

Adaptation and Examination of the Greek General Formative Assessment Framework P.D./152/2013 in Physical Education Teachers

Georgia Arampatzi^{1*}, Vassiliki Derri², Kyriaki Emmanouilidou³, Pavlos Kyrgyridis⁴¹Ph.D. candidate, Democritus University of Thrace, Department of Physical Education and Sport Science, Greece,²Prof. Department of Physical Education and Sport Science, Democritus University of Thrace, Greece,³Dr. Special Educational Staff, Department of Physical Education and Sport Science, Democritus University of Thrace, Greece⁴Dr. Department of Physical Education and Sport Science, Democritus University of Thrace, GreeceDOI: [10.36348/JASPE.2019.v02i08.001](https://doi.org/10.36348/JASPE.2019.v02i08.001)

| Received: 21.10.2019 | Accepted: 28.10.2019 | Published: 31.10.2019

*Corresponding author: Georgia Arampatzi

Abstract

The purpose of this research was to adapt and clarify the criteria of the Greek general assessment framework of the Presidential Decree (P.D.)152/2013 in the physical education subject and to examine the tool's validity and reliability. Secondary aims were a) the discrimination of the teachers according to their observed performance, and b) the examination of the effect of school grade in teaching effectiveness. The P.D. 152/2013 includes five categories analyzed in 15 criteria on a mixed four-level rubric (Incomplete, Sufficient, Very Good, Excellent). In the current research, the 10 observable criteria of three categories that relate to teaching were used. The criteria were analyzed in detail, clarified and adapted, according to international literature on physical education (PE) teaching effectiveness. Following the step-by-step upgrade of the rubric's levels (the same indicator at all levels but each time with more qualitative characteristics), 41 indicators were developed. In order to examine the validity and reliability of the adapted tool, trained evaluators observed in person 40 PE lessons taught in early and in upper grades by 20 PE teachers. The lessons were assessed in every indicator of the rubric, and classified in one of the four performance levels. Principal Component Analysis, descriptive statistics, chi-squared and ANOVA tests were carried out to examine the construct validity of the tool. The results revealed that the adapted rubric to the PE subject included the three initial categories, seven criteria and 26 indicators and can be considered as a valid and reliable instrument for assessing PE teachers' effectiveness.

Keywords: evaluation, rubric, physical education teacher, observation tool, assessment, effective teaching.

Copyright © 2019: This is an open-access article distributed under the terms of the Creative Commons Attribution license which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use (NonCommercial, or CC-BY-NC) provided the original author and source are credited.

INTRODUCTION

Teacher evaluation is a controversial and central issue in current educational policy in many countries [1, 2], since the way the teachers design and create learning opportunities in their classrooms influence largely the students' learning [3, 4]. Indeed, the teaching quality offered by the teachers is considered as the most important of the within-school and extracurricular factors that influence the achievement and the development of the students [5]. Effective teachers, compared to less effective, act supportively for all the students of their classroom, in order to improve them to the greatest possible extent, independently of any heterogeneity in their background (Organization of Economic Cooperation and Development, OECD) [6, 7].

Although assessment is considered an inherent element of the educational process [8], there is a

considerable heterogeneity of programs, methods and instruments in the assessment of teacher effectiveness, internationally [9, 1, 2]. This heterogeneity may be due to the difficulty in identifying the concept of effectiveness in teaching [10] which is considered as an interaction among teacher, students, content, and context [11, 12]. An important aspect is the purpose that teacher evaluation is designed to meet. This involves an accountability (summative) function and an improvement or developmental (formative) purpose, and usually some combination of both [13, 14, 1, 2]. In the first case, the key concern is to distinguish between effective and less effective teachers, with positive and negative consequences attached to the results. In a formative approach the main goal is to help teachers improve their performance through diagnosis and feedback [15].

A well-organized teacher evaluation system should include precisely established content, methods

and metrics of assessment, as well as with feedback and continuous support plan [16]. At the same time, the results of the evaluation should be used for the improvement of teaching quality, of students' learning and of the whole educational system [17].

Research on effective teaching in physical education (PE) was largely related to the work being done in the classroom [10]. Over the last decades, researchers have developed systematic observation tools which have been used both in Greece and internationally to evaluate one or more qualitative aspects of PE teachers' instruction. Systematic observation is considered the most reliable and objective method not only for teaching assessment [18] but also for the level of student engagement, the interaction between the students and the teacher, the focus of the subject, and the depth of instruction [19, 20]. Consequently, observation systems could be able to discriminate effective from less effective teachers and can help them to become more effective [10].

