, we had to outdo special challenges faced, such as changing the standard classroom contract, reflecting on teaching methods to elicit student thinking and furthermore addressing the need to respond constructively to evidence from student performance during the lesson.

On the other hand, the students through their engagement in the following tasks, felt that they worked in a positive learning environment, which helped them identify their misconceptions as well as develop a capacity to work on a metacognitive level.

In this paper we present three didactic proposals employing different formative assessment techniques.

Theoretical issues

How educators view formative assessment in the classroom will determine the ultimate affect it has on students. Burkhardt et al (2019) argue that formative assessment for learning mathematics goes beyond surfacing of the students' understandings misconceptions and thought processes and using them in modifying the course of instruction. It entails a change in pedagogy, that is changes in classroom roles and culture must follow for both students and teacher.

A large body of evidence suggests that one of the major topics to be investigated in this field is self-assessment and hetero-assessment, as research has revealed that a student through self-assessment:

- has the ability to think about learning issues and thus enhance the development of metacognitive skills
- has several opportunities to focus on specific areas. He is able to understand the way of evaluation (and through the rating), does not consider the points arbitrary nor does he try to improve his performance solely by this criterion
- empowers his self-esteem
- through the feedback it offers the student and the teacher understand where he is in relation to the learning objective (Black et al, 2004).

In the meanwhile, hetero-assessment activities can help develop the objectivity required for effective self-assessment (Black et al, 2004, p.8). All these lead to the autonomy of the student, which is the cornerstone of the educational process (Black et al., 2004; Hattie and Timperley, 2007).

Embedding Formative Assessment methods in mathematics classroom

In the following three tasks indicatively and for the sake of variety each one concerns a different technique of evaluation applied by the authors.

Task1: Evaluating the execution of an activity

At the High School of Nea Chalkidona, Eleni chose to present the evaluation of her students' ability to perform an activity that she had prepared for this purpose. The mathematics class comprised 20, 8th grade students divided in 5 groups, who were asked to calculate the height of the wall of the mathematics classroom from the edge of the ceiling to the floor. For the indirect calculation of the height the concept of the tangent to an angle is necessary. The tools at hand were measuring rule, goniometer, paper, pencil and pocket calculator.

Each member of the group took on a specific role: one organized the mathematical modelling, another took measurements, the third recorded them while the fourth one carried out the calculations and extracted the results (Figure 1).

Figure 1: Division of labour

The learning objectives assessed by the teacher were the students' ability to design and execute the solution steps, to pinpoint any difficulties arising during the process, to find a solution to them and finally evaluate the outcome. The task was considered to be successful as long as the students were able not only to complete the task but also interact harmonically during the whole process.

In the next phase, each group reflected on its own solution to the problem in the light of the discussion and unanimously concluded that they will be able to calculate the height of any inaccessible point, as long as they apply the same method. When they were asked to reflect upon the assessment process, they pointed out that it helped them identify their weaknesses and the points they need to strengthen, but most of all that they worked

in a positive learning environment.

During the performance of the activity the teacher discussed with each group the difficulties they encountered and challenged them to redefine the steps to take. Thus, students applied self-evaluation, as they were asked to investigate how effective the plan they implemented was, as well as hetero-evaluation, when exploring the effectiveness of the suggestions of the other members of the group.

Task 2: Written Assessment- Students critiquing student work examples

At the 2nd High School of Tavros, Kalliopi presents a set of Self-assessment and Hetero-assessment activities that 12, 9th grade students carry out individually during the lock down 2020 period, to revise the solution to a rational equation.

In the first phase the students engage in the task displayed in a hot potatoes interface where the steps to solving the equation are mixed up and must be rearranged.

The next phase asks students to identify the prerequisites needed to solving a rational equation and then to critique carefully four selected designed solutions to the same rational equation, presented by 4 supposed students (A, B, C and D), to point out mistakes, assess and grade them. The student's task is to understand each sample solution, to work out its strengths and limitations (none is correct) and to explain in written his critiques to the teacher.

By unpacking the errors made in the four sample solutions (Figure 2) we can see that they vary from not multiplying some of the terms of the equation with the LCD to omitting part of the LCD, or other to misusing the distribution of negative signs property or finally to applying a combination of the above mistakes.

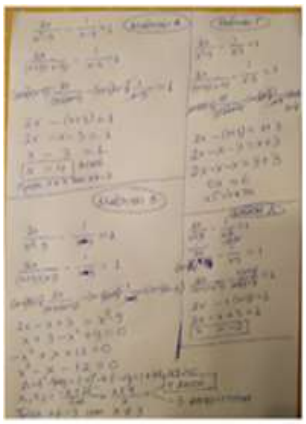


Figure 2: The four solutions to the rational equation

We quote one students detailed grading to Student's A solution, as her justification has followed the exact steps the teacher had applied when she solved a similar equation in classroom prior the lockdown (Figure 3): "*Total score 15/ 20: he factored correctly all the denominators 4/4, identified the restrictions 2/2 determined the LCD but didn't apply it correctly on both sides 2/3, cleared correctly the fractions on the terms he had applied the LCD 1/2 and did the multiplication correctly on the terms he applied the LCD 2/3. Finally, he solved correctly the regular equation, but found the wrong result 4/6*"

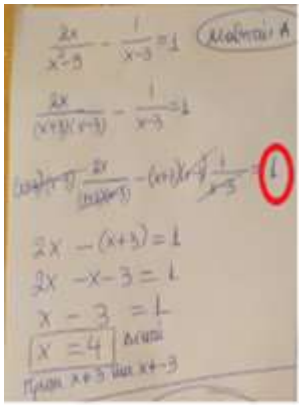


Figure 3: Student A solution

There was a wide range of grading assigned to each solution. The students apart from determining the errors tried also to justify them. Their justifications to the same mistake ranged from a "*serious mistake*" to a "*mistake of carelessness*" (version B), "*He has probably just forgotten it or is blind*" (version C), "*He may have made this mistake out of sheer momentum or carelessness*" (version D).

Through the process of peer assessment, the teacher had the opportunity to deepen on the way of thinking of the students and on their misconceptions and help them debugging happen naturally.

Task 3: Written Assessment- Students developing metacognitive thought

At Varvakeio Model Junior High School, Eirini, presents an activity in which students engage in a self-assessment project. Two 8th grade math target classes comprising of 24 students each, after having completed proportionality sit an exam where they have to work out a percentage problem. In the next lesson each student is asked to correct part of his test, having as a guide an exemplar solution to the problem. The exam was graded using the Analytical method where the rating was distributed to the individual sections of the problem (Figure 4).

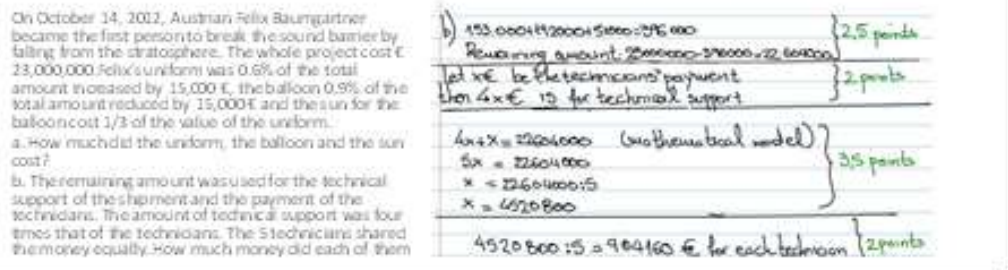


Figure 4: The test assigned to target class 1 and the answer key to part b

The students worked diligently the whole class hour. When their process of solving the problem was different from the given one, they tried on the one hand to find those mathematical elements that legitimized the solution to the problem and on the other to decide on the relevant rating, while the teacher walked around the classroom to help and guide the students as they worked.

In the next phase was conducted a small-scale research. Each student's assignment was to reflect, in written, on his experience of self-assessment part of his test and to share it with the whole class. The results are:

Target classes 1 and 2 reasoning lie under the following categories:

- Understanding teacher evaluation criteria (8/24, 7/24): "*I understood the reasoning of the teachers when they correct, their… criteria*".
- Understanding misconceptions (13/24, 14/24): "*I could see my mistakes better, because when I get the test corrected I wrongly, of course, pay more attention to the grade and I do not see my mistakes globally*"
    - Creating a positive learning culture *"Apart from the fact that it reduces stress it gives you the opportunity to correct your mistakes, yourself and be honest to yourself"*

    There was differentiation in the following categories. In Class 1 from the data processing emerged a fourth category:

    - Enhancing metacognitive skills: "*We managed to sharpen our critical ability*", whereas in class 2 the self – esteem factor arose from the data:
- Image of oneself-Difficulty of self-assessment: "*the most difficult thing was as we saw our careless mistakes is that we could not correct them, but it was a very good way to judge ourselves objectively*",

Discussion

To be able to implement formative assessment techniques we had to change the "classroom contract", that is to change drastically the roles assigned. While the students came to fore, stepping into our shoes, took responsibility for and managed their own learning, we receded into the background, monitoring the process, being helpful when needed, encouraging them whenever they met a difficulty.

Moreover, we faced challenges such as how to elicit student thinking and

furthermore how to respond constructively to evidence from student performance during the lesson. For that reason, with the feedback generously provided by our group leaders and the rest colleagues- participants in the project during our discussions at the university, we resorted to adopt new didactic approaches, with the aim that students through their engagement would enhance the development of metacognitive skills, all of which support our professional development.

On the other hand, the student's reflection upon their engagement in the formative assessment tasks are in line with the findings of the related research to this area. He is able to understand the way of evaluation, while has the chance to think about learning issues and thus enhance the development of metacognitive skills. These contribute to his better feedback regarding the learning process and thus by improving his self -esteem moves towards his autonomy from the teacher.

In Conclusion

We feel benefited by participating in this project.

It is helpful to evaluate each step in the educational process and to know how to set the criteria on which the assessment will be based. By the same token, the students had to change from behaving as passive recipients of the knowledge offered by the teacher to becoming active learners who can take responsibility for and manage their own learning.

Moreover, the formas professional development project helped us realise how a complex array of features and conditions must come together if teachers and districts are to benefit fully from professional development on formative assessment. We stress the need for ongoing, professional learning that invests in teachers so that we can develop more the skills required for formative assessment implementation in mathematics classroom.

Acknowledgements:

We would like to thank Dr. George Psycharis (National and Kapodistrian University of Athens) and Dr. Elissavet Kalogeria who, through their support and constructive feedback during our participation in the FORMAS project helped us improve our skills in formative assessment and reflect on our practices but above all they helped us realize that we need to change our daily practices in the classroom and to turn our lens to the students, the

great protagonists of the educational process, if we want to help them become autonomous.

References

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. Phi delta kappan, 86(1), 8-21.

Burkhardt, H., Schoenfeld, A. H., Bennett, R., Andrade, H., & Cizek, G. (2019). Formative assessment in mathematics. *Handbook of formative assessment in the disciplines. New York: Routledge*.

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

Using Formative Assessment in Helping Students Understand Geometrical Concepts: The Case of Area

Polyxeni Tsitsa

Secondary Mathematics Teacher at 9th Junior High School of Acharnes, Athens, Greece.

Email: polyxtsitsa@gmail.com

Abstract: This paper deals with formative assessment in Mathematics. Its purpose is the presentation of formative assessment activities used in classroom in order to help students understand the concept of area of 2-Dimensional shapes and assist them with their learning. The tasks were designed for the students of 8th grade of 9th Junior High School of Acharnes during a professional development course for mathematics teachers working in lower secondary schools at the National and Kapodistrian University of Athens under the coordination of the Professors Theodossios Zachariades and Giorgos Psycharis (in the context of FORMAS implementation in Greece).

Assessment for learning

While teaching mathematics to students in the 8th grade the teacher-author noticed that many of them had difficulty calculating areas of 2-Dimensional shapes, even though they knew the formulas. According to Earl and Katz (2006) a teacher can use assessment tasks in order to find out what the students really know or can do (assessment for learning). For this the teacher decided to use formative assessment activities to diagnose students' needs and help her students understand the elements used in calculating areas. In addition, she supported her students' learning with differentiated teaching, a decision that agrees with the personalized instructions for learning of Tomlinson (2014).

Three Formative Assessment Activities

The teaching scenario included a series of tasks (shape detection in a Malevich's painting, gamified tests, etc.). The tasks presented in this article are the main three formative activities:

- The correction of a test and the formation of the corresponding assessment criteria by the students

- An activity of peer-assessment where each student randomly selects the worksheet of another student.
- An activity of students' self-assessment.

1rst Activity of Formative Assessment

In the first activity of formative assessment students were asked to correct a test of an imaginary fellow student and create assessment criteria for that test.

Description and objectives of the activity

For the sake of the classroom dialogue and the simulation in real conditions, three different work samples were given to the students. Each worksheet contained shapes (an amblygonal triangle, a trapezium and a parallelogram) and answers with the most characteristic errors about the identification and the area of the shapes.

The activity aimed (through the change of roles as the students become teachers-assessors) to arouse the interest of the students, to involve them in the process, to make them ask their own questions, to help them identify shapes and use the appropriate formula to calculate the area, to help them decide on the correctness of the solutions, to make them think about the necessity of having assessment criteria, to encourage them to collaborate. In other words, to enhance students' motivation and commitment to learning (Earl & Katz, 2006). For the teacher the activity was an opportunity to identify learning gaps, misconceptions and the students' needs.

The students worked on the test in groups of two, enhancing communication and collaboration among them. They identified errors, found the correct answers, justified their corrections and suggested assessment criteria. The teacher's strategy was to provide differentiated individual assistance and appropriate feedback, but not the complete solutions. If for example a student had difficulty in identifying the shape in the exercise, he/she was encouraged to rotate the page to see it from another angle or to make a list of what he/she knows about the sides and the angles of the shape. In some cases, the students got auxiliary material (worksheets with shapes and formulas). Besides the suggestions, the teacher posed guiding questions: If a student, for example, found it difficult to locate the corresponding height, he/she was asked about the angle that the height forms with the side.

Highlights of the Activity:

(1) The way of working and the mistakes of the students were revealed to a great extent.

According to James (2017) in order for learning to be secure, all answers must be probed and the misconceptions explored. Therefore, the teacher discussed with the whole class: (a) the mistakes that were made (e.g. a student did not identify correctly the height that corresponds to the base of a triangle), (b) the misconceptions (e.g. a student cannot distinguish between a rhombus and a plain parallelogram), but also (c) the good practices (e.g. a student who is not certain that a shape is a trapezium, is able to calculate correctly the area of the shape by using analysis or synthesis of other shapes, in this case a rectangle and a triangle).

During the activity (from the learner´s responses) and by examining the worksheets the teacher made useful observations about the way her students work and the mistakes they make. That was helpful for designing her next teaching steps. It is also in line with Black and William (1998) and their comments regarding students' involvement in the assessment process.

(2) Creating assessment criteria proved to be very helpful for the students in two ways: To understand their mistakes concerning the calculation of the area and to know what they have to do while solving the exercise.

For example, a student admitted that someone must have studied before doing the exercises and suggested the following criteria: "To be able to identify the shape, to use the correct formula, to avoid making mistakes when calculating, to identify correctly the corresponding height, to explain and justify satisfactorily the results."

As a result, the students' involvement in the assessment process reflected on their own learning. They gained a better understanding of the requirements expected in an exercise and improved their knowledge on area of 2-Dimensional shapes.

2nd Activity of Formative Assessment

The second activity of formative assessment was a peer-assessment activity, a teacher's strategy for placing the work in the hands of the students (Black et al., 2003). The students had already corrected a test and were familiar with the formation of the criteria, so they could assess the work of a fellow student and give advice in a non-threatening way.

