Computers & Education 128 (2019) 353–364

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Supercomputers to improve the performance in higher education: A review of the literature

Álvaro Fernández[a,*], Camino Fernández[b], José Ángel Miguel-Dávila[b], Miguel Ángel Conde[b], Vicente Matellán[a]

[a] University of Leon – Supercomputing Center of Castile and Leon (SCAYLE), Spain
[b] University of Leon, Spain

ARTICLE INFO

Keywords:
Supercomputing
Computational science
Training
Education
Systematic literature review

ABSTRACT

The use of Supercomputers is currently very widespread, constituting an essential component in many fields of science. The interest in the use of high performance computation is increasing in a wider and more diverse population of higher education students, mainly senior undergraduates and postgraduates, because the use of these infrastructures allows learners to improve their skills and the results of their training. For this reason, the demand of courses related to supercomputing increases continuously. In this paper we propose, through a wide review of primary studies, several questions that have been considered as a way of knowing the most widely-used contents in Supercomputing training. We have focused on the factors considered for improving training in Supercomputing, in order to improve the results of researchers in higher education organizations, to identify the limitations of Supercomputing training, and to provide solutions for these limitations. During the search procedure for answering research questions, 1911 studies were considered in the first selection. Through the definition of inclusion and exclusion codes in the results of searching databases, 136 published articles were studied. Finally, using quality criteria, 34 studies were identified as relevant in answering the research questions. Several factors were described, such as the way in which courses related to Supercomputing are organized, the adaptations that are currently being applied in curricula related to the students of these techniques, the use of problem-solving training and the qualification of teachers, among the most relevant ones, as well as several limitations of this type of training and the identification of solutions for these limitations. Data was collected by searching keywords related to Supercomputing training and education in the most important databases used in Computational Science, finding empirical evidence to support the positive effect of High Performance Computers (HPC) on educators and researchers. The implications of this study are: first, it provides a summary of the most relevant factors in improving training, as well as the factors that improve the results through the use of a Supercomputer; and second, it provides the analysis of the limitations found for a better performance of learners and the solutions for these limitations.

1. Introduction

The concept of a Supercomputer is dynamic, and could be described as an instance of the most highly-performing computer

* Corresponding author.
E-mail addresses: aferg@unileon.es (Á. Fernández), camino.fernandez@unileon.es (C. Fernández), jam.davila@unileon.es (J.Á. Miguel-Dávila), miguel.conde@unileon.es (M.Á. Conde), vicente.matellan@unileon.es (V. Matellán).

https://doi.org/10.1016/j.compedu.2018.10.004
Received 11 June 2018; Received in revised form 4 October 2018; Accepted 5 October 2018
Available online 10 October 2018
0360-1315/

available at a given time, and which changes as advances in computer science and technology are made (Infante, 1986). A Supercomputer is cutting-edge technology, based on large computing systems providing close to the best currently achievable sustained performance on demanding computational problems (Verschelde, 2012) based on simulations requiring high computing power (Resch, 2005). Supercomputing means running “large” projects or applications (using at least half of the available resources of a “large” system) which cannot be run on small or average sized systems (Jones & Nitzberg, 1999). The processing power of a large number of computers is distributed in the form of computer clusters and a large number of processors are used in close proximity to each other (Seagall, 2013).

The history of supercomputers dates back to the 1940s when the first High-Performance Computers (HPC) were introduced. In the 60s, there was a radical change in the way in which these infrastructures were understood, because of the highly complex simulations used in many projects and their introduction in industry, when the first parallel computers were built (González, Rosillo, Dávila, & Olivera, 2015). During recent decades, educational practice has faced a growing adoption of computer tools for the support of learning processes (Jonassen, 1999; Smeets, 2005). The fast introduction of Supercomputing technologies in education (Bote-Lorenzo et al., 2008) is a key challenge that provides a broader range of opportunities, mainly due to the possibility of managing a high volume of data and information, and has prompted educators and researchers to take a pedagogical view to promote, in the best possible way, teaching and using these infrastructures. All this creates new and critical tasks for the Higher Education System, which require extensive and complete training, using many tools for problem solving