However, there is a limited number of observation tools with effective teaching standards for the evaluation of Physical Education Teachers (PETs). According to Rink [10], the difficulty lies in developing observation systems able to distinguish effective from the less effective teachers. For instance, the tool of NASPE [21] is an observation tool which includes critical elements of effective PE teaching such as instructional variables, evidence of learners' achievement, class management and climate, and professionalism. Nevertheless, it is necessary to clearly define and exemplify the structural elements of the tool in order to become valid and ready to use.

In addition, the up to now developed generic tools for teachers' assessment such as the Framework for Teaching Evaluation Instrument (FFT) [22] present a weakness in implementation to PE. Specifically, the FFT is one of the most comprehensive tools and provides indicators for four domains (design and preparation, educational environment, instruction, professional responsibility) only two of which can be observed, and corresponding subcategories, which are presented with examples and assessed by a comprehensive scoring rubric. However, in order to be used in PE, its observed behaviors as well as the description and the examples in each category should be adapted. In line, the observers should practice to discriminate the observed behaviors. The Physical Education and Lesson Observation Tool (PELOT) [23] developed in Singapore is an example of the generic tool (FFT) that was adapted to PE [10].

In Greece, several tools of systematic observation have been used in order to evaluate specific single or few components of effective teaching in PE, such as the Qualitative Measures of Teaching Performance Scale (QMTPS) [24], the Academic

Learning Time – Physical Education (ALT-PE) [25] and the Observation System for Content Development – Physical Education (OSCD-PE) [26]. Studies [27] also identified a number of characteristic elements of PE effective teaching, and contributed to the development of the Questionnaire of Self-Evaluation of Teaching Effectiveness in Physical Education (SETEQ-PE) [17], a five level scale with 43 criteria which assess 11 thematic units of effective teaching in PE (1. PE content, 2. Lesson implementation, 3. Classroom management and organization, 4. Learning environment, 5. Teaching strategies, 6. Teacher communication skills, 7. Teacher adaptability, 8. Teacher feedback, 9. Student assessment, 10. Teaching assessment, and 11. Use of technology).

Despite the efforts of the researchers and at times legislation, the issue of instructional evaluation remained equivocal and largely unresolved in our country. Only the last few years, a great interest has been expressed about the instructional evaluation, with the publication of the Presidential Decree (P.D.) 152/2013 [28], which is directly associated with the necessity of continuous improvement of teaching quality. Specifically, it is about a general framework designed to examine and evaluate the Greek teachers' performance in all school subjects, including PE, which defines that the purpose of teachers' assessment is the contribution to the continuous improvement of their teaching quality (article 2) for the benefit of the teachers, of the students and of the society. The evaluation of the teachers takes place in connection to the sociocultural context, in which they perform and the self-assessment of the school unit in which they serve, and not according to established standards and norms at national level. As Darling Hammond [13] noticed, we are moving toward a “conception of a teacher evaluation as part of a teaching and learning system that supports continuous improvement for individual teachers and the profession”.

The general assessment framework of P.D. 152/2013 takes into account the Greek educational system purposes, the existing and the additional curricula, the roles and the tasks of the teachers, the modern pedagogical theories and the educational research of international and Greek literature on teacher evaluation. Specifically, it focuses on the educational process and on the framework of the effective teaching and examines and evaluates the educational work with five categories analyzed in 15 criteria, on a four-level rubric (Incomplete-Sufficient-Very Good-Excellent).

In conclusion, from the review of the international and Greek literature emerged that the assessment of the instructional work is related to a continuous process of reinforcement, feedback and improvement of teachers, in the frame of their professional development, and contribute to the improvement of the students and the school unit [29,

30]. However, the criteria of the general assessment framework of P.D. 152/2013 should be clarified and adapted to PE. Also, the validity and reliability of the tool should be examined before its application in the evaluation-training process of the PE teachers.