Description and objectives of the activity

Each student got a fellow student's answer sheet by random selection in order to feel comfortable with it and not to be affected by the other student's academic performance.

Initially each student worked individually on an exercise from the textbook with a complex shape. When the time was up, all worksheets were put in a box and each student chose randomly one of them. Subsequently, the correct solutions were presented by students on the whiteboard and discussed with the whole class. In addition, the sides of each shape were drawn with different colors as a differentiated representation of the exercise, a visual aid. Finally, the students had to correct the fellow student's answer sheet using criteria and to provide adequate feedback.

The aim of the activity was for the teacher to involve students actively in the process and, at the same time, to see how well the students do in identifying shapes and elements needed to calculate the area after having received feedback for the first activity. For the students the exercises were an opportunity to test the depth of their knowledge and their ability to form assessment criteria as well as to promote their comprehension of shapes' areas.

Highlights of the Activity:

(1) Students became actively involved, developed strategies and made suggestions to fellow students.

The following examples give an idea of the work done:

- A student wrote the following comment suggesting a strategy: "You must study the formulas. Also, it might help you if you split (and supplement) the shapes."
- Some students found the requested area by adding or subtracting areas of basic 2-Dimensional shapes (such as squares, rectangles or triangles) and by doing so some of them found a way to overcome their difficulties in calculating area of a parallelogram or a trapezium.
- Students liked the assessment of other students' work, the fact that their mistakes were made by classmates too and the opportunity to offer feedback. They felt that the process helped them learn. A student's response to an evaluative for the activity question was: "The process of correcting another person's worksheet helped me a lot to figure out how to correct small but important mistakes."
- Students' feedback comments reflected on their individual learning. They saw more clearly what they had to do in order to clarify concepts, to tighten their

learning. “Explain what you do when you do not use the formula. Try to adjust your time. Avoid multiplication errors.”

(2) The teacher’s suggestion to color each shape, in order to see it clearly and easily distinguish one from another, proved to be very helpful for many students and was embraced by the whole class.

A student wrote: “The tip with the colored shapes was the most helpful to me. I could locate the height and the base in order to calculate the area.”

3rd Activity of Formative Assessment

The third activity of formative assessment was a self-assessment activity, a tool for promoting learning autonomy (James, 2017).

Description and objectives of the activity

The students worked individually on a worksheet at their own pace with minimum help. The worksheet contained three exercises designed by the teacher after taking in consideration the two previous activities and the persisting errors. The students had to replace the correct elements in the area formula, verify the results, develop criteria to assess their work and decide on the next learning steps. As Miranda and Hermann (2015) suggested, the information gathered from a formative task can be used to address students’ errors and misunderstandings by changing the instructions. Thus, the teacher adjusted the time and modified the instructions accordingly.

Highlights of the Activity:

(1) Some students continued to confuse the shorter side of a parallelogram with the distance between its parallel sides.

The teacher had to adjust the activities that followed, so that those students practiced by constructing altitudes, which is in line with the strategies proposed by Chappuis (2015). The students worked on a worksheet containing triangles, trapeziums and parallelograms in different positions and orientation. An interactive digital application (in Geogebra) was recommended for the students to test if a height corresponds to a particular side, in order to achieve understanding.

(2) From their comments it was clear that the students developed their critical and evaluative capacities, deepened their understanding of the subject, tightened their learning and gained confidence.

A student wrote: “Yes, I can assess the work I have done. At first, I check the formula, if it is the wright one. Then I check the operations and verify the result. I examine the way I have presented the solution. I go through the same exercise again until I figure it out and solve it correctly.”

As a result, the students’ grades improved in the test that followed the activities (the class did better than other classes where the activity was not implemented). Also, students became accustomed to using control techniques. An example of that was the comment of a student during the following lessons in Pythagorean Theorem: “So, when looking for a vertical side, I have to find a number less than the length of the hypotenuse.”

In addition, the teacher adjusted an activity to check (again) upon the students understanding of the subject (Fisher & Frey, 2014a). The students had to create their own problem involving shape area (with at least two different shapes) and they were allowed to work in groups. The aim of the task was to check for understanding, but also to enhance skills such as creativity and critical thinking that are useful in learning.

Summarizing

Three formative assessment activities, part of a teaching scenario in the context of FORMAS implementation in Greece, are presented in this paper. The tasks aimed to help students understand the concept of area of 2-Dimensional shapes and promote a deeper level of learning.

As William (2000) suggested, using formative assessment activities the teacher focused

(1) on her role, by identifying needs, setting goals, asking appropriate questions, giving effective feedback, adjusting the activities when needed, supporting the students with differentiated instructions, and

(2) on the learners’ role, by sharing assessment criteria with them, asking them to create their own criteria during peer and self-assessment activities and encouraging them to monitor their own progress towards the learning goals.

References

Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2003). *Assessment for*

Learning: Putting it into practice. Maidenhead: Open University Press.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-71. Doi:10.1080/0969595980050102

Chappuis, J. (2015). *Seven strategies of assessment for learning* (2nd ed). Upper Saddle River, N.J: Pearson Education.

Earl, L., & Katz, S. (2006). *Rethinking Classroom Assessment with Purpose in Mind: Assessment for Learning, Assessment as Learning, Assessment of Learning.* Manitoba Education, Citizenship and Youth. Retrieved April 20, 2021, from: https://www.edu.gov.mb.ca/k12/assess/index.html

Fisher, D., & Frey, N. (2014a). *Checking for understanding: Formative assessment techniques for your classroom* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

James, M. (2017). Embedding Formative Assessment in Classroom Practice. Doi:10.1007/978-981-10-3654-5_31. Retrieved April 20, 2021, from: https://www.researchgate.net/publication/316733898

Miranda, R. J., & Hermann, R. S. (2015). Teaching in real time: Integrating continuous formative assessment into inquiry-based classroom instruction. *Science & Children, 53*(1), 80-85. Retrieved from ERIC database. (EJ1116136)

Tomlinson, C.A. (2014). *The Differentiated Classroom: Responding to the Needs of All Learners.* 2nd Edition, ASCD, Alexandria.

William, D. (2000). Formative Assessment in Mathematics, Part 3: The Learner's Role. *Equals: Mathematics and Special Educational Needs*. 6. 19-22. Retrieved April 20, 2021, from: https://www.researchgate.net/publication/258423288

My Experience in Applying Formative Assessment Strategies in my Classes: An Application in the Addition of Rational Numbers.

Aggela Ellina

Secondary Mathematics Teacher, Ministry of Education, Culture, Sport and Youth, Nicosia, Cyprus

Email: chrisangel188@hotmail.com

Abstract: During the time I participated in the programme "Promoting Formative Assessment: From Theory to Policy and Practice-FORMAS," I tried to implement different courses of action for formative assessment in my classroom. This paper describes my experience in applying formative assessment strategies in my classes. It also describes an application in the addition of rational numbers which took place in one of my classes which consisted of 21 thirteen-year-old students in December 2020. The goals were to apply different assessment techniques in my lesson (a real-life problem, math tiles, the number line, rules) to help my students understand the addition of real numbers, to determine assessment criteria for the assessment activities and to develop improvement strategies and personal action plans with a variety of ways and procedures including checklists. In addition, I gave constructive feedback to my students and at the same time I kept a portfolio to provide feedback to them. These different courses of action resulted in encouraging the students to work with more appetite and enthusiasm, enhancing the students' participation in the assessment process, improving the students' learning abilities, showing what students know and what they can do using a variety of ways and procedures and improving and acquiring the expected learning outcomes.

Introduction

Black and William (1998) define assessment broadly to include all activities that teachers and students undertake to get information that can be used diagnostically to alter teaching and learning. Under this definition, assessment encompasses teacher observation, classroom discussion, and analysis of student work, including homework and tests.

Students' assessment is a continuous process which constitutes an integral part of effective teaching. It includes:

(1) Developing and selecting appropriate assessment tools and strategies
(2) Implementing assessment strategies
(3) Using assessment tools
(4) Recording assessment results
(5) Analysing, interpreting and utilising the assessment results
(6) Reporting assessment results to the students and their parent

The main purposes of students' assessment are:

- to provide information to the teachers as well as to the parents regarding the level of a student in relation to the other students (comparative assessment).
- to help teachers assess the effectiveness of their teaching
- to support teachers in identifying students' learning needs in order to help them improve their learning outcomes (formative assessment)
- to provide information about a school's performance (evaluative assessment)

The assessment becomes formative when the information is used to adapt teaching and learning to meet student needs. When teachers know how students are progressing and where they are having trouble, they can use this information to make necessary instructional adjustments, such as reteaching, trying alternative instructional approaches, or offering more opportunities for practice. These activities can lead to improved student success (Boston, 2002). For example, a teacher can use the results of a written assessment to identify and to handle students' learning needs in an attempt to help them improve their learning outcomes.

Formative assessment is ongoing and is expected to be used often so as to ensure that learning needs are identified early enough and that the necessary corrective actions are taken.

Different Assessment Strategies

Learning is multidimensional and it cannot be measured adequately with only one strategy. Consequently, the teacher needs to implement a variety of assessment strategies in an attempt to give students the opportunity to show their knowledge and skills using different ways and procedures. The choice of an assessment strategy depends on the learning goal to be assessed, since different assessment strategies work best for different learning goals.

Research on teacher effectiveness has shown that successful teachers tend to be those who are able to use a range of teaching strategies and who use a range of interaction styles, rather than a single, rigid approach to teaching and learning (Darling-Hammond 2000).

The use of assessment strategies not only provides a structure that enhances teaching and learning, but makes that learning explicit to the teacher and students in a way that suggests assessment may provide a pathway to enhance the participants' perspectives on their mathematical education experiences (Ní Chróinín & Cosgrave, 2013).

Applying formative assessment strategies in my classes

This paper refers to my experience in applying formative assessment strategies in my classes and shows an application in the addition of rational numbers.

I have decided to take part in the FORMAS programme because I considered it to be a great opportunity which would help me improve the quality of the assessment of my students and their learning results.

In the beginning of the programme, I examined my teaching methods in relation to the way I evaluated my students. I realised that I needed to develop improvement strategies and personal action plans, to apply formative assessment to my lesson and to give my students constructive feedback.

Along the way, I encountered several difficulties while applying my action plans. On the one hand, a big obstacle was the time required to implement these action plans, which was not a part of the time allocated for the implementation of the curriculum scheduled in the Comprehensive Syllabus.

On the other hand, my students were not at all familiar with these types of assessment and they were facing my new assessment strategies for the first time. It took me quite a while to explain to them why and how we were going to work. Then, when my students got used to this new way of conducting a lesson, they worked with an appetite and enthusiasm. That was when I started to see their learning abilities change for the best.

During this programme, I implemented different courses of action for formative assessment in my classroom (Figure 1):

(1) Different assessment techniques.
(2) Determination of assessment criteria.
(3) Using a variety of ways and procedures to assess my students' attainment.
(4) Giving constructive feedback to the students.
(5) Keeping a portfolio.

Figure 1. Action in my classes.

(1) Different assessment techniques

Considering the learning goals, which I set up for each lesson, I managed to use different assessment techniques in my teaching methods efficiently and systematically: written, oral as well as performance assessment.

After that, I recorded the results and identified the advantages and disadvantages of these techniques. For their implementation, I used presentations, oral questions, performance activities and projects.

An application in the addition of rational numbers.

For example, in the addition of rational numbers, I assessed the learning achievement of my students using the following assessment techniques:

(a) <u>*Orally*</u>

Starting with a problem from real life, my students began to think of a solution and verbally describe it to me (Figure 2).

For example,

Mary had €2 in her account. Using her visa card, she bought goods worth €5 from the supermarket in her neighbourhood. What is the current balance of her bank account?

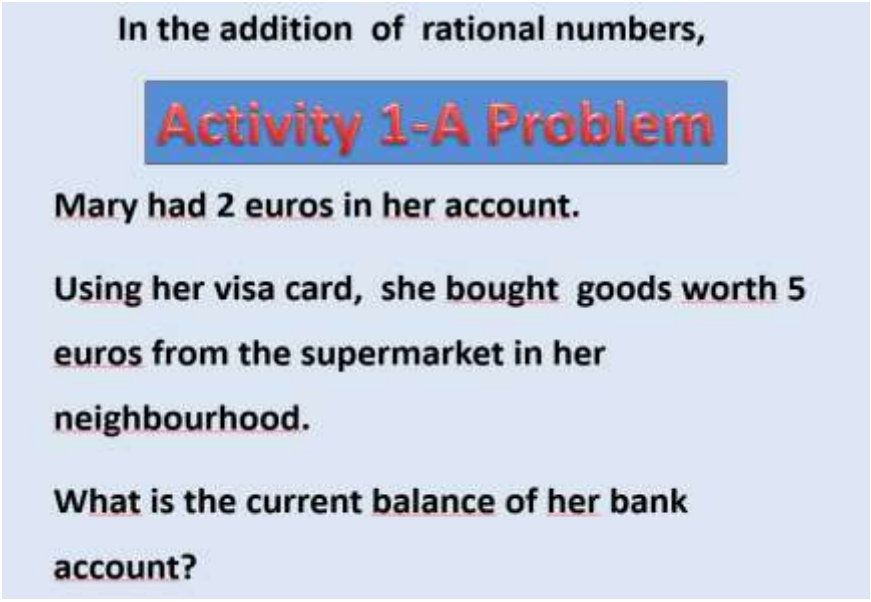


Figure 2. Using a real-life problem

(b) Using math tiles, (+) and (-)

I gave each student math tiles, considering the (+) tile to represent a positive number and the (-) tile to represent a negative number. Knowing that when a (+) tile meets a (-) tile, they neutralise each other, I asked my students to present the solution of the above problem using the math tiles they had in front of them and then to draw the answer in their notebook (Figure 3).

So, 2 positive tiles and 5 negative tiles, give us 3 negative tiles.

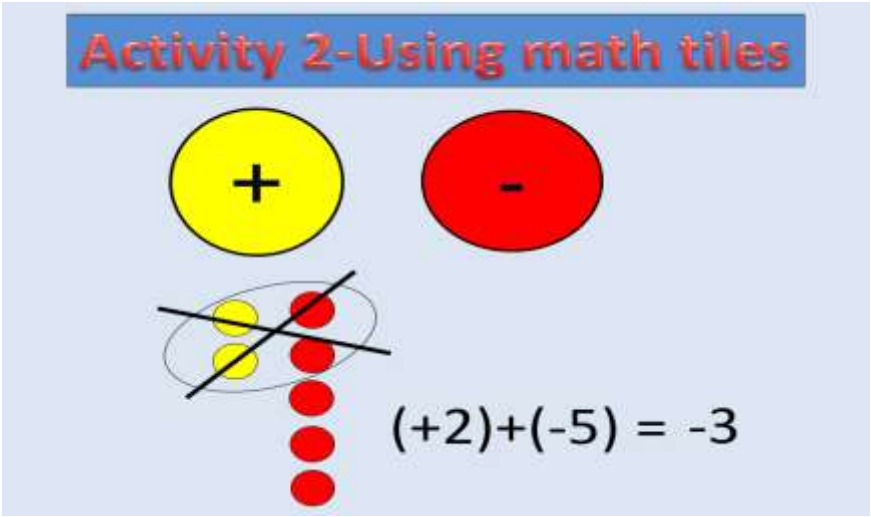


Figure 3. Using math tiles

(c) Using the number line

Next, I asked my students to represent the above mathematical relationship on the number line as well (Figure 4).

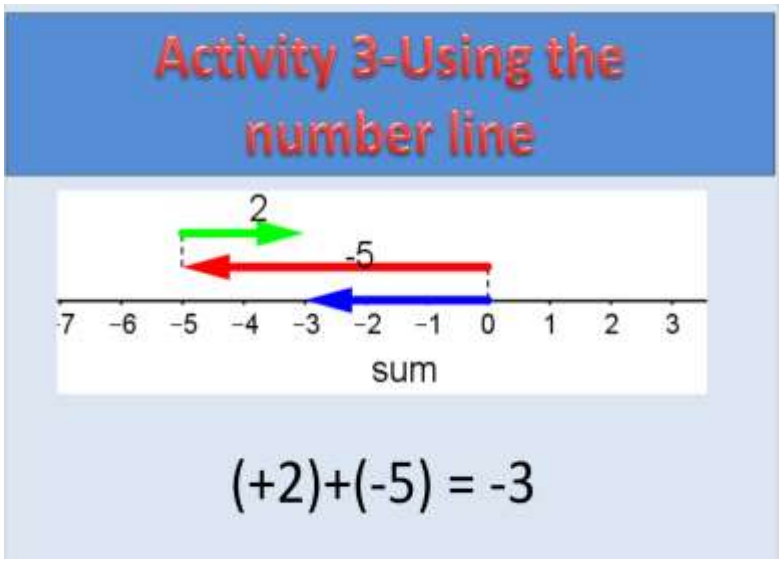


Figure 4. Using the number line

(d) Using rules

After covering all the possible cases with the ways previously mentioned, my students discovered the rules of the addition of rational numbers on their own. To determine the degree to which they acquired this knowledge, I gave each of my students a small whiteboard. Then, I wrote an addition of rational numbers operation on the board, asking my students to solve it on their small whiteboard and raise it towards me to see whether each student was in a position to answer correctly (Figure 5).

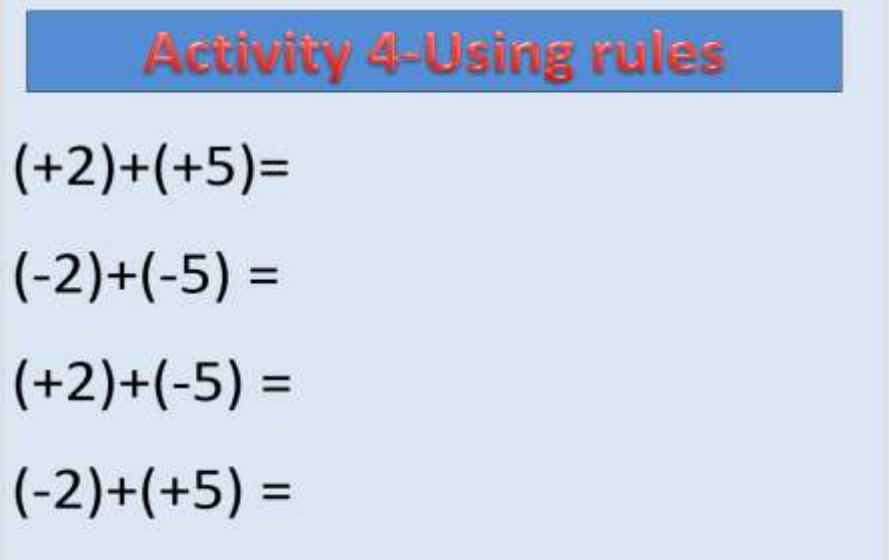


Figure 5. Using rules

Therefore, my students had several opportunities to show what they know and what they can do using different representations (Figure 6)

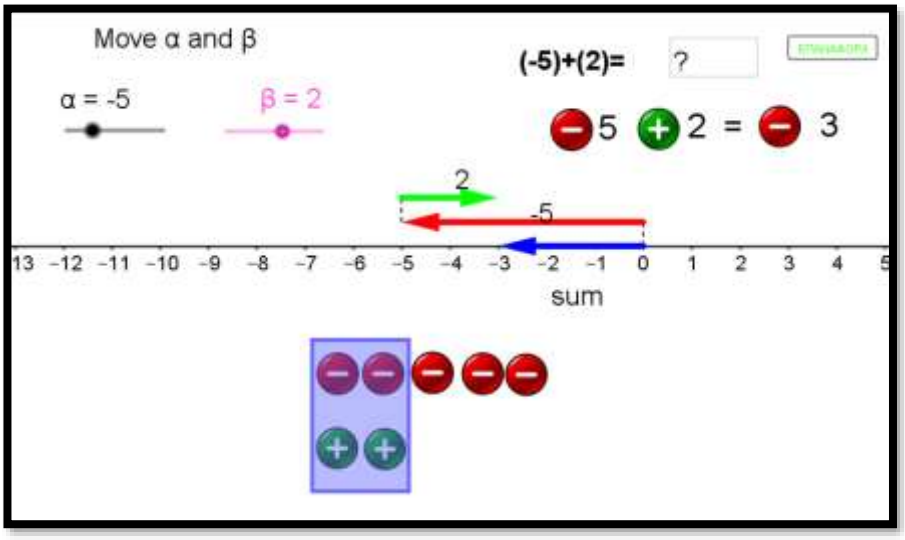


Figure 6. Using different assessment techniques

(2) Determination of assessment criteria

From the programme's training, I learned to determine assessment criteria for the assessment activities, either by myself or with the help of my students.

Moreover, I prepared checklists wherever possible in order to register the outcomes resulting from the different assessment techniques that I applied and to understand what my students know and what they can do in relation to the expected learning outcomes. The checklists made it easier to inform the students and their parents of the results.

I feel that I would like to use the rubrics as well, but unfortunately, there was not enough time for implementing them in my classrooms. I hope I will be able to do this at a later stage.

(3) Using a variety of ways and procedures to assess my students' attainment

My students' attainment was assessed using a variety of ways and procedures:
I presented the process to my students and asked them to define the assessment criteria. Furthermore, I presented them with activities at different stages to help them determine how an activity evolves. I also assigned my students the assessment of their work according to the assessment criteria we had already established.

The above actions regarding the involvement of my students in the assessment process resulted in the enhancement of their participation in this process.

(4) Giving constructive feedback to the students

I gave constructive feedback to my students using appropriate mathematical language, to help them improve and acquire the expected learning outcomes.

(5) Keeping a portfolio

I kept a portfolio in which I noted comments and observations in relation to the methods used to provide feedback to the students. These included the use of different assessment techniques and ways in which the students used the defined assessment criteria.

Conclusion

Formative assessment is tightly linked with instructional practices. The teacher needs to consider how his/her classroom activities, assignments, and tests support learning aims and allow students to communicate what they know, to give constructive feedback to the students, to make the necessary adjustments to his/her lessons, and then use this information to improve teaching and learning.

The determination of assessment criteria for the assessment activities, the use of a variety of ways and procedures and the constructive feedback that the educator gives to the students, help the students to work with an appetite and enthusiasm, enhance the students' participation in the assessment process, help to see the students' learning abilities change for the best, show what students know and what they can do using a variety of ways and procedures and improve the expected learning outcomes of the students.

Acknowledgements

I would like to mention that after the application of the action plans, all us educators made as a team, we met to discuss our experiences in the scheduled meetings of the programme.

With the help of our educationists: Leonidas Kyriakides, Yiannis Ioannou and Margarita Christoforidou, who I would like to thank very much, we exchanged views and made suggestions for the improvement of our action plans.
I truly feel very fortunate that I was given the opportunity to participate in this programme. I believe I have learnt a lot and I hope there are more beneficial programmes like this for us, active educators, to participate.

References

Black, P. & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *The Phi Delta Kappan, 80*(2), 139-148. https://www.jstor.org/stable/20439383

Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research and Evaluation, 8*(9).
https://doi.org/10.7275/kmcq-dj31

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1). https://doi.org/10.14507/epaa.v8n1.2000

Ní Chróinín, D. & Cosgrave, C. (2013). Implementing formative assessment in primary physical education: Teacher perspectives and experiences. *Physical Education and Sport Pedagogy, 18*(2), 219-233. http://dx.doi.org/10.1080/17408989.2012.666787

How I Create a Positive Learning Environment and Effectively Use Homework Assignments for Formative Assessment

Katerina Nicolaou

Secondary Mathematics Teacher, Ministry of Education, Culture, Sport and Youth, Nicosia, Cyprus

Email: kanicolaou@hotmail.com

Abstract: Students focusing more on grades rather than on learning has always been one of my concerns. So, when I found out about the program "Promoting Formative Assessment: From Theory to Policy and Practice" (FORMAS), I felt that it was an opportunity for me to learn more on how to help students shift their attention from grades to the learning process and as a result make my teaching, including my assessment methods, more effective. As we learned during our meetings, what is very important in effectively implementing formative assessment is to first create a positive learning environment. The key in achieving this is fostering a growth mindset. In the first of the two parts of this paper, I will describe two strategies that I use to foster a growth mindset in my students. Namely, the teaching of perseverance and the power of mistakes. In the second part of my paper, I will explain how to effectively use homework assignments for formative assessment. The aforementioned strategies did indeed result in some positive outcomes. Several students became more interested in learning, the number of students turning in their own work substantially increased, students were more willing to participate even if they did not know the answer and more students were not afraid of making mistakes or asking questions. In conclusion, the FORMAS program gave me a chance to reflect on my teaching and assessment methods and, consequently, to make my teaching more effective. The emphasis given on student assessment during the learning process helped students become more interested in learning and hence to improve their learning outcomes.

Introduction

How many times do we teachers hear: "How can I get an A?", "What can I do to improve my grade?", "What will my final grade be?", "My parents will punish me if I don't get a high enough grade." Unfortunately, way too often. Students focusing more on grades rather than on learning has always been one of my concerns. Afterall, historically, grades were not created with the students in mind. They were created to ease communication between institutions. So, it is not surprising that, rather than motivating students to learn, grades can dampen existing intrinsic motivation, give rise to extrinsic motivation, enhance fear of failure, reduce interest, decrease enjoyment in class work, increase anxiety, hamper performance on follow-up tasks, stimulate avoidance of challenging tasks, and heighten competitiveness (Schinske & Tanner, 2017).

When I found out about the program "Promoting Formative Assessment: From Theory to Policy and Practice" (FORMAS), I felt that it was an opportunity for me to learn more on how to help students shift their attention from grades to the learning process and as a result make my teaching, including my assessment methods, more effective. It turns out that I was right. As Sheppard (2005) also argues, one of the goals of formative assessment is to counteract students' obsession with grades and to redirect interest and effort toward learning.

According to Black and William (2009), "practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited" (p. 7). Moreover, formative assessment consists of five key strategies (Black & William, 2009):

(1) clarifying and sharing learning intentions and criteria for success;
(2) engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding;
(3) providing feedback that moves learners forward;
(4) activating students as instructional resources for one another; and
(5) activating students as the owners of their own learning.

For the above strategies to be successful, it is important to maintain students' willingness to stay focused on their learning goals and persist in the face of difficulty. This depends on their awareness of and access to volitional strategies (metacognitive knowledge to interpret strategy failure and knowledge of how to buckle down to work) (Black & William, 2009). In other words, to effectively implement formative assessment, it is essential to foster a growth mindset in our students. As Dweck (2016), a Stanford psychologist and lead researcher on the topic, states "In a growth mindset, people believe that their most basic abilities can be developed through dedication and hard work – brains and talent are just the starting point. This view creates a love of learning and a resilience that is essential for great accomplishment."

In the first of the two parts of this paper, I will describe two strategies that I use to foster a growth mindset in my students and hence create a positive learning environment. Namely, the teaching of:

(1) perseverance; and

(2) the power of mistakes.

In the second part of the paper, I will explain how I effectively use homework assignments for formative assessment. According to Marsh (2007), formative assessment involves providing "useful" feedback on tests and homework. That is, rather than a grade, information is provided about specific errors and suggestions for improvement. This encourages students to focus their attention thoughtfully on the task rather than on simply getting the right answer.

How I create a positive learning environment

I begin with a 40-minute lesson in which I talk about perseverance and the power of mistakes using quotes, video and examples of famous people accompanied of course with discussion. Thereafter, I remind my students of these strategies whenever I feel it is necessary.

Perseverance

To teach perseverance, the first thing I tell my students is that what matters the most is to try their best and to learn. But they must bear in mind that learning takes time and effort. Even Einstein once said, "It's not that I'm so smart, it's just that I stay with a problem longer."

So, it helps to have a positive attitude and to not give up when things get hard. My favorite in getting this point across is a video I present to the students about two frogs who fell into a large vessel full of milk. The frogs tried hard to get out but could not. After some time one of the frogs became very tired and he gave up the hope of escape. He slowly sank to the bottom and drowned. Whereas the other frog did not lose heart, he kept on swimming. To his surprise, his motions slowly turned the milk into thick butter. He got delighted and tried even harder. Soon the frog was able to climb out of the vessel and jumped to safety (Infobells, 2019).

After explaining that learning requires persistence, we move on to explain that learning is a process. When a new concept is introduced in class, it is normal not to understand everything right away, but this does not mean that they never will. Provided that they listen carefully in class, study the class notes again at home, solve the in-class examples again, do their homework, ask questions and practice. Quoting Paul Halmos, "the only way to LEARN mathematics is TO DO mathematics."

The Power of mistakes

This brings us to the second strategy that I use to foster a growth mindset. The teaching of the power of mistakes. Because when you do mathematics, you make mistakes. They are an integral part of the learning process. In Einstein's words, "a person who never made a mistake never tried anything new."

There are numerous examples of famous people who experienced massive failure at one point in their lives. Like, Thomas Edison (who as a child was thought to be dumb and was told by many of his teachers that he would never be a success), Michael Jordan (who was not able to stay on his high school team), Elvis Presley (who failed his music classes and whose first manager told him that he should go back to driving a truck) and the list goes on.

So, the secret lies not in avoiding mistakes, but in how we react to them and how we use them to learn. Mistakes can be viewed as the stepping stones to success. As Thomas Edison nicely put it "I have not failed 700 times. I have not failed once. I have succeeded in proving that those 700 ways will not work. When I have eliminated the ways that will not work, I will find the way that will work."

How I effectively use homework assignments for formative assessment

As already mentioned, homework assignments, if appropriately used, can be an extremely valuable tool in formative assessment.

My first step is to explain to the students the importance of homework. First and foremost, it is a means to achieving their goal to learn. In addition, it gives them the practice that we keep emphasizing over and over again, it improves their understanding and it helps them to retain this understanding.

Having the aforementioned in mind, I do not grade homework assignments. I only return them with constructive feedback. I tell my students that "I know that you can have someone else do the assignment for you, or copy it from one of your classmates, but this will not help you learn. Because it will not give me the opportunity to see where you have difficulties and how I can help you."

This is the reason why it is important to always remember to praise the effort. As mentioned by Shepard (2005), feedback that focuses on a student's level of effort, evidence of alternative reasoning strategies used, and the specifics of work products fosters incremental beliefs about ability and results in more constructive behavior in the

face of learning obstacles. Thus, praise focused on effort and strategies increases students' resilience and learning.

To be formative, assessment insights must be used immediately as part of the instructional process (Shepard, 2005). Technology turns out to be extremely important in assisting me to provide prompt feedback. Particularly, I ask the students to turn in daily assignments electronically. This gives me the chance to look at most of the students' work before going over it in class. In this way, I know what to give emphasis on during our next meeting. In addition, I do not only ask students who solved an exercise correctly to present the solution in class, but also students who had difficulty with that exercise. In this way, we will all (teacher and students) give them the help they need. I even have students tell me "Can I attempt to do this exercise because I had difficulty with it and I want to understand it."

Results

The strategies described in this paper to create a positive learning environment and to effectively use homework assignments for formative assessment did indeed result in some positive outcomes. More precisely, several students became more interested in learning, the number of students turning in their own work substantially increased, students were more willing to participate even if they did not know the answer and, lastly, more students were not afraid of making mistakes or asking questions.

Of course, I did encounter some difficulties along the way. For instance, not all students submitted their homework and it was quite time-consuming to assess the students' assignments on a daily basis.