Purpose

The purpose of this study was to adapt and clarify the criteria of the general assessment framework of the P.D. 152/2013 for the PE subject, based on international research data on effective teaching, and to examine its validity and reliability. Secondary aims were a) to classify the teachers according to their performance in the assessment framework, and b) to examine the effect of grade on teacher effectiveness. The hypotheses of the study were the following:

- The general assessment framework of the P.D. 152/2013 would have clearly adapted criteria with adequate adjustment indicators.
- The general assessment framework of the P.D. 152/2013 would be able to separate each participant's teaching practice in high and low quality.
- There would be a statistically significant difference in teaching effectiveness, as evaluated through the general assessment framework of the P.D. 152/2013, due to school grade.

Experimental section

Participants and Data Collection

Twenty elementary PETs (9 males and 11 females) with teaching experience from 5 to 25 years ($M=11.6$ years, $SD=5.7$) participated. Each one of them was observed and assessed live by two experienced evaluators in two lessons, one in an early (1st-3rd) and one in an upper (4th-6th) grade. A total of 40 lessons were divided in 172 teaching episodes and were observed and assessed via the rubric of the P.D./2013. A teaching episode is a task cycle that includes the presentation of the task, the organization of the class for the task, and the response to the students' task engagement [31]. The study was conducted after the consent of the directors of the schools and the teachers. Also, the study was approved by the Democritus University Ethics Committee.

Rubric adaptation

The original framework of the P.D. 152/2013 as evaluation tool uses a rubric with mixed system, step-by-step, additive, and upgrade mode descriptors to define the levels of each criterion, and includes five categories analyzed in 15 criteria on a 4-level scale (Incomplete, Sufficient, Very Good, Excellent). In this research, only the 10 criteria of the three categories related to the teaching observation were used: I-Educational environment (3 criteria), II-Planning, Design and Preparation of Teaching (3 criteria) and III-Implementation of Teaching and Student Assessment (4 criteria). Initially, the authors studied the bibliography on which the design of the assessment framework was

based, and then they analyzed in detail, clarified and adapted the criteria to the PE subject, according to their relation with effective teaching, as this was described in the international literature. Afterwards, the indicators for each criterion were identified, following the step-by-step upgrade where the same indicator appears at all levels, but each time with more qualitative characteristics. In the next phase, the observation tool, namely the adapted form of the 4-level rubric with the 10 criteria and 41 indicators, was examined for its construct validity and internal consistency reliability.

STATISTICAL ANALYSES

The inter-rater reliability (IR) of the evaluators was calculated for each item and for the whole rubric using their percentage of agreement: $\% IR = (\text{number of agreements} \div [\text{number of agreements} + \text{number of disagreements}]) * 100$ [32].

Principal Component Analysis (PCA) with oblique rotation (assuming correlated factors) was used to identify the underlying factor structure of the general assessment framework of the P.D. 152/2013 for the PE subject. Indicators' loading greater than .35 were considered to load to a particular factor as it is suggested for a sample size of approximately 200 cases [33]. The Bartlett's test of sphericity [34], the Kaiser-Meyer-Olkin (KMO) [35] measure of sampling adequacy value and the communalities of the indicators for the test of their multicollinearity were checked as assumptions that the data had to meet in order for PCA to give valid results. Internal consistency of the rubric was estimated by Cronbach's alpha with minimum acceptable value .70 [36]. Also, the construct validity of the rubric was examined further with a) descriptive statistics, b) x2 analysis, to identify if the rubric could discriminate the participants' instructional practices in the factors of the tool in the early and upper grades, and c) t-test for independent samples for the possible discrimination of the participants in groups, according to their scores in each factor of the rubric, as those emerged from the PCA.

RESULTS

The 40 observed lessons included 172 teaching episodes which were analyzed. The inter-observer reliability was 98%.

The Rubric

The test of assumptions for PCA indicated that KMO measure of sampling adequacy value was .83 supporting the appropriateness of data for PCA and the Bartlett's test of sphericity was statistically significant at .001 level, showing adequate correlation among the indicators which allowed the analysis to proceed.

The initial PCA revealed 10 factors which explained 72.36% of the total variance. However, indicators with negative loadings as for example OC1-OC3 (Category II: Objectives and Content), SP3

(Category III: Students' preparation) and PS4 (Category III: Participation of students), with low loadings as for example EC1 - EC4 (Category I: Educational climate), IA1 – IA5 (Category II: Instructional activities and educational materials) or with loadings in multiple factors as for example OC1 – OC3 (Category II: Objectives and Content) and IA3 (Category III: Instructional activities and educational materials) were eliminated. The next round of PCA revealed seven factors with 26 indicators which explained the 61.27% of the total variance.