Conclusion

In conclusion, the FORMAS program gave me a chance to reflect on my teaching and assessment methods and, consequently, to make my teaching more effective. The emphasis given on student assessment during the learning process resulted in students becoming more interested in learning and hence improving their learning outcomes.

Acknowledgements

I wish to extend my special thanks to our educationalists (Leonidas Kyriakides, Yiannis Ioannou and Margarita Christoforidou) for giving us the chance to participate in this stimulating learning experience.

References

Black, P. & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31. https://doi.org/10.1007/s11092-008-9068-5

Dweck, C. S. (2016). *Mindset: The new psychology of success.* Ballantine.

Infobells. (2019, March 21). *The Resourceful Frog Moral Story/Bedtime Stories for Kids/Infobells.* [Video]. YouTube. https://www.youtube.com/watch?v=xxFcXky8klc

Marsh, C.J. (2007). A critical analysis of the use of formative assessment in schools. *Educational Research for Policy and Practice, 6*(1), 25-29. https://doi.org/10.1007/s10671-007-9024-z

Schinske, J. & Tanner, K. (2014). Teaching more by grading less (or differently). *CBE-Life Sciences Education, 13*(2), 159-166.
https://doi.org/10.1187/cbe.cbe-14-03-0054

Shepard, L.A. (2005). Formative assessment: Caveat emptor. ETS Invitational Conference 2005. *The Future of Assessment: Shaping Teaching and Learning*, New York.
https://web.archive.org/web/20111007220347/http://www.cpre.org/ccii/images/stories/ccii_pdfs/shepard%20formative%20assessment%20caveat%20emptor.pdf

Searching for stages of teacher assessment skills to improve assessment practice: A European study

Margarita Christoforidou* & Jitske de Vries[b]

[a]*Department of Education, University of Cyprus, Nicosia, Cyprus;* [b] *Faculty of Behavioural, Management and Social Sciences, Department of ELAN, University of Twente, Enschede, The Netherlands*

* corresponding author

Email: margo.chr@gmail.com & j.a.devries@utwente.nl

Abstract: This paper reports the findings of the first phase of the Erasmus+KA3 project entitled *"Promoting Formative Assessment: From Theory to Policy and Practice (FORMAS)".* A framework for examining teachers' assessment skills is proposed. Based on this framework, a questionnaire for measuring assessment skills is developed. To examine the properties of the questionnaire, a validation study was conducted in the four participating countries (i.e., Cyprus, Greece, The Netherlands, and Belgium) in June 2019 with a sample of 574 math teachers. Data elicited are analysed using the Rasch and Saltus models. Based on the across country analysis, assessment skills are grouped into three types of behaviour which are discerned in a distinctive way and move gradually from skills associated with everyday assessment routines to more advanced skills. The developmental scale is consistently identified in both measurement periods of the second phase of the study (i.e., at the beginning and at the end of the intervention) which provides further support to the initial findings. Implications for the development of educational policy are drawn.

Introduction

Student assessment has long been recognized as an important element of teachers' practice (Black, 2015; Hayward, 2015; Hopfenbeck, 2018; Hopfenbeck & Stobart, 2015; Kingston & Nash, 2011; 2015; Panadero et.al., 2019). When students are assessed for formative purposes, assessment can have positive impact on students' learning (Hattie & Temperley, 2007; Herman et al., 2006; Kyriakides & Creemers, 2008; Kyriakides et al., 2021; Wiliam et al., 2004). At the same time, studies conducted in different countries reveal that although teachers appear to hold positive views towards formative assessment (Brown et al. 2019; Brown, 2008; Yates & Johnston, 2018), their practice is still focused on conducting assessment for summative purposes (Randel et al., 2016; Wiliam, 2017). These findings suggest that emphasis on how effective assessment practice can be achieved is required.

The European study reported here, is part of the Erasmus+KA3 project entitled *"Promoting Formative Assessment: From Theory to Policy and Practice (FORMAS)".* This 3-year project aims to generate guidelines on how to use assessment for formative purposes and how to develop teacher professional development (TPD) mechanisms to support teachers in the implementation of assessment for learning. The first phase of the project (i.e., reported here) has two major aims: a) to develop a theoretical framework for measuring teacher assessment, and b) to establish valid instruments to measure teachers' professional needs in student assessment. A description of the theoretical framework proposed is presented below. Then, the instrument developed (i.e., teacher questionnaire) to measure teachers' assessment skills is described and information on the validation study conducted is provided. Finally, the contribution of findings for the development of educational policy regarding formative assessment is discussed.

Theoretical framework

The framework developed is based on previous studies examining teacher assessment skills conducted in various European countries (see Christoforidou & Xirafidou, 2014; Christoforidou, Kyriakides). The proposed framework (see Figure 1) examines assessment looking at three main aspects.

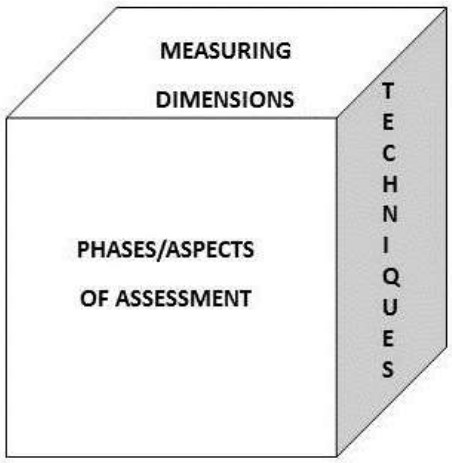


Figure 1. A framework for measuring teacher assessment skills

First, the dynamic nature of education is taken into account (Scheerens, 2016) and skills associated with the main phases of the assessment process are examined (i.e., *Constructing/ Selecting Assessment Tools/ Processes, Administering Assessment*

Tools/Processes, Recording Assessment Results, Analyzing, Interpreting and Using Assessment Results, Reporting Results to Intended Users). Second, assessment skills are defined and measured in relation to teachers' ability to use the main assessment techniques (i.e., written assessment, oral assessment, performance assessment) as well as their skills to implement not only teacher but also self-, peer- and co- assessment. Finally, the measurement framework suggested in the *Dynamic Model of Educational Effectiveness* (Creemers & Kyriakides, 2008) is adopted, allowing us to examine both quantitative and qualitative characteristics of the assessment process (i.e., frequency, focus, stage, quality, and differentiation). For a more detailed presentation of the framework please see *Output 2: A Comprehensive Framework for Measuring Teacher Assessment Skills* of the FORMAS project.

Developing an instrument to measure assessment skills

By considering the theoretical framework, a teacher questionnaire was developed. The questionnaire consisted of 119 items, designed to measure teachers' assessment skills in mathematics across the three aspects of the theoretical framework (i.e., phases of assessment, techniques of assessment, measurement dimensions). Each assessment technique was examined in relation to the five phases of the assessment process and for each phase of the assessment process, each of the five measurement dimensions was applied. A five-point Likert scale (1 to 5) was used, and teachers were asked to indicate the extent to which they behave in a certain way during mathematics teaching in their classroom. For example, an item asked teachers to indicate from 1 to 5, if they construct items/exercises/questions for a written test considering their students' abilities. As you can see, this item refers to the first phase of the assessment process *Constructing/ Selecting Assessment Tools/ Processes,* in relation to the technique of *written assessment*. The dimension examined is *differentiation*. Similarly, another item asked teachers to indicated from a scale 1 to 5, if they orally assess students to check to what extent the results correspond to the results of the written test. This item refers to the fourth phase of the assessment process *Analyzing, interpreting, and using assessment results* and it examines it in relation to the technique of *oral assessment*. The dimension examined here is *quality*, since it examines if teachers use multiple sources to validate assessment results.

The questionnaire was initially developed in English and a double translation process was used to develop the Greek and the Dutch versions of the instrument. A

validation study of the teacher questionnaire took place in the four participating countries (i.e., Cyprus, Greece, The Netherlands, and Belgium) in June 2019. Table 1 shows the sample obtained from each country.

Country	Number of Teachers
Belgium	104
Cyprus	188
Greece	96
The Netherlands	186
Total	574

Table 1. Validation study teacher sample

Methodology

Data elicited from the administration of the teacher questionnaire were analyzed using the extended logistic model of Rasch and the Saltus model to examine the scaling and developmental structure of teachers' abilities in assessment. Within and across countries analyses were performed. However, in this paper only the results of the across countries analysis are presented. It is important to mention that some items were removed to ensure a better fit of the model. The revised version of the questionnaire included a total of 115 items.

The results of the various approaches used to test the fitting of the Rasch model to our data revealed that there was a good fit to the model when teachers' performance in these assessment skills was analysed. Specifically, all teaching skills were found to have item infit with the range 0.83–1.18, and item outfit with the range of 0.79–1.36. In addition, all the values of infit t for both persons and assessment skills were greater than - 2.00 and smaller than 2.00. The procedure for detecting pattern clustering in measurement designs developed by Marcoulides & Drezner (1999) was used to find out whether assessment skills are grouped into levels of difficulty that may be taken to stand for types of teacher behaviour which move from relatively easy to more difficult. By applying this method to segment assessment skills based on their difficulties that emerged from the Rasch model it was found out that they are optimally grouped into three clusters. Furthermore, by conducting an across-country analysis, the three-cluster solution emerged as the best solution especially since this solution was able to explain 69% of the total variance. The Saltus model was then used to specify the developmental structure of

the skills. The results provided support to the scaling and developmental structure of teachers' abilities in assessment. In fact, the Saltus solution was found to represent a better fit to the actual data rather than the Rasch model and offered a statistically significant improvement over the Rasch model which is equal to 391,6 chi-squared units at the cost of 12 additional parameters (i.e., 4 t values, three means, three standard deviations, and two independent proportions). Table 2 shows the Rasch and Saltus parameter estimates for teachers' assessment skills.

Table 2. Rasch and Saltus parameter estimates for teachers' assessment skills

Teachers' Assessment Skills	**Rasch**	**Implied within-stage difficulty (Saltus)**		
	All	**Level 1**	**Level 2**	**Level 3**
Freq Construction Written	-2,34	-3,27	-3,27	-3,27
Stage Construction Written	-2,32	-3,25	-3,25	-3,25
Freq Administration Written	-2,31	-3,22	-3,22	-3,22
Stage Administration Written	-2,29	-3,18	-3,18	-3,18
Focus Construction Written	-2,28	-3,16	-3,16	-3,16
Freq Reporting Written	-2,27	-3,14	-3,14	-3,14
Freq Analysis Written	-2,26	-3,11	-3,11	-3,11
Stage Reporting Written	-2,25	-3,09	-3,09	-3,09
Freq Recording Written	-2,24	-3,06	-3,06	-3,06
Stage Recording Written	-2,23	-3,05	-3,05	-3,05
Freq Administration Oral	-2,22	-3,01	-3,01	-3,01
Focus Recording Written	-2,21	-2,98	-2,98	-2,98
Stage Analysis Written	-2,19	-2,95	-2,95	-2,95
Freq Construction Oral	-2,17	-2,93	-2,93	-2,93
Freq Recording Oral	-2,15	-2,91	-2,91	-2,91
Stage Recording Oral	-2,14	-2,88	-2,88	-2,88
Focus Administration Written	-2,13	-2,85	-2,85	-2,85
Quality Construction Written	-1,33	-0,83	-2,77	-2,79
Quality Administration Written	-1,31	-0,81	-2,73	-2,76
Frequency Reporting Oral	-1,28	-0,78	-2,75	-2,71
Freq Construction Performance	-1,26	-0,76	-2,69	-2,73
Freq Administration Performance	-1,25	-0,75	-2,66	-2,68
Focus Reporting Written	-1,24	-0,73	-2,61	-2,66
Quality Recording Written	-1,22	-0,71	-2,62	-2,64
Quality Reporting Written	-1,22	-0,69	-2,59	-2,58
Stage administration performance	-1,21	-0,67	-2,57	-2,55
Focus analysis written	-1,19	-0,66	-2,55	-2,51
Frequency analysis oral	-1,18	-0,62	-2,51	-2,49
Quality analysis written	-1,13	-0,61	-2,48	-2,47
Stage administration oral	-1,09	-0,59	-2,45	-2,46
Stage analysis oral	-1,08	-0,57	-2,41	-2,44
Focus administration oral	-1,05	-0,55	-2,38	-2,39
Focus analysis oral	-1,01	-0,52	-2,36	-2,37

Quality administration oral	-0,98	-0,49	-2,32	-2,34
Stage construction oral	-0,95	-0,51	-2,29	-2,28
Focus construction oral	-0,93	-0,47	-2,25	-2,26
Stage reporting oral	-0,91	-0,42	-2,27	-2,21
Focus reporting oral	-0,88	-0,46	-2,16	-2,19
Focus recording oral	-0,86	-0,39	-2,12	-2,16
Freq construction peer	-0,84	-0,33	-2,11	-2,14
Freq administration peer	-0,81	-0,37	-2,03	-2,12
Freq administration self	-0,79	-0,35	-2,07	-2,11
Stage construction peer	-0,72	-0,29	-1,95	-2,09
Stage administration peer	-0,71	-0,21	-1,91	-2,07
Freq recording peer	-0,68	-0,26	-1,82	-2,04
Diff/tion administration written	-0,66	-0,24	-1,79	-1,95
Diff/tion administration oral	-0,65	-0,19	-1,84	-1,92
Freq analysis peer	-0,64	-0,12	-1,73	-1,88
Freq recording self	-0,63	-0,09	-1,67	-1,85
Freq reporting performance	-0,63	-0,08	-1,66	-1,85
Freq recording performance	-0,62	-0,07	-1,65	-1,84
Stage reporting performance	-0,61	-0,07	-1,65	-1,84
Stage recording performance	-0,61	-0,06	-1,64	-1,83
Freq analysis performance	-0,60	-0,06	-1,64	-1,83
Stage analysis performance	-0,60	-0,05	-1,63	-1,82
Quality administration perf/ce	0,73	1,19	0,11	-1,75
Quality recording performance	0,74	1,19	0,11	-1,76
Diff/tion administration perf/ce	0,75	1,20	0,12	-1,77
Quality construction performance	0,75	1,20	0,12	-1,77
Quality reporting performance	0,76	1,21	0,13	-1,78
Quality construction oral	0,76	1,21	0,13	-1,78
Quality reporting oral	0,79	1,27	0,16	-1,74
Quality analysis oral	0,81	1,23	0,18	-1,72
Freq reporting peer	0,83	1,25	0,19	-1,69
Stage recording per	0,84	1,34	0,22	-1,65
Quality recording oral	0,85	1,38	0,25	-1,63
Focus recording peer	0,86	1,41	0,27	-1,61
Diff/tion construction oral	0,87	1,42	0,29	-1,57
Diff/tion construction written	0,89	1,48	0,32	-1,54
Diff/tion analysis oral	0,92	1,51	0,35	-1,51
Diff/tion recording oral	0,95	1,55	0,38	-1,48
Diff/tion reporting oral	0,98	1,58	0,41	-1,45
Diff/tion reporting written	0,99	1,61	0,43	-1,44
Focus administration peer	1,03	1,64	0,49	-1,41
Diff/tion recoding written	1,05	1,68	0,51	-1,38
Quality administration peer	1,08	1,72	0,53	-1,36
Diff/tion analysis written	1,09	1,76	0,55	-1,34
Diff/tion administration peer	1,11	1,79	0,59	-1,31
Focus construction peer	1,14	1,81	0,61	-1,28
Focus administration self	1,16	1,86	0,63	-1,26
Quality construction peer	1,17	1,89	0,66	-1,22
Stage analysis peer	1,19	1,93	0,68	-1,19
Diff/tion construction peer	1,21	1,96	0,72	-1,16

Focus analysis peer	1,23	1,99	0,75	-1,15
Stage Reporting peer	1,25	2,03	0,79	-1,09
Focus Reporting peer	1,27	2,06	0,81	-1,07
Quality analysis peer	1,29	2,09	0,83	-1,05
Quality Reporting peer	1,31	2,12	0,86	-1,02
Stage Recording peer	1,32	2,17	0,89	-0,99
Diff/tion analysis peer	1,33	2,19	0,93	-0,96
Diff/tion reporting peer	1,35	2,23	0,96	-0,94
Freq construction self	1,36	2,25	0,99	-0,91
Stage construction self	1,37	2,28	1,01	-0,88
Quality recording peer	1,38	2,32	1,03	-0,84
Focus construction self	1,39	2,36	1,06	-0,82
Diff/tion recording peer	1,41	2,38	1,08	-0,79
Stage administration self	1,42	2,42	1,12	-0,77
Quality construction self	1,43	2,44	1,15	-0,74
Quality administration self	1,44	2,48	1,19	-0,72
Freq analysis self	1,45	2,51	1,21	-0,69
Freq reporting self	1,46	2,56	1,23	-0,67
Focus analysis self	1,47	2,57	1,28	-0,63
Stage analysis self	1,49	2,61	1,32	-0,61
Stage recording self	1,51	2,65	1,35	-0,57
Diff/tion administration self	1,52	2,69	1,39	-0,55
Diff/tion construction self	1,53	2,72	1,41	-0,52
Quality recording self	1,54	2,76	1,44	-0,49
Quality reporting self	1,55	2,79	1,46	-0,47
Quality analysis self	1,57	2,81	1,48	-0,45
Diff/tion reporting self	1,58	2,85	1,53	-0,41
Diff/tion recording self	1,63	2,89	1,55	-0,39
Diff/tion analysis self	1,69	2,92	1,57	-0,37
Focus reporting self	1,75	2,99	1,59	-0,33
Stage reporting self	1,81	3,01	1,61	-0,29
Focus recording self	1,89	3,05	1,62	-0,26

As we can see from Table 2, the Rasch and Saltus analyses provided support for the classification of assessment skills based on their level of difficulty. This classification suggested the existence of three different groups of skills that move gradually from easier to more advanced skills (i.e., Stage 1, Stage 2, and Stage 3). A brief description of each stage is presented below.

Stage 1: Using mainly written assessment to measure achievement in mathematics for summative purposes.

The assessment skills included in this stage reveal that teachers demonstrating this type of behaviour usually use assessment for summative purposes. Even though it is possible that teachers acknowledge the importance of formative assessment, they have not yet

managed to establish a classroom culture that can successfully foster formative assessment practices. They usually use ready-made assessment tasks. The quality of their assessment practice needs improvement in term of its representativeness and of its internal and content validity. Oral and performance assessment are not systematically used to assess students' learning and assessment tasks used are mainly written. Regarding the written assessment tasks these are usually of the same type which raises questions about their concurrent validity. Homework is provided but it is not used for formative purposes (i.e., to identify students' learning needs to inform their teaching practice).

Stage 2: Using different techniques of assessment to measure achievement in mathematics but without defining appropriate success criteria and providing constructive feedback.

The assessment skills included in this stage reveal that, teachers demonstrating this type of behaviour give feedback to students about their learning and attempt to use assessment for formative purposes. However, the feedback provided is usually evaluative instead of constructive. In addition, teachers at this stage use different assessment techniques to assess students in mathematics but this is not done in a way that enables them to compare the results which emerge from the use of different types of assessment and in this way to test the internal validity of their assessment. In addition, they usually keep records of information elicited from only the written forms of assessment and do not systematically utilize records instruments such as checklists and rubrics. There is also space for improving their skills in formulating appropriate learning goals and criteria for success.

Stage 3: Using assessment techniques to measure specific and more complex educational objectives to provide constructive feedback but without involving students in the assessment process and differentiating their assessment practice. Teachers demonstrating this type of behaviour can use a variety of assessment techniques to measure students' learning and usually keep records of information elicited not only from written assessment but from other techniques as well. However, recording is usually not done in ways that facilitate the formative use of the information available. For example, they record information per exercise instead of per learning goal, thus making it difficult to draw conclusions of whether the student has achieve the goal or not. In addition, teachers at this stage assess group work but not in a systematic way and their assessment

is primarily concerned with the group's overall performance rather than with each student's contribution to the teamwork. Teachers situated at this stage have already established a culture that encourages students' involvement in the process of assessment. However, both peer and self-assessment for formative purposes are not yet systematically and efficiently introduced. In addition, teachers at this stage have not yet managed to introduce differentiated assessment practices in their teaching.

The stages identified were used to make decisions in relation to the content and design of the Teacher Professional Development (TPD) program that was implemented during the 2nd phase of the FORMAS project (i.e., intervention study). The developmental scale was consistently identified in both measurement periods (i.e., at the beginning and at the end of the intervention) which provided further support to the initial findings. In addition, by comparing the two measurements of teachers participating in the intervention it was observed that in the cases where change occurred, this change was towards the next demanding level (i.e., from stage 1 to stage 2, from stage 2 to stage 3 etc.). This stepwise movement further confirms the developmental character of the assessment skills examined.

Contribution of findings for educational policy development in formative assessment

The theoretical framework and the teacher questionnaire developed allowed us to examine teacher assessment behaviour and identify specific skills involved when assessing students' learning. As mentioned earlier, the FORMAS-project aimed to encourage policy makers to reform assessment policies and to establish teacher support mechanisms for the effective implementation of formative assessment. The developmental stages of assessment skills identified are expected to support the project's aims by providing a basis for theory-driven and evidence-based policy decisions. First, these findings can help policy makers to clarify what constitutes sound assessment practice. The critical policy analysis performed under the FORMAS-project showed that in participating countries no policies that require teachers to be knowledgeable and skilful in assessment are present (see FORMAS -Project's *Output 1: A Critical Review of National Policies on Formative Assessment*). Therefore, the stages can aid the definition of specific expectations to be met by teachers in relation to assessment. This will in turn,

help teachers understand how to implement effective assessment in their everyday teaching practice, identify possible shortcomings in their practice, while at the same time holding them accountable. Moreover, for teachers to shift their emphasis on effective assessment practices the necessary support should be provided. The proposed stages can be used to guide decision regarding initial teacher training and professional development. In this way, education offered to teachers both at pre- and in-service contexts will address the professional needs of specific groups of teachers each time supporting them to improve their assessment practice. The stages identified can also enable the identification of teachers' specific needs in assessment for appropriate corrective actions to take place. Thus, stages can be used to identify teachers' professional development needs per country and phase of education based on which decisions on the emphasis, content and duration of support can be made. Finally, the developmental stages identified can also be useful for formative evaluation purposes. Given that student assessment is recognized as an important factor of teacher effectiveness, stages can be used to guide the process of gathering data of a teacher's performance for formative evaluation purposes and provide a reliable and unbiased basis for decision making. At the same time, it is acknowledged, that using the suggested questionnaire to measure assessment skills in another context (i.e., country, subject, grade etc.) may yield different results both in terms of the numbers of stages identified and of the specific skills included in each stage.

Acknowledgements

This project, entitled "Promoting Formative Assessment: From Theory to Policy and Practice (FORMAS)" has been funded with support from the European Commission. This communication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

References

Black, P. (2015). Formative assessment–an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice, 22*(1), 161-177.

Brown, G. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York: Nova Science Publishers.

Brown, G.T., Gebril, A., & Michaelides, M.P. (2019). Teachers' Conceptions of Assessment: A Global Phenomenon or a Global Localism. *Frontiers in Education, 4,* 1-13.

Christoforidou, M., Kyriakides, L., Antoniou, P., & Creemers, B. P.M. (2014). Searching for stages of teacher's skills in assessment. *Studies in Educational Evaluation, 40,* 1-11.

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.

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

Hattie, J., & Temperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Hayward, L. (2015). Assessment is learning: The preposition vanishes. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 27-43.

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).

Hopfenbeck, T.N. (2018). Classroom assessment, pedagogy and learning–twenty years after Black and Wiliam 1998. *Assessment in Education*, *25*(6), 545-550.

Hopfenbeck, T.N., & Stobart, G. (2015). Large-scale implementation of assessment for learning. *Assessment in Education, Principles, Policy and Practice, 22*(1), 1–2.

Kingston, N.M., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*(4), 28–37.

Kingston, N. M., & Nash, B. (2015). Erratum. *Educational Measurement: Issues and Practice, 34*(1), 55.

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., Panayiotou, A., & Charalambous, E. (2021). *Quality and Equity in Education: Revisiting Theory and Research on Educational Effectiveness and Improvement.* London and New York: Routledge.

Marcoulides, G. A., & Drezner, Z. (1999). A procedure for detecting pattern clustering in measurement designs. *Objective measurement: Theory into practice, 5*, 261-277.

Panadero, E., Broadbent, J., Boud, D., & Lodge, J. M. (2019). Using formative assessment to influence self-and co-regulated learning: the role of evaluative judgement. *European Journal of Psychology of Education*, *34*(3), 535-557.

Randel, B., Apthorp, H., Beesley, A. D., Clark, T. F., & Wang, X. (2016). Impacts of professional development in classroom assessment on teacher and student outcomes. *The Journal of Educational Research*, *109*(5), 491-502.

Wiliam, D. (2017). Assessment for learning: meeting the challenge of implementation, *Assessment in Education: Principles, Policy & Practice, 25*(6), 686–689.

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.

Yates, A., & Johnston, M. (2018). The impact of school-based assessment for qualifications on teachers' conceptions of assessment. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 638-654.

The Implementation of Formative Assessment Strategies in the Mathematics Classroom in the Context of Professional Development Program

Danai Dafnopoulou

Department of Mathematics, National and Kapodistrian University of Athens, Athens, Greece

Email: danaidaf@math.uoa.gr

Abstract: This research aims to study the enactment of formative assessment practice in the mathematics classrooms by two teachers participating in a professional development program. Specifically, the way they implement formative strategies in their classrooms. Group meetings' observations of the program, classroom observations and personal semi-structured interviews with teachers produced the data, which were analysed following inductive methods based on principles of Grounded Theory. Results revealed that the emergent strategies mainly concerned students' engagement with assessment. Strategies that were part of the professional development material were also noticed.

Introduction

Assessment is not only a means of measuring the knowledge but also a basic element for supporting teaching (Black & William, 2003; Heritage, Kim, Vendlinski & Herman, 2008; Christoforidou, Kyriakides, Antoniou & Creemers, 2014). Formative is a type of assessment that differentiates from the summative and diagnostic types. Formative assessment is characterised as the one whose evidence is used to adjust teaching according to students' needs (Black & William, 1998). Different aspects of formative assessment have been the focus of research studies in education. Pointing to the need to create tools, since summative methods are not necessarily appropriate for directing learning (Black & Wiliam, 2003), multiple studies have focused on the analysis of quality strategies, their improvement in the classroom and on creating a unifying theoretical basis for them (Black, Harrison, Lee & Marshall, 2004; Wiliam, 2007; Black & Wiliam, 2009). Moreover, there has been a focus on how formative assessment has affected students. This kind of assessment has been characterised as a central factor of students' learning success (Christoforidou et al., 2014). More specifically, it has been studied how formative assessment strategies can influence students' development of mathematical writing communication skills (Santos & Semana, 2014).

Furthermore, the study of teachers' assessment skills indicated four types of behaviours, while it has been observed that teachers using more advanced types of assessment have greater results (Christoforidou et al., 2014). Moreover, it has been pointed out that formative assessment practice enactment can be affected from teachers' beliefs and attitudes about multiple educational issues, like learning (Marshall & Drummond, 2006) and the assessment (Brown, 2004).

In studies on assessment, the focus on the formative type has been neglected and the way formative assessment practice is enacted by mathematics teacher in secondary education is lacking. Thus, the aim of the study is to understand the formative assessment practice in the mathematics classroom, through the implementation of strategies from two mathematics teachers in secondary education in the context of a professional development program. Specifically, which formative assessment strategies are used and how they are implemented in the mathematics classroom.

Conceptual Framework

The notion of formative assessment was initially defined in relation with summative assessment, according to Bloom, Hastings & Madaus (1971), as a different evaluation type which would help all the participants (learners, teachers, curriculum makers) to improve what they wish to do (Black & Wiliam, 2003). Later in the literature, Sadler (1989) viewed formative assessment as a way of eliminating the randomness and inefficiency of learning as the assessment of students' answers can support and improve their learning through feedback. Feedback can provide information about the achievement of a learning goal, helping the gap between the current and the desirable learning goal to be recognised (Sadler, 1989). More recently, Black & Wiliam (2009), refer to the formative assessment as a practice in the classroom, that evidence about students' achievement are elicited, interpreted and used from teachers, students or their peers, in order for decisions to be made for the next teaching steps (pp. 9). Santos & Semana (2014), provided a definition that points out the learning goals and teacher's role on that. Specifically, they mention that assessment for learning or formative assessment, is characterised by everyday practices in the classroom that include elicitation and interpretation of evidence for students' learning and the use of this evidence for better decisions to be made about teaching, aiming to support students learning. These decisions are made mostly by teachers, but students can also be a part of this process.

Strategies regarding the achievement of formative assessment practice has been indicated in the literature. Informal methods for collecting evidence about students can be teacher's *spontaneous questions* regarding students solving method, observation, questions, dialogue, or students written *answers* (Ginsburg, 2009). Formal strategies regarding formative assessment were developed by Wiliam & Thomson (2007). They combined three basic processes of learning and teaching by Ramaprasad (1983), specifically where the student is in learning, where they want to go and what needs to be done to go there, with three different agents of those processes (teacher, learner, peer). They formulated five basic strategies that formative assessment can be perceived, that are presented in Table 1.

Table 1. Aspects of formative assessment (Black & William, 2009, pp. 8)

<table>
<tr><th></th><th>Where the student is going</th><th>Where the student is</th><th>How to go there</th></tr>
<tr><td>Teacher</td><td>1. Clarify the aims and success criteria</td><td>2. Effective classroom discussions and other learning tasks that elicit evidence about students understanding</td><td>3. Provide feedback that helps learners move forward</td></tr>
<tr><td>Peer</td><td>Understand the learning goals and success criteria</td><td colspan="2">4. Activate students as instructional resources for each other</td></tr>
<tr><td>Student</td><td>Understand the learning goals and success criteria</td><td colspan="2">4. Activate students as instructional resources for each other</td></tr>
</table>

These five strategies can be implemented in the classroom through multiple activities and tools. Specifically, the strategy of *clarification of the aims and success criteria* can be achieved by providing scripts to students. Scripts are written texts which describe learning goals that students are expected to achieve, as well as brief directions on where students should focus to reach those goals (Santos & Semama, 2014). Moreover, questions in the classroom are a way to facilitate the strategy of *effective classroom discussions* (Black & Wiliam, 2009). *Questions in the classroom* should be critical for student's development of understanding and aim to provide information teacher is searching for or issues that

students should think about (Black et al., 2004). Thus, students understand that learning is not about giving instantly the correct answer, but they should be ready to express their own understanding. Regarding the strategy of *providing feedback that helps students to improve*, feedback can be understood as the information provided by an agent, like the teacher, a peer, the book or a parent , regarding the performance or the understanding of a student (Hattie & Temperley 2007). It can also be achieved through *feedback only marking*, that can help students and parents to focus on learning issues and lead to student's reflection on those issues (Black & Wiliam, 2009). According to Santos & Pinto (2009) the form and length of written feedback have also been noticed to play a critical role for different students, while it has been observed that sort feedback has a positive contribution on students' focus on specific parts of a task. Student's strengths should be noted through feedback and directions on how to improve should be provided otherwise feedback has not a formative aim (Black et al., 2004; Wiliam, 2007). To achieve the strategies of *activating students as instructional resources for each other and for themselves* peer and self-assessment can be used (Black & Wiliam, 2009). Tools that are used for peer and self-assessment, specifically rubrics, may not be understood by students (Leahy, Lyon, Thompson, & Wiliam, 2005), for that reason assessment criteria should be transparent (Black & Wiliam, 2005), either by example through tasks or by simplified rubric forms or by students creating their own ones (Black et al., 2004). Finally, encouraging students to remember the aims and use them to assess their work can lead them to independently lead their work through assessment (Black et al., 2004).