Specifically, the 26-indicators rubric in that phase consisted of the following factors (criteria): 1. Interpersonal relations and expectations (IR) (4

indicators: IR1, IR2, IR3, IR4), 2. Classroom organization (CO) (4 indicators: CO1, CO2, CO3, CO4), 3. Perception of students' capabilities and needs (PC) (3 indicators: PC1, PC2, PC3), 4. Students' preparation for the instruction (SP) (3 indicators: SP1, SP2, SP3), 5. Instructional activities and educational materials (IA) (5 indicators: IA1, IA2, IA3, IA4, IA5), 6. Participation of students in the learning process (PS) (4 indicators: PS1, PS2, PS3, PS4) and 7. Consolidation of new knowledge and students' assessment (CA) (3 indicators: CA1, CA2, CA3). The loadings of the indicators fluctuated between .39 and .87 which is considered acceptable [33] and grouped in the factors which they theoretically represent (Table 1).

Table-1: The Rotated Loading Matrix from the Principal Component Analysis

Categories	Criteria	Indicators	Factors							
			1	2	3	4	5	6	7	
I: Educational Environment	1. Interpersonal relationships & expectations (IR)	1. IR1. The T* develops teacher-student relationships.	.77							
		2. IR2. The T* develops a spirit of coherent class, mutual and reciprocal acceptance among students.	.89							
		3. IR3. The T*has learning expectations for students.	.77							
		4. IR4.The T*has interpersonal expectations for students.	.78							
	2. Classroom organization (CO)	5. CO1. The T* takes care of the safe and functional sports area's arrangement.		.75						
		6. CO2.The T* applies behaviour protocols.		.58						
		7. CO3.The T* manages instructional time.		.42						
		8. CO4.The T* manages the students' problem behavior.		.54						
II: Planning & Preparation of Teaching	3. Degree of perception of students' capabilities and needs (PC)	9. PC1. The T* takes into consideration the students' specific psychological and developmental characteristics.			.45					
		10. PC2.The T* takes into consideration the students' learning readiness.			.48					
		11. PC3. The T* takes into consideration the students' knowledge and skills of prior lessons or classes or extracurricular experiences.			.54					
III: Implementation of Teaching & Student Assessment	4. Students preparation for the instruction (SP)	12. SP1.The T* sets objectives and developmentally appropriate activities.				.77				
		13. SP2. The T* informs and prepares students.				.41				
		14.SP3. The T*activates students' prior knowledge.				.77				
	5. Instructional activities & Educational materials (IA)	15. IA1. The T* includes in teaching more than one purposes of two PE sectors.					.51			
		16. IA2. The T* applies the curriculum content and differentiated teaching.					.45			
		17.IA3. The T* motivates all students to participate and provides them appropriate feedback.					.73			
		18.IA4. The T* completes teaching with summary and consolidation questions/exercises.					.55			

6. Participation of students in learning process (PS)	19. IA5. The T* completes teaching with appropriate time management.											.48		
	20. PS1. The T* sets objectives and ensures success experiences for all students.												.41	
	21. PS2. The T* assigns activities to students using instructions, clarifications and directions.												.70	
	22. PS3. The S* are interesting in learning and respond to questions and activities.												.67	
	23. PS4. The S* participate in learning activities.												.43	
7. Consolidation of new knowledge and students' assessment (CA)	24. CA1. The T* uses assessment methods for students' evaluation.													.80
	25. CA2. The T* assesses all the lesson objectives systematically.													.87
	26. CA3. The T* takes into consideration the age, the needs, the abilities, the prior knowledge and the social-cultural data for students' assessment.													.41

T*=Teacher S*= Students

The chi squared analysis revealed statistically significant differences between the teaching practices in

the early and upper grades in all the seven criteria of the rubric (Table 2).