Furthermore, Andersson (2017) suggested the addition of another component in the framework formulated by Wiliam and Thomson, which is the adjusted teacher instruction as a part of the strategy refers to *feedback that moves students forward*. This new categorisation occurred since the activity of adjusted instruction aims to improve learning even though it does not include only feedback (Andersson, 2017). Specifically, as Andersson (2017) explains, this activity can be achieved if summative results are used for picking or creating personalised tasks for different students.

Another activity that is related to the formative assessment strategies presented above is the *formative use of summative assessment*. Black et al. (2004) indicated that formative and summative assessment cannot be easily distinguished in practice. Moreover, they believe that the use of summative assessment for formative purposes can improve the classroom practice, as students can reflect on their work to schedule an

efficient revision. Finally, summative assessment results can be used by teachers to provide *feedback* to students (Black & Wiliam, 2009) and by clarifying the success criteria students can better understand the assessment (Black et al., 2004; Black & Wiliam, 2009).

Methods

This research is a qualitative research of two case studies (Baxter & Jack, 2008) of secondary mathematics teachers who, participated in the professional development program FORMAS was conducted. Firstly, the group that included teachers with some formative assessment skills was selected since it could provide a variety of data. From this specific group two mathematics teachers were selected according to their different work setting and due to their previous active engagement when participating to other professional development programs. Mary is a mathematics teacher at a public secondary school in Greece, teaching grades 8 and 9 (ages 13-15). John is a mathematics teacher at an experimental secondary school in Greece, teaching grade 9 (ages 14-15). Both teachers did not have previous training or formal institutional directions regarding formative assessment and their knowledge was mostly due to their personal searching, other professional programs not related to assessment or their previous collaboration with academics.

The context of the study was the professional development program FORMAS and specifically the group dedicated to teachers with some formative assessment skills which took place in Greece, in the period of 2019-2020. Briefly, the content of the specific group was multiple assessment techniques, assessment criteria and student's engagement is assessment process, results record of multiple assessment techniques and feedback. The basic tool of the program was the action plan, which is a document every teacher used to organise the strategies implemented in the classroom, according to aims, actions, time schedule, tools and their reflection on the implementation in the classroom. In the group meetings teachers was sharing their action plan and their experiences or their reflections on that was discussed collectively and also the professional development material was presented by the educators and discussed collectively as well.

Data collection conducted in a seven-month period, when the program took place and was on pre and in COVID-19 period, thus for data collected through face to face interaction or through video conferencing platform. Multiple resources were used for the

collection of the data. Specifically, observation of the PD meetings, semi-structured interviews and informal discussions with the cases were video or audio recorded and transcripted. Field notes of classroom observations of the action plan implementation and classroom material were also collected.

Data analysis followed the principles of Grounded Theory, specifically from the data collected the data referring to formative assessment practice focusing on teachers were selected through sensitizing questions (Vollstedt & Rezat, 2019). Following line by line coding strategies were deteceted and in vivo codes were uses for the implementation of the strategies. Finally, codes were checked again and the systematic network was created.

Results

The analysis of the data for each case, indicated multiple formative assessment strategies being used throughout the program, mostly in combination, when teachers implemented their action plan in the mathematics classroom. Following the categorisation provided by teachers' action plans, strategies will be presented accordingly.

In the case of Mary, self-assessment, observation of students when working individually and oral feedback was included in the first action plan. Peer-assessment and assessment criteria were implemented as part of the second action plan. Written feedback was also implemented in the online teaching not as a part of an action plan. All the strategies were a part of the professional development material and were discussed in the group meetings before or/ and after the implementation in the classroom, expected of the observation of students which appeared in the classroom as a spontaneous strategy of the teacher.

In the case of John, self- assessment and comment only marking were implemented followed by assessment differentiation and feedback in the first action plan. Peer and self- assessment as well as success criteria were part of the second action plan. Differentiation of assessment was part also of the summative assessment practice that John implemented.

The implementation of the strategies will be presented according to the aims, the tools and the setting enacted, while setting is presented according to the topic structure and the supervision rules of the specific strategies in the action plan. The action plan that was planned according to the professional development material will be presented.

In the case of Mary the systematic network of the implementation of formative assessment practice in the classroom is presented below.

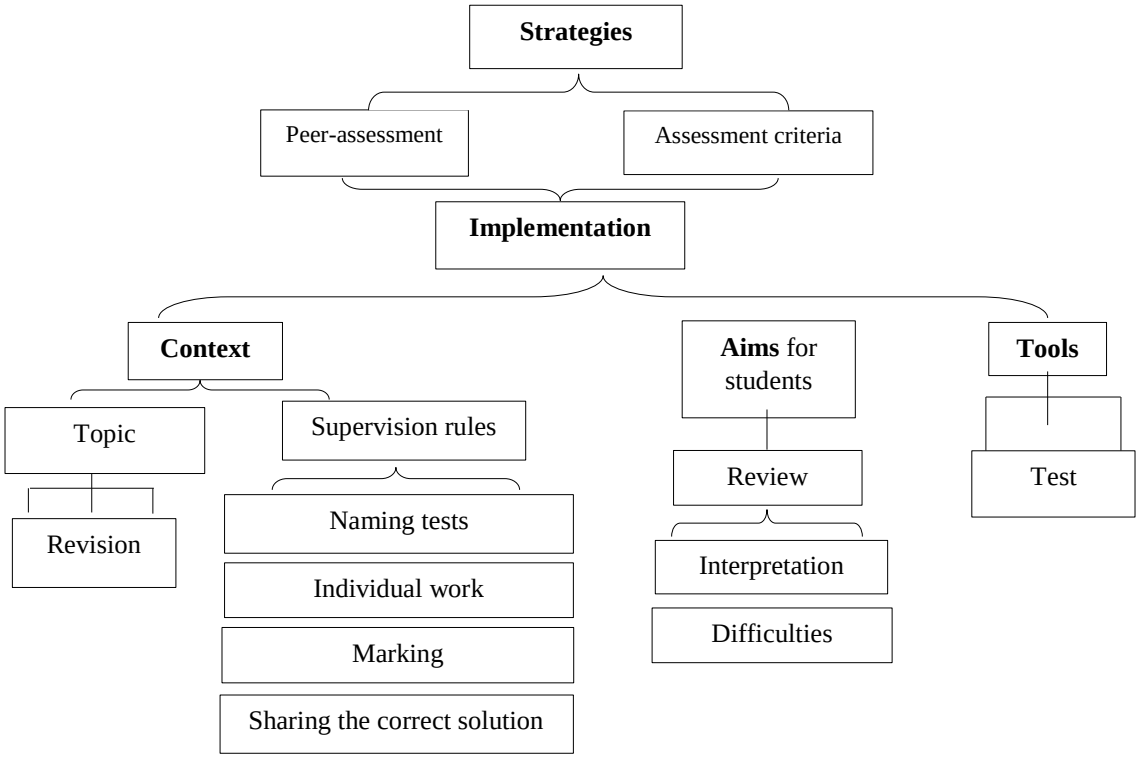


Figure 1. Systematic network of the implementation of formative assessment practice in the classroom.

The strategy of peer-assessment was implemented in the classroom after its formal presentation in the group meeting of the program and after the implementation and group reflection of the self-assessment. The strategy where uses in the *topic of revision* of the irrational numbers at grade 8 and as test was used as *tool* for that. For the implementation of the peer- assessment strategy the *supervision rules* was students' individual work when working on the test. Moreover, sharing of the correct solutions on the task in the classroom before the process of peer- assessment was another rule. Finally, peer-assessment conducted through students marking their classmate's solutions on the task and the finale step was marking of teacher as well, as she explained in the group meeting.

> *'With students help we will share the correct solutions on the broad, students exchange their worksheets and assess and mark each other's solution. Finally, I mark the solutions and give them back to see'*

Naming the test was a rule that also applied in the process of peer-assessment for students to have the opportunity to see the results according to teacher, as it is indicated from personal interview ('I will also mark their solutions and I will give them back to see how I marked (…), this is the most important to see how I also mark').

The strategy of peer-assessment was also combined with other strategies in the design process of the action plan and its implementation, like the recording assessment criteria and sharing them with students after peer- assessment process, as presented at image 1.

Image 1. Assessment criteria second action plan

Κριτήρια αξιολόγησης
Κατανόηση άρρητων
Εφαρμογή ιδιοτήτων τετρ. ρίζας
Ικανότητα εκτέλεσης πράξεων με άρρητους
Χρήση άρρητων σε προβλήματα

Moreover, feedback was implemented in the classroom to help students mark their peer answers.

The aims connected with the peer-assessment was about cognitive aspects as presented in the assessment criteria, but also the formative aim of students recognising their difficulties and understand their importance was posed, presented in image 2.

Image 1. Aims for second action plan

Στόχος/οι: Οι μαθητές θέλω :

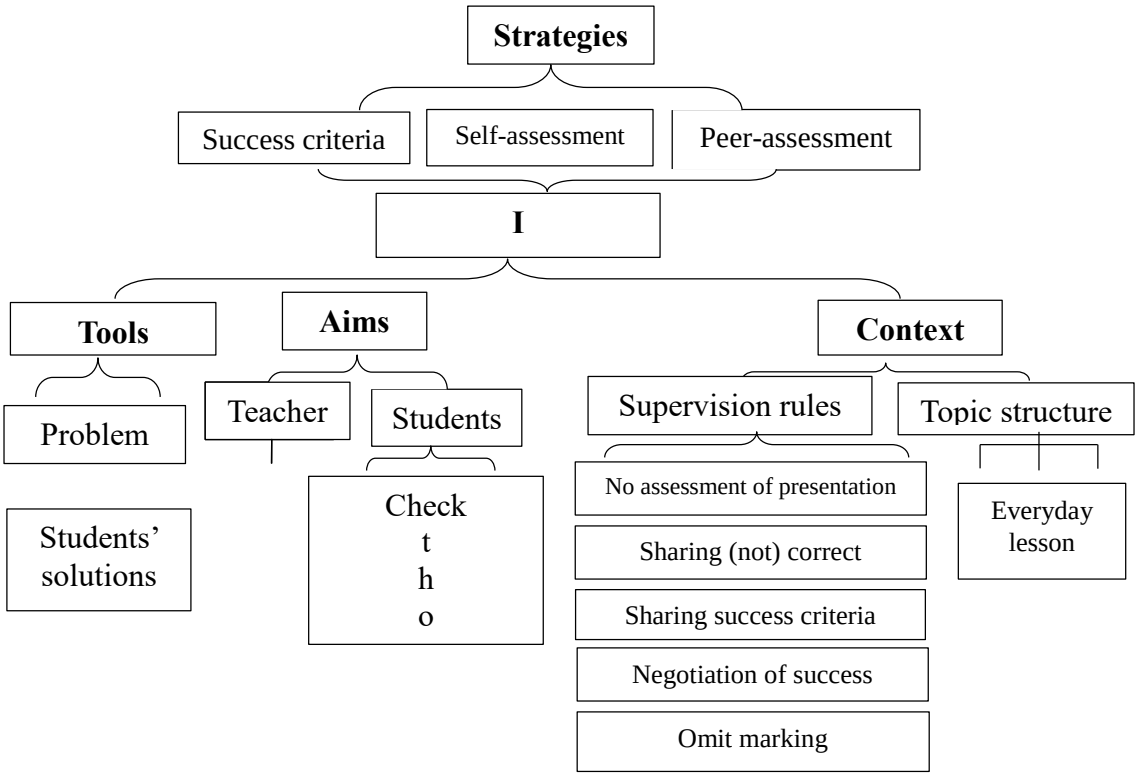
1. Να εφαρμόζουν ιδιότητες δυνάμεων για να εκτελούν πράξεις με άρρητους.

2. Να ξέρουν σε ποια περίπτωση θα χρησιμοποιούν το ΠΘ.

3.Να κατανοήσουν μέσα από τα προβλήματα ότι οι άρρητοι υπάρχουν αλλά δεν μπορούμε να τους προσδιορίσουμε ακριβώς.

4. Με την ετεροαξιολόγηση που θα ζητηθεί από τους μαθητές θα έχουν την ευκαιρία να δουν μόνοι τους τη βαρύτητα των λαθών- παραλήψεών τους.

For the case of John, the systematic network of how he implemented the formative assessment in the classroom is presented accordingly below.



The action plan that implemented by John, combined peer-assessment and sharing the success criteria in the classroom and self-assessment as a homework ('I then asked them to use the criteria to assess their own work'). The strategies were implemented in the *context* of the everyday lesson in the *topic* of linear systems in grade 9 and specifically problems regarding linear systems. The *tools* used for those strategies were a word problem (image 3), the success criteria (image 4) and four students answers collected by teacher in previous lesson.

Image 2. Word problem on linear systems

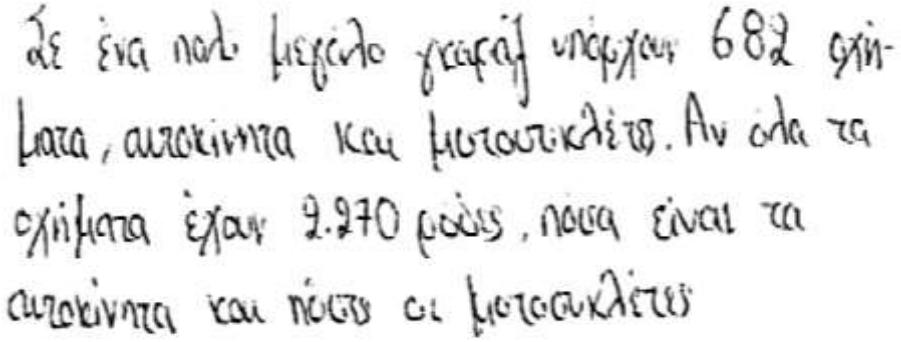


Image 3: Success criteria of linear systems problem

Λύση προβλημάτων με την βοήθεια συστήματος εξισώσεων

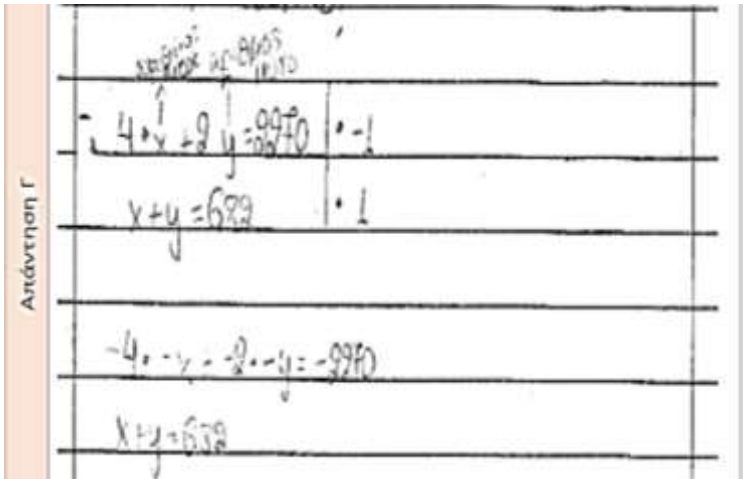
Κριτήρια επιτυχίας ως προς την διαδικασία					
	Ποσοστό επιτυχίας ανά εργασία				
	Ιδανική εργασία	Α	Β	Γ	Δ
Καταγραφή των δεδομένων και των ζητουμένων	5 %				
Σωστός καθορισμός των μεταβλητών	15 %				
Μετατροπή των σχέσεων που εκφράζει το πρόβλημα σε κατάλληλες εξισώσεις	30 %				
Αλγεβρική λύση του συστήματος των εξισώσεων	30 %				
Απάντηση στα ερωτήματα του προβλήματος	10 %				
Επαλήθευση της απάντησης αριθμητικά ή Επαλήθευση της απάντησης γραφικά, με την χρήση λογισμικού	10 %				
Σύνολο	100 %				

Regarding the supervision rules of the action plan, for students understanding the success criteria the teachers when sharing the correct answer to students with their participation in the process, he used and explained the criteria. Then students were asked to assess the solutions given using the criteria, as indicated in personal interview.

> *"We discussed the problem, with the help of students, but using the steps presented above (success criteria), we solve it on the board. Then I gave students their peers work to assess according to the criteria we posed earlier."*

Not only the correct answer was shared but also a mistake (Solution Γ- image 4) was pointed out and asked to be compared with the correct solution, when students could locate it at peer-assessment process. ('Finally, I told them, to observe the multiplication with -1 and asked them what would occur otherwise, so they compared with the correct answer').

Image 4: Solution C



Moreover, in some classrooms this process implemented, it was discussed the importance of the criteria and students gave their own explanations and interpretations, as he explained in group meeting of the professional development program.

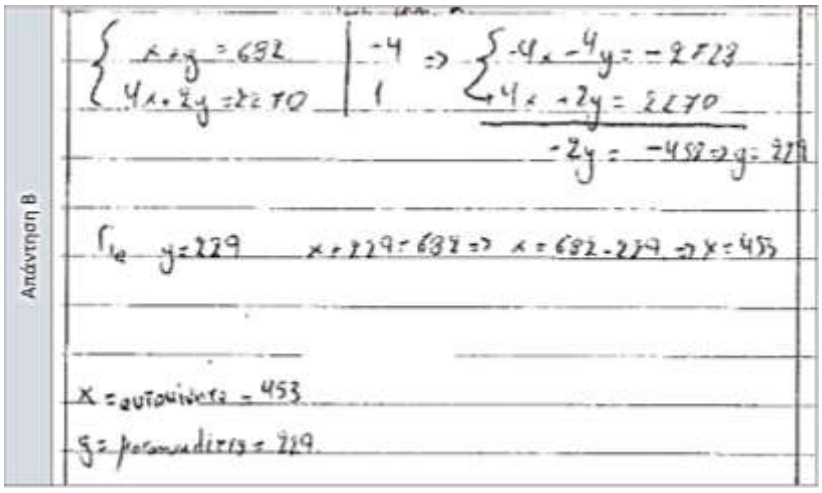
> "We discussed in two classrooms which are the most important steps according to them and they said that it was the expression of the equation and the algebraic solution in this problem. Thus, they decided independently to change the percentage."

The main aim of the teacher was students to get involved in a process to assess different solutions and check the toughness of them following specific methods and criteria ('An assessment process but with specific criteria'). Moreover, his aim was to engage with different solutions that may or may not correspond to the correct answer but also

according to his *supervision rules* of the process students solution should not follow the same order of the criteria given (‘ It is not compulsory a solution to presented in the same order. It is assessed only if the solution completes the criteria’). This aim was achieved by providing students a solution with a different form than the correct solution given but the criteria were achieved, as presented in personal interview and in image 5.

> *“In solution B, there are mixed things, and students were asked if the criteria exist in the solution. I started the discussion on purpose to show them that the criteria should not presented in specific order necessarily.”*

Image 5. Solution B



Finally, his aims was to provide to students criteria that can be generalised and used to assess any problem solution and not only the specific topic (‘The criteria are simply to indicate a complete answer to the problem, they are general to correspond to any problem’).

Discussion

This work is a case study of two teachers in secondary education, regarding the formative assessment practice enactment in the classroom, in the context of professional development program.

The strategies observed were peer and self-assessment, observation of students, oral and written feedback, comment only marking, sharing assessment and success criteria and assessment differentiation, which implemented in the classroom mostly in combinations. More specifically, peer and self-assessment, oral and written feedback, sharing assessment and success criteria were part of the professional development

material. Students' observation and comment only marking were not part of the program, but they can be located in the literature. Moreover, the strategy of assessment differentiation is not detected in the literature but is connect with the strategy of adjusted instruction (Andersson, 2017) which take into consideration students different needs, as well as the category of differentiation of assessment referred by Christoforidou et al. (2014).

Moreover, results revealed that formative strategies and aims can be integrated in the classroom practice with formative but also summative assessment process or tools which has been stated also by Black et al (2004). Specifically, mostly at revision sessions were self or peer- asssement were used even though summative assessment tools were used like tests or marking, a formative aim was posed for helping students reflect on their understanding and plan their individual revision better (Black et al., 2004) before the summative exams.

Marking or grading was part of both teachers' strategies implemented in the classroom, but it was either with an aim for students to focus on their difficulties or to indicate the importance of different solution processes. Thus, grading have received a different meaning in the formative assessment practice. Finally, the aims of the strategies used were mostly concerning students. Specifically, for them to engage and understand assessment and active them as owners of their own learning (Black & William, 2009).

Differentiation in the implementation of formative strategies in two cases is observed not only when strategies not detected in the program were used, but also when similar strategies were implemented. Even though teachers used the same strategies they adjusted them in different ways according to their interpretation. Further research can be conducted to study how different interpretations of formative assessment is implemented in the classroom by different ways and what factors contribute to those differences.

Acknowledgments

The work reported inhere was part of the project, titled "Promoting Formative Assessment: From Theory to Policy and Practice (FORMAS)" which has been funded with support from the European Commission.

References

Andersson, C. (2017). Formative assessment: and the component of adjusted teacher instruction. *In 10th Congress of European Research in Mathematics Education* (pp. 3419-3426). Dublin: DCU Institute of Education and ERME.

Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The qualitative report, 13(*4*),* 544-559.

Black P. , William D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, International Journal of Policy, Practice and Research*, 6-31.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi delta kappan, 86(*1*),* 8-21.

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., Kyriakides, L., Antoniou, P., & Creemers, B. P. (2014). Searching for stages of teacher's skills in assessment. *Studies in Educational Evaluation,* 40, 1-11.

Ginsburg, H. (2009). The Challenge of Formative Assessment in Mathematics Education : Children's Minds , Teacher's Minds. *Human Development,* 52, 109-128.

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

Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2008). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice, 28(3),* 24-31.

Marshall, B., & Jane Drummond, M. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research papers in education, 21*(2), 133-149.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral science, 28*(1), 4-13.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science,* 18(2), 119-144.

Santos, L. & Semana, S. (2014). Developing mathematics written communication through expository writing supported by assessment strategies. *Educational Studies in Mathematics,* 88(1), 65-87.

Santos, L., & Pinto, L. (2009). Lights and shadows of feedback in mathematics learning. In M. Tzekaki, & M. S. Kaldrimidou, *Proceedings of the 33rd Conference of International Group for the Psychology of Mathematics Education, vol. 5,* (pp. 49-56). Thessaloniki, Greece.

Vollstedt, M. & Rezat, S. (2019). An introduction to grounded theory with a special focus on axial coding and the coding paradigm. In G. Kaiser, & N. Presmeg (Eds.), *Compendium for Early Career Researchers in Mathematics Education* (pp. 81-100). Springer.

William D. (2007). Keeping learning on track. In F. Lester Jr. (Ed.), *Second handbook on mathematics teaching of learning* (pp. 1053-1098). Charlotte: Information Age Publishing.

The Impact of an Assessment for Learning Teacher Professional Development Programme on Student Learning

Jitske de Vries[a*], Andria Dimosthenous[b], Kim Schildkamp[a], Adrie Visscher[a]

[a]*University of Twente, Enschede, The Netherlands;* [b]*University of Cyprus, Cyprus*

Email: j.a.devries@utwente.nl

Introduction

Assessment for Learning (AfL) is a promising educational innovation. AfL refers to forms of assessment focused on generating feedback on the performance of students in order to improve their learning (Sadler, 1998).

However, teachers still find it difficult to implement AfL in their classroom, and as a result AfL does not always lead to improved student achievement (Heitink et al., 2016; Kippers et al., 2017). Teachers need professional development (TPD) to support them in developing and implementing AfL in their classrooms.

The focus of this study is to measure effectiveness of an AfL-TPD programme student achievement.

Theoretical framework

In AfL, three formative questions are answered by teachers and their students: “Where is the learner going?”, “where is the learner now?”, and “how is the learner going there?”. When these three formative questions are answered in a coherent way, AfL can result in better student achievement (e.g., Black & Wiliam, 1998).

Teachers need intertwined, complex competences to answer these questions effectively (Heitink et al., 2016). The Dynamic Approach (DA) towards TPD programmes can be suitable to help teachers develop these complex competencies (Creemers, et al. 2013). The reason can best be explained by the principles of the DA, which are:

(1) addressing professional needs. Teachers may best advance their varying AfL-skills if the TPD programme is suited to teachers' individual professional needs.
(2) integrating skills. As teacher competencies for AfL depend on each other, it can help transfer to the classroom, when the required skills for AfL are addressed as one coherent skill-set.
(3) explaining underlying mechanism. Teachers not only need to know that AfL can be beneficial for improving student learning, but also *how* it is beneficial. Such understanding might help teachers appreciate AfL, and be more motivated to implement it in their own classroom.
(4) supporting teachers. During TPD, teachers are asked to experiment with the learnt material in their own classroom practice. Without the help of an AfL-expert, it will be hard for teachers to persevere in their professional development as AfL-results are often only visible after correct and long-term implementation.

Methods and data sources

A total of 73 mathematics teachers was randomly assigned to the experimental group (i.e., the TPD programme) or the control group. The TPD programme included five sessions, the first one took place in September 2019 and the last one was planned in April 2020.

However, due to COVID-19, we had to reschedule the last session to September 2020. This also caused a drop-out of 20 teachers and 700 students.

Mathematics tests were developed and validated (N = 393) for the purpose of this study. These tests included 11 exercises, which were aligned to the curriculum of the students. The mathematics tests were administered before and after the TPD programme took place (N = 599).

Results

Using IRT techniques, it was possible to test the validity of the battery of tests and generate scores of student achievement in mathematics at the beginning and at the end of the intervention. Multilevel analyses revealed that the intervention had a statistically significant effect on student achievement gains in (d = 0.27).

Discussion

The results of this study show that the DA can be used to promote student learning through teacher professional development in AfL. This is in line with other findings (e.g. Antoniou & Kyriakides, 2011), which show that the four design guidelines can be beneficiary for the development of complex teacher skills, such as AfL.

Further research may include long-term investigation of the DA in TPD effects on student learning. In addition, this study may need to be replicated due to the fact that while it took place, there was a pandemic, which caused schools to be closed and the last session of the TPD to be rescheduled to the next schoolyear.

References

Antoniou, P., & Kyriakides, L. (2013). A dynamic integrated approach to teacher professional development: Impact and sustainability of the effects on improving teacher behaviour and student outcomes. *Teaching and Teacher Education*, *29*, 1-12.

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

Christoforidou, M., Kyriakides, L., Antoniou, P., & Creemers, B. P. (2014). Searching for stages of teacher's skills in assessment. *Studies in Educational Evaluation*, *40*, 1-11.

Heitink, M. C., van der Kleij, F., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational research review*, *17*, 50-62. https://doi.org/10.1016/j.edurev.2015.12.002

Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and teacher education*, *75*(October), 199-213. Doi: https://doi.org/10.1016/j.tate.2018.06.015

Creemers, B., Kyriakides, L., & Antoniou, P. (2012). *Teacher professional development for improving quality of teaching*. Springer Science & Business Media.

Flemish Student Perspectives on Evaluation and Feedback

Vlaamse Scholierenkoepel vzw

Pupil volunteers, VSK (Flemish Pupils' Council), Belgium

Email: info@scholierenkoepel.be

Abstract: As the representative organisation for Flemish school students, we often hear from students on the topic of evaluation in school. Based on this, we formulate positions which we try to inject into the national debate. From the point of view of school students, we feel it is important for schools to think about ways to guarantee that the way they evaluate their students actually contributes to their learning, that teachers are open to trying new things and that everyone plays their part in ensuring everything is manageable for students. We also make a few specific policy recommendations regarding the situation in Flemish education: greater focus on evaluation in school inspections, more focus on formative evaluation skills in teacher training and professionalisation, more knowledge exchanges between schools and more opportunities for experimentation in digital evaluation tools.

Introduction

Today I want to share with you a brief summary of what students themselves think on the subject of evaluation and how they regard the way evaluation currently happens in Flemish schools. In 2019 we heard over 13.000 students about a variety of subjects to put together a manifesto for the new minister of education. As you can imagine, the topic of evaluation was quite popular, so this gave us quite a bit of input to work with.

Looking at what students tell us, three main threads emerge.

1. Evaluation = education

First, and most importantly, evaluation equals education. The prime focus of evaluation procedures should be to help students learn. To show them what they are doing well, and what they are struggling with, but also: how they can learn from their mistakes, and what they can do to improve. For this reason we are glad to be a part of this conference, entirely focused on formative evaluation, because as you will have noticed this expectation is fairly close to the definition of formative evaluation.

Unfortunately, it is the other type of evaluation, summative evaluation, that is still very powerful in Flemish school culture, with grades often being seen as a means to an end. Not only by schools, but also by parents and even by students themselves.

Two years ago, we got a taste of the way this culture focused on grades is still ingrained in Flanders. In a media interview, our president at the time made the statement that grades should not even be necessary in evaluation. The result was a big backlash, featuring many caricatures of formative evaluation and scare stories of percentage scores being replaced by flowers and smileys. Schools that do experiment with getting rid of grades or dialing back their importance, even just by small intervention like getting rid of displaying the class average, often also have to deal with a local backlash, mainly from parents. Ambition is often still confused with aiming for high grades, when true ambition would be to make sure that all students are learning as much as possible.

Nonetheless, students mostly agree that a mentality switch would be a good thing, but this has to start with the way evaluation is approached by the schools and the teachers. It is essential that students understand in advance what they are being asked to study, how they will be evaluated, what the criteria will be for grading, and most importantly: what part these evaluations play in their learning process.

And of course, everything eventually comes down to the quality of feedback. Many Flemish students report not always understanding where the grades they receive are coming from. If we pass, it's okay; if we fail, it's a problem. Without meaningful feedback these grades mean little to nothing. I say 'meaningful', because common comments such as "study harder", "do better" or even "good job" do not add a lot. Focus on the competences that have been acquired or are still lacking and offer a clear path towards improvement.

2. Evaluation = experimentation

A clear path towards improvement is a great start, but we can do one better: multiple paths towards improvements. That's why students believe evaluation should also be about experimentation.

As evaluation's main purpose is to give an overview of the extent to which students have acquired certain skills, why not allow them some choice in the way to demonstrate this. Rather than teachers deciding on a multiple choice exam, an essay, a

spoken presentation or an open book exam; students can pick the form they are most comfortable with.

Teachers should be willing to think outside the box and look beyond the standard procedure of questions and answers. Many students know the material, but for whatever reason, be it nerves, lack of self-confidence, or less strong verbal skills do not perform well in the type of examination set up. This should not rob them of the opportunity to demonstrate what they have learned.