Table-2: Chi-square test indices for the 26-indicator and 7-factor rubric

Grades	Interpersonal relations and expectations		Classroom organization		Perception of students' capabilities and needs		Students' preparation for the instruction		Instructional activities and educational materials		Participation of students in the learning process		Consolidation of new knowledge and students' assessment	
	χ^2	df	χ^2	df	χ^2	df	χ^2	df	χ^2	df	χ^2	df	χ^2	df
Early	2.95	6	10.93	6	10.29	7	17.04	4	16.21	9	16.68	6	16.67	5
Upper	33.04	8	10.31	7	2.58	5	11.35	5	8.69	8	4.38	5	6.65	6

The mean score in the total of 172 teaching episodes (M=1.92, SD=.32) of the 7-factor rubric was used to distinguish participants' teaching quality into two groups (below and above average). Participants with mean score greater than 1.92 joined Group 1 (above average) while those with score lower than 1.92 joined Group 2 (below average). Eighty-eight teaching episodes were classified in Group 2 and 84 teaching episodes in Group 1. The t-test for independent samples showed a statistically significant difference between the two groups (M1 = 2.16, M2 = 1.66, t (170) = 17.01, p<.001).

Also, the results revealed satisfactory construct validity of the rubric with acceptable values for the

internal consistency (factors' Cronbach's a >.70) [37]. Specifically, the total rubric showed an acceptable reliability value (Cronbach's a=.84). Also, the internal consistency of each factor was found satisfactory: (1) Interpersonal relations and expectations .78, (2) Classroom organization .82, (3) Perception of students' capabilities and needs .80, (4) Students' preparation for the instruction .79, (5) Instructional activities and educational materials .77, (6) Participation of students in learning process .81, and (7) Consolidation of new knowledge and students' assessment .78. The factors of the observation tool, the number of indicators and Cronbach's a values as well as means and standard deviations for each grade are shown in Table 3.

Table-3: Descriptive statistics and Cronbach's a values of the 7 factors of the rubric for early and upper grades of elementary school

Factors (criteria)	Indicators	Cronbach's a	early grade		upper grade	
			M	SD	M	SD
Interpersonal relations and expectations	4	.78	1.52	.32	1.36	.43
Classroom organization	4	.82	1.36	.64	1.77	.60
Perception of students' capabilities and needs	3	.80	2.29	.72	2.46	.39
Students' preparation for the instruction	3	.79	1.54	.31	1.76	.33
Instructional activities and educational materials	5	.77	1.88	.35	2.14	.34
Participation of students in the learning process	4	.81	2.08	.28	2.13	.24
Consolidation of new knowledge and students' assessment	3	.78	1.36	.64	1.77	.60

Note: Scale: 1=Incomplete, 2=Sufficient, 3=Very Good, 4=Excellent

DISCUSSION

The purpose of this study was to adapt and clarify the criteria of the general assessment framework of the P.D. 152/2013 for the PE subject, based on international research data on effective teaching, and to examine its validity and reliability. Secondary aims were a) to classify the teachers according to their performance in the assessment framework, and b) to examine the possible effect of grade on teaching effectiveness. Only the three of the five categories of the P.D. 152/2013 general assessment framework (Category I - Educational Environment, Category II - Planning, Design and Preparation of Teaching, and Category III - Implementation of Teaching and Student Assessment) were related to teaching and were included in the current research. The adapted form of the framework included the 10 criteria of the original form, after they were analyzed in detail and clarified for the PE subject, according to the international literature on effective teaching. These 10 criteria were expressed in 41 indicators, following the step-by-step upgrade, on a 4-level scale (Incomplete, Sufficient, Very Good, and Excellent).

Afterwards, the construct validity and the internal consistency reliability of the tool were tested. The Exploratory factor analysis revealed 26 indicators which represented seven of the criteria of the tool (1. Interpersonal relations and expectations, 2. Classroom organization, 3. Perception of students' capabilities and needs, 4. Students' preparation for the instruction, 5. Instructional activities and educational materials, 6. Participation of students in the learning process and 7. Consolidation of new knowledge and students' assessment. Three to five indicators loaded well on each criterion, verifying the first hypothesis of the study.

The tool can be regarded as a valid instrument for the assessment of the PE teachers' effectiveness as it can distinguish the teaching practice of the teachers in all seven criteria for both grades as well as their teaching quality. Specifically, it can evaluate the ways in which the PETs: a) interact with their students, and encourage and cultivate the interaction among students, in order to create a positive learning environment, in which there are high expectations and all students feel valued and safe, b) establish and teach systematically behavior protocols with well-defined consequences, in collaboration with students as well as efficient routines and procedures for the smooth operation of the classroom and the maximization of instructional time, c) take into consideration the learning readiness, the needs, the capabilities, and the interests of all students to design the lesson, d) activate students' prior knowledge by linking it to previous lessons or experiences, to their instructional needs as well as to the social-cultural class composition, e) encourage and motivate all the students to participate, and observe and provide appropriate feedback to all of them, f)

encourage all the students to participate in teacher-centered and in student-centered activities, and g) assess systematically all the lesson objectives.

The second and the third hypotheses of the study were verified by the results which supported the validity of the adapted general assessment framework of the P.D. 152/2013 to PE, as the 20 participants' performance was categorized according to the average of the 40 observed lessons in two groups with significant difference in overall quality. Moreover, it was indicated that the criteria of the tool were able to discriminate the teaching practice of each teacher in the early and upper elementary grades. Therefore, the adapted assessment framework of the P.D. 152/2013 can be considered a valid instrument for assessing teaching effectiveness in PE.

According to the maximum possible score on the 4-level rubric, that reflects the implementation of the criteria during the lessons, PETs teaching was classified into medium quality in the criteria "Perception of students' capabilities and needs", and "Participation of students in the learning process" in both grades, as well as in the criterion "Instructional activities and educational materials" but only in upper grades. Conversely, the participants' performance was of lower quality in all the other criteria ("Interpersonal relations and expectations", "Classroom organization", "Students' preparation for the instruction", "Consolidation of new knowledge and students' assessment) in both grades, and in "Instructional activities and educational materials" only in early grades.

Three of the ten criteria and 15 of the 41 indicators were eliminated, even though they were considered essential for effective PE teaching: Category's I criterion "Educational Climate" (4 indicators), the Category's II criteria "Objectives and content" (4 indicators) and "Instructional activities and educational materials" (5 indicators), as well as indicators of the Category's III, namely, one indicator of the criterion "Instructional activities and educational materials" and another one of the criterion "Participation of Students in the learning process".

The eliminated indicators of the three criteria and the low values in the participants' performance in many criteria's indicators, in which they were assessed, is likely to relate to a lack of targeted training and evaluation on recent data of effective teaching. Indeed, the design of the general assessment framework of the P.D. 152/2013 is based on modern pedagogical theories and recent international and Greek educational research. The participating PETs were not trained on all those elements that constitute an effective PE teaching. Specifically, low performance in the criterion "Interpersonal relations and expectations" in both

grades, and especially in the upper ones, is probably related to the fact that the PETs focused mainly on teaching for the development of motor and sport skills. It is well known that apart from motor skills the PE subject can contribute with a unique manner to the development of cognitive, affective and social skills and values which are required in an interactive and effective lesson [38-40]. It is obvious that targeted PE programs are required for the development of such skills.

In line, the above low scores might be attributed to the traditional teaching strategies that mostly the PETs used in the observed lessons, which seem not to contribute to student's social-emotional development [41] since they provide few opportunities to students to interact, to disagree, to help each other and to cooperate. Also, the low values in the criterion "Classroom organization" is probably related to a lack in establishing and teaching behavior rules with well-defined consequences in collaboration with students. PETs should use systematically efficient routines and procedures for the smooth operation of the classroom and the maximization of instructional time. Students' understanding of the routines but also their contribution to the formation of instructional groups, transition from activity to activity and handling the PE supplies are considered critical [22].

In regards to the low performance in the criterion "Students' preparation for the instruction", the PETs should activate students' prior knowledge by linking it to their past experiences and instructional needs as well as to the social-cultural composition of the class. Also, the values of the criterion "Consolidation of new knowledge and students' assessment" were low, although the assessment of students demonstrates to the teacher whether and to what extent the goals and the objectives of the lesson were achieved, and enable the provision of respective instructional feedback as well as the direction of future instruction [42].

The adapted assessment framework of the P.D. 152/2013 to PE was used to observe two lessons of each teacher in different elementary grades. The results would probably have been different if the observed lessons were more.

Future studies could also investigate the influence of theoretical and practical training seminars on PE teaching effectiveness. Also, the combination of results derived by PETs' self-assessment and assessment by experienced evaluators could provide teachers with a key role in their evaluation-training process and enhance further teaching effectiveness in PE.