3. Evaluation = organisation

And finally, evaluation is also about organisation. It will probably not surprise many people that many Flemish students feel they have to do many tests. As it turns out, the statistics back them up. Almost nowhere in the world is so much time spent on evaluation. At the same time, teachers have a lot of freedom in how to go about this evaluation, and the national school inspection has for many years pointed out that most schools do not really have a solid policy on evaluation. The result is a lot of testing, for comparatively little benefit.

Simple changes can make a big difference. Consult with students on how to approach testing and homework. Agree school wide limits on daily workloads and spread out the moments of evaluation. A common problem in schools happens in the final days before a holiday, when all teachers scramble to get their grades in and overload students with tests.

Of course, students themselves also have a part to play in the matter of organisation. The school can lend a helping hand in teaching students how to go about planning their work, and deal with multiple deadlines. As most of you will know, this is a skill that will probably be beneficial for their entire working lives.

Policy recommendations

To summarise, from the point of view of Flemish school students, we feel it is important for schools to think about ways to guarantee that the way they evaluate their students

actually contributes to their learning, that teachers are open to trying new things and that everyone plays their part in ensuring everything is manageable for students.

For this reason, we have the following policy recommendations, specifically for the situation in Flanders, though we feel most of these will be broadly applicable.

- We want the school inspection to place even greater focus on evaluation policies in schools. School leadership has to think about how their testing contributes to students learning processes.
- Every teacher must be or be willing to become an expert in formative evaluation. Their objective should always be to achieve as much growth and learning gains in each individual student, not to ‘filter ’out students that do not meet the bar, as is now too often the case. Teacher training programmes and teacher professionalisation services obviously have a big part to play in this.
- School should have more opportunities, but also be more willing to learn from each other. School umbrella organisations are ideally placed to let schools reap the benefits and lessons from each other’s practices.
- A final thing, that is not often mentioned, but can be more important than many people realise: the way digital learning platforms are set up. In Flanders, a few players dominate the market and we have heard from schools that do experiment with evaluation that these systems are not prepared for this. The evaluation modules commonly feature fields that are to be filled in with numbers, and not much more, leaving little space for meaningful descriptionary feedback. This of course further ingraines a very limited way of thinking about evaluation.

Feedback can also go the other way…

Before wrapping up, I want to turn your attention to another initiative VSK has taken over the last years. Yesterday and today you have all hopefully learned a lot about evaluating students. But feedback does not have to be a one way street. Ideally it can travel in both directions. That’s why we have developed a scientifically validated feedbacktool to be used by teachers.Through it, we intend to lead by example. The feedback is not intended to grade teachers, but focuses on learning outcomes experienced by students. It is intended to serve as a jumping off point for a conversation with the class. We encourage you all to check it out.

Student Assessment Processes in Cyprus: The Transition to Formative Assessment, Policy Perspectives and Practice

Ioannis Ioannou

Inspector (Mathematics), Ministry of Education, Culture, Sport and Youth, Nicosia Cyprus

Email: ioioannou@schools.ac.cy

Abstract: Assessment is an indispensable and integral practice in the education process. Students and teachers, school heads, and all actors in the process of education are being assessed. The dominant assessment, mainly for secondary schools in our educational system, is summative, measuring what students have learnt through testing and written examinations. Simultaneously, this process is holding schools and teachers accountable for student performance, although school or teacher assessments are not directly connected or used for appraisal purposes. Recently, however, we have acknowledged the dynamics of formative assessment. We are in the process of seeking effective practices which will enable us through a successful shifting from the currently used summative approaches to blinding methodologies where teachers could successfully apply both formative and summative assessment approaches, each one for its contribution. This policy is part of a bunch of measures taken to improve children's learning attainment. The country's poor ranking in international educational surveys, mainly PISA, gave rise to coordinated efforts for better schools, teaching and assessment policies, and hopefully improved learning outcomes. Thankfully, the country's better ranking in the last TIMSS survey was a relief, and it has provided an impetus for intensifying our efforts for even better learning outcomes.

Introduction

Our decision to partake as active partners in the FORMAS project is part of many efforts that the Ministry of Education takes for better education for all pupils and students in Cyprus. The Ministry of Education (MOECSY) decided not only to call selected schools and teachers to participate in the project. Instead, it has been actively involved in all individual phases and activities of this project. Notably, in the development of the training material and the actual training of teachers, the development of the measuring instruments, the analysis of data. We wanted to earn in know-how and be able to transform it into praxis when time is ready.

Current assessment process in Cyprus

To articulate the challenges, we are to phase in this transition process, we need first to picture the assessment policies in action for our educational system. A report carried out by a committee set up by academics, ministry officials, and delegates from the teacher's

unions have studied student assessment practices in place and is suggesting a comprehensive scheme for student assessment.

Comprehensiveness means adopting student assessment policies that will uniformly apply from primary to lower and upper secondary education and inform about the students' learning attainment at each level of education. At the transition between levels of education (e.g., between pre-primary and primary, or primary and lower secondary etc.), lower levels ought to be in place to inform the successive level of education whether and to what extent students have attained the core curriculum indices as well as on non-cognitive characteristics of the whole child/student.

As regards the country's current student assessment policies, the committee has identified:

(1) There is an absence of a general and uniform educational student assessment policy between levels of education. Assessment policy, either formative or summative, is almost absent for Pre-Primary and Primary education. In Secondary education, the prevalent assessment method is summative.

(2) To a large extent, student learning and development are either determined from tests and final exams or is not adequately evaluated during semesters. As a result, students' weaknesses are not identified on time, and consequently, intervention or compensation programs cannot be planned effectively.

(3) Oral assessment practices, especially at the Secondary level, do not provide for reliable student achievement measurement. Instead, written tests and final examinations determine students' learning performance as it is reflected in their progress reports. As a result, student motivation for day to day actively engaged in learning activities is reduced.

(4) Assessment emphasises mainly procedural understanding, algorithmic knowledge, memorisation and retrieving of information described in the student's textbooks.

(5) Student assessment practices, and mainly their comparative nature, fail to provide reasonable documentation about students' achievement regarding the success and adequacy indicators of the school curriculum. As a result, the education system lacks the necessary information and feedback on curriculum implementation, which could drive improvement efforts.

(6) The assessment of academic achievement is predominant to the detriment of the cultural development of the whole child.

The committee's study for a comprehensive student assessment and evaluation was handed in in December 1918. The suggested measures were designed to combat the challenges as these are described above. The study suggests adopting a blinding approach to student assessment where both summative, comparative, diagnostic, and formative assessment practices will be enacted, each for its contribution. Diagnostic and formative assessments are essential for measuring attainment of intended learning outcomes and informing teachers for reshaping their teaching approaches. Comparative assessment is necessary for schools to meet their obligation for providing students and parents with the end of the semester and the end of year progress reports. Summative assessments are essential in measuring learning outcomes and whether and to what extent students have grasped the core curriculum success criteria. Notably, the committee's leading suggestions are:

(1) Formative assessment is to be the bulk of assessment practices at all educational levels. To meet this goal the report suggests the introduction of new assessment instruments such as the Individual Progress Report, (IPR). The IPR is defined as an instrument which summarises student's progress as regards the intended learning outcomes described by the success criteria of the curriculum and provides a description of student's strengths and weaknesses. Is characterized as a qualitative report unlike end-of-year reports which are quantitative. The instrument is to be used as a guide for any compensation measures to be designed for students at school level and inform teachers for the need to use differentiation teaching approaches in favour of students.

Of course, the IPR is not the solution in its own, rather is to be taken as a vehicle which will drive teachers in formative directions and stimulate the everyday use of formative assessment practices.

(2) Teaching methodologies have to be improved. The report underlines the importance of authentic teaching and learning, approaches which are more suitable for improving student's conceptual understanding and critical thinking.

(3) Assessment needs to be multifaced using both oral, written and performance forms. Emphasis is given to the day-to-day assessment where a list of possible

forms is suggested (e.g., homework, short quizzes, discourse in the classroom, short projects, self and peer evaluation, etc.).

(4) Introduction of semester central examinations for core subjects at lower and upper Secondary schools. Semester examinations are to be designed to inform teachers and schools about each student's attainment, identify learning needs and facilitate reshaping teaching to meet these needs. They serve also as an instrument for summative and comparative student assessment. In so doing, they are to be used for formative evaluation, at the central level, as regards the curriculum, textbooks, teaching methods, teacher professional development needs, etc.

(5) Teacher professional development in assessment methodologies and effective teaching practices is a precondition for a successful implementation of the comprehensive student assessment plan.

(6) Setting up a standing committee to monitor the implementation of the suggested comprehensive plan for student assessment at all levels of education.

While teacher professional development (TPD) is regarded as a precondition for putting in action the new suggested measures of a comprehensive scheme for student assessment, the MOECSY has proceeded to a parallel and simultaneous implementation of individual suggestions of the committee's report. For example, as form 2019, semester central examinations are gradually introduced in upper Secondary schools replacing the end of year summative examinations. Serval TPD activities have been organised by the Pedagogical Institute in cooperation with the Administrations of Education for targeted groups of actors in the educational system, (e.g., school inspectors, school leaders, senior teachers etc.). The MOECSY has reached a mutual agreement with primary teacher's union for the introduction, for the first time in the history of primary pupil assessment, of the suggestions described in the committee's comprehensive report. However, the journey to a shifting from summative student assessment and school and system evaluation to a formative one is long and concerted actions need to be taken some of which are described hereafter.

Implications for Further policy measures

Encouraging links between research, policy, and praxis

The Ministry is promoting and encouraging research into the area of formative assessment as well as the multifaced assessment practices and forms of assessment for individual school subjects as is for example the FORMAS project whose results and suggestions are presented in this international conference. The best practices developed by research programs should find their place in our policies. Moreover, there is a need for investing in training practitioners and policy officials to disseminate the research results. In a more general sense, we need to strengthen the capacity of our practitioners to draw upon research findings and be able to develop bespoke teaching material suitable for each learning subject. For example, the training material developed by the FORMAS project focus on Mathematics, while the principles guiding the training is applied in any school learning subject.

Invest in training for formative assessment.

Many actors in the process of education in our educational system still share the false assumption, as described by Margaret Heritage, professor at UCLA, that formative assessment is "a particular kind of measurement instrument, rather than a process that is fundamental and indigenous to the practice of teaching and learning", (Heritage, 2010, p. 1). International literature reviews show that formative assessment is much more than a set of best practices. In essence, teachers using formative assessment change the culture of their classrooms, (OECD, 2005). Therefore, beyond describing measures and even tools, as the report for a comprehensive student assessment scheme does, which is of course essential for an educational reform to be initiated, we are in need for a description of policy principles for formative assessment that can promote wider, deeper, and sustained practice of formative assessment investing on the role of teachers.

The FORMAS project, by emphasising the role of teachers, it has developed a framework for empowering them in ways of practical implementation for formative assessment as an integral part of their teaching. Furthermore, by using experimental design this project has found causal effects between improved student learning outcomes and the framework employed for the training of teachers. We have, as educational system, to draw from this methodology and successfully incorporate it in our policies for teacher training in formative assessment.

There are certain features of this approach that we have to consider when designing our TPD programmes as regards to formative assessment First, the approach acknowledges diversified needs for training among teachers. By using a validated instrument, this project was able to identify these differences and set up three distinctive groups of training needs among teachers. Then it has developed a training material suitable for each group and effectively delivered it organising five 2.5-hour long workshops. Understanding that formative assessment is foremost a way of teaching than a way of evaluation we need to take steps to enable our teachers enact the principles of formative assessment in their classes.

Develop links and reinforcement between formative and summative assessment.

The report for a comprehensive assessment framework has suggested the adoption of a blinding approach for assessment where both summative as well as formative practices will be enacted, each for its contribution. Yet, for effectively have in place both forms of student assessment, all actors in the educational system must have a clear understanding about the individual role of each assessment approach.

Moreover, it must also get clear that the two forms of assessment do not compete each other. On the contrary, when summative assessments are not merely design for comparative reasons, they can, and should be in place to inform and add to the whole picture of the learning attainment of students and aggregately for schools. This information can then be used at upper level for decisions to be taken for designing compensation programs, for individual students, or decisions for measures to improve schools' functionality.

The strong point of formative assessment is providing constructive feedback to students. Feedback on how to perform a task more effectively and how students can improve their work. Hattie and Temperly, (2007) studying the effect sizes for different kinds of feedback, they concluded that students given formative feedback have better results than their peers who receive praise, rewards, or punishments. Teachers, to be able to provide constructive feedback need data about student's ability to reason and apply knowledge to new situations. If standardised tests are well designed to measure student's ability, then formative and summative assessments can reinforce each other.

Providing feedback is also linked with the forms of informing parents and students about progress made. The proposed, interim to the end of semester Individual Progress

Report, (IPR) could be designed to serve as a formative assessment report. In turn, this measure will create the conditions and encourage teachers to the transition to formative assessment.

The committee's report on student's assessment has underlined, as a drawback in current assessment practices, that day-to-day assessments are merely based on unit written tests. As a remedy, the report suggests the introduction of multifaced forms of assessment. However, for any form of assessment activity, (oral, written, performance assessment) to be informative both for teachers and students, (promoting in that way metacognitive skills like self-assessment), assessment activities should be linked to carefully identified intended learning outcomes and success criteria. Moreover, assessment activities ought to inform both actors, teachers, and learners about the level of reasoning and concept understanding. Assessment activities should, also, enable teachers to record the results in a way that facilitates constructive feedback. Furthermore, teachers should be empowered in using assessment activities in an interactive manner orchestrating discourse in their classrooms. Concluding, the use of multifaceted assessment activities should be coated with formative characteristics to be able of usage both as summative as well as formative source of information.

Standardised tests, like the semester central examinations, pose the risk of hindering teachers from helping students develop their abilities, reach their potential, and explore their creative interests. Moreover, it is argued that high-stakes testing leads to narrowing the curriculum as teachers often jump quickly into ready-made solutions to address a specific issue, destroying, as a consequence, how and what students are taught. It is also underlined that a focus on testing demotivates and de-professionalised teachers. To avoid falling into this trap, end of semester examinations should be designed very carefully examining intended learning outcomes and success criteria while at the same time give room for non-trivialities.

Build coherence between assessment and evaluation.

If assessment's data is confined only to assigned grades to students and is not used for improvement purposes, then they are of little value. Teachers, school leaders and other actors in education will be capable using data in evaluating their job. We need invest in developing a culture of evaluation at all levels (teachers, schools, system).

Training actors in educations on how to analyze or correctly read assessment data is essential but is not enough.

School internal evaluations and school's action plans need to transform these analyses into actions for improvement. A policy for encouraging schools to follow self-evaluation strategies and develop action plans for improvement based on data is in place for a few years now. Yet, there is a long way to travel for meeting our aims. A perseverance on this goal is essential for being able to develop a culture of evaluation and conditions of sustainment.

References

Cirillo, M., & Herbel-Eisenmann, B. A. (Eds.). (2009). *Promoting purposeful discourse: Teacher research in mathematics classrooms*. National Council of Teachers of Mathematics

Clark, I. (2011). Formative assessment: Policy, perspectives and practice. *Florida journal of educational Administration & Policy*, *4*(2), 158-180.

Hattie, J., & Temperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.

Kaur, B., & Wong, K. Y. (Eds.). (2011). *Assessment in the Mathematics Classroom: Yearbook 2011*, Association of Mathematics Educators. World Scientific.

Organisation for Economic Co-operation and Development, (OECD). (2005). *Formative Assessment: Improving learning in secondary classrooms*. Centre for Educational Innovation and Research. Paris: OECD.

Wiliam, D., & Leahy, S. (2016). *Embedding formative assessment*. Hawker Brownlow Education.

Υπουργείο Παιδείας και Πολιτισμού, (2018). *Ενιαίο σύστημα αξιολόγησης μαθητή/μαθήτριας*, ΥΠΠΑΝ, Λευκωσία