CONCLUSION

The general assessment framework of the P.D. 152/2013 was not previously tested. The adapted tool to

PE is considered as a valid instrument which can be used both by external evaluators as well as teacher themselves, with ultimate goal to improve teaching in PE. This observational tool could also contribute to the professional development of PETs and therefore to the students' whole development and lifelong learning, as the evaluation of the observed lessons identifies teachers' strengths, weaknesses and needs in the school environment. In addition, the needs identified by the results of the tool's use should be taken into account in the organization of theoretical and practical training seminars that aim at increasing teaching effectiveness [43, 44], as well as in the reformation of the curricula of the Departments of Physical Education and Sport Science, in order to upgrade the PE teaching quality.

Funding

The present study is part of a research which is implemented through the Operational Program "Human Recourses Development, Education and Lifelong Learning" and is co-financed by the European Union (European Social Fund) and Greek national funds.

REFERENCES

1. Organization for Economic Co-operation and Development (2013a). Synergies for Better Learning: An International Perspective on Evaluation and Assessment. OECD Reviews of Evaluation and Assessment in Education. Paris, France: Author. doi: 10.1787/9789264190658-en. Retrieved from website of the Organization for Economic Co-operation and Development Library: <http://dx.doi.org/10.1787/9789264190658-en>.
2. Organization for Economic Co-operation and Development (2013b). Teachers for the 21st Century: Using Evaluation to Improve Teaching. Paris, France: Author. Retrieved from website of the Organization for Economic Co-operation and Development Library: <http://www.oecd.org/site/eduistp13/TS2013%20Background%20Report.pdf>.
3. Darling-Hammond, L., & Bransford, J. D. (Eds.) (2005). Preparing teachers for a changing world: What teachers should learn and be able to do. San Francisco, CA: Jossey-Bass.
4. Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454–499.
5. Caena, F. (2011). Literature Review: Quality in Teachers' Continuing Professional Development. European Commission.
6. Organization of Economic Cooperation & Development. (2005). Teachers matter: Attracting, developing and retaining effective teachers. Paris: OECD.
7. Nye, B., Konstantopoulos, S., & Hedges, L.V. (2004). How large are Teacher Effects?

- Educational Evaluation and Policy Analysis, 26, 237-257.
8. Papadopoulou, V. (2013). Teaching observation: Theoretical framework and applications. Thessaloniki: Publishing, Kyriakidis.
 9. Isoré, M. (2009). Teacher Evaluation: Current Practices in OECD Countries and a Literature Review. OECD Education Working Papers, 23. OECD Publishing. doi:10.1787/223283631428. Retrieved from website of the Organization for Economic Cooperation and Development Library: <http://dx.doi.org/10.1787/223283631428>.
 10. Rink, J. E. (2013). Measuring teacher effectiveness in physical education. *Research Quarterly for Exercise and Sport*, 84(4), 407-418.
 11. Rose, M. (2011). The mismeasure of teaching and learning: How contemporary school reform fails the test. *Dissent*, 58(2), 32-38.
 12. Ward, P. (2013). The role of content knowledge in conceptions of teaching effectiveness in physical education. *Research Quarterly for Exercise and Sport*, 84(4), 431-440.
 13. Darling-Hammond, L. (2012). Desarrollo de un enfoque sistémico para evaluar la docencia y fomentar una enseñanza eficaz. *Pensamiento educativo*, 49(2).
 14. Kennedy, M. (Ed.). (2010). *Teacher assessment and the quest for teacher quality: A handbook*. John Wiley & Sons.
 15. Taut, S., & Sun, Y. (2014). The Development and Implementation of a National, Standards-based, Multi-method Teacher Performance Assessment System in Chile. *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas*, 22, 1-30.
 16. National Board for Professional Teaching Standards (2016). What teachers should know and be able to do. Retrieved 10 December, 2017 from <http://accomplishedteacher.org/>
 17. Kyrgiridis, P., Derri, V., Emmanouilidou, K., Chlapoutaki, E., & Kioumourtoglou, E. (2014). Development of a questionnaire for self-evaluation of teacher effectiveness in physical education (SETEQ-PE). *Measurement in Physical Education and Exercise Science*, 18(2), 73-90.
 18. Rink, J. E. (2002). *Teaching Physical Education for Learning*. New York: McGraw-Hill.
 19. Erickson, F. (2006). Definition and analysis of data from videotape: Some research procedures and their rationales. *Handbook of complementary methods in education research*, 3, 177-192.
 20. Jaeger, R. M. (1993). Live vs. Memorex: Psychometric and Practical Issues in the Collection of Data on Teachers' Performances in the Classroom.
 21. National Association for Sport and Physical Education. (2007). *Physical Education Tool*. Reston, VA: Author. Retrieved from <http://www.aahperd.org/naspe/publications/TeachingTools/observepe.cfm>
 22. Danielson Group. (2013). *The Framework for Teaching Evaluation Instrument (FFT)*. Princeton, NJ: Author.
 23. Physical Education Lesson Observation Tool, by the Ministry of Education, Singapore. (2013). Copyright 2013 by the Physical Education & Sports Teacher Academy, Ministry of Education, Singapore.
 24. Vasiliadou, O., Hristakis, E., Derri, V., & Emmanouilidou, E. (2006). Elementary Physical Educator Assessment in Qualitative Teaching Elements. *Inquiries in Sport & Physical Education*, 4(1), 29-38.
 25. Derri, V., Emmanouilidou, K., Vassiliadou, O., Tzetzis, G., & Kioumourtoglou, E. (2008). Relationship between academic learning time in physical education (ALT-PE) and skill concepts acquisition and retention. *Physical Educator*, 65(3), 134.
 26. Coules, E. F., & Tzetzis, G. (2005). Systematic Observation of the Lesson of Physical Education with the Use of OSCD-PE. *Inquiries in Sport & Physical Education*, 3(3), 204 – 211.
 27. Kyrgyridis, P., Derri, V., & Kioumourtoglou, E. (2006). Factors that contribute to effective teaching in physical education: a review. *Inquiries in sport & physical education*, 4(3), 409-420.
 28. Presidential Decree 152/2013 (FEK 240/05/11/2013).
 29. Burstow, B. (2011). School effectiveness and improvement. In: J. Dillon & M. Magure (eds), *Becoming a Teacher: Issues in Secondary Education* (4th Edition) Buckingham, Open University Press
 30. Pasiardis, P. (2008). Strategic planning, innovations and assessment in education. Volume 3: Observation and assessment of the instructional work. Nicosia: Open University of Cyprus.
 31. Chen, W., Hendricks, K., & Archibald, K. (2011): Assessing pre-service teachers' quality teaching practices. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 17(1), 13-32.
 32. Van der Mars, H. (1989). Systematic observation: An introduction. In P.W. Darst, D.B. Zakrajsek & V.H. Mancini (Eds.), *Analyzing physical education and sport instruction* (pp. 3-18). Champaign, IL: Human Kinetics.
 33. Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate Data Analysis*. 7th Edition, Pearson, New York.
 34. Barlett, M. S. (1954). A note on the multiplying factors for various chi square approximations. *Journal of the Royal Statistical Society*, 16 (Series B), 296-298.
 35. Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36.
 36. Nunnally, J.C. (1978). *Psychometric theory*. 2nd Edition, McGraw-Hill, New York

37. Spector, P. (1992). Summated rating scale construction. Thousand Oaks, CA: Sage.
38. Derri, V., & Lozza Olave, E. (2006). 'Developing social values through physical and sport activities'. Comenius 2.1 (2003-2006) «Les activites Physiques et Sportives: Media de la Culture Generale et Scientifique».
39. Dyson, B. (2014). Quality Physical Education: A Commentary on Effective Physical Education Teaching. Research Quarterly for Exercise and Sport, 85, 144–152.
40. SHAPE America – Society of Health and Physical Educators. (2014). National Standards and Grade-Level Outcomes for K-12 Physical Education. Champaign, IL: Human Kinetics.
41. Sharpe, T., Brown, M., & Crider, K. (1995). The effects of a sportsmanship curriculum intervention on generalized positive social behavior of urban elementary school students. *Journal of Applied Behavior Analysis*, 28(4), 401-416.
42. Black, P., & D. Wiliam, D. (2012). "Developing a Theory of Formative Assessment." In *Assessment and Learning*, 2nd ed., edited by J. Gardner, 206–229. London: Sage.
43. Vasiliadou, O., Derri, V., Galanis, N., & Emmanouilidou, K. (2009). Training in-service physical educators to improve class time management. *Revista Internacional de Ciencias del Deporte*, 17(5), 33-43.
44. Derri, V., Vasiliadou, O., & Kioumourtoglou, E. (2015). The effects of a short-term professional development program on physical education teachers' behavior and students' engagement in learning. *European Journal of Teacher Education*. <http://dx.doi.org/10.1080/02619768.2014.947024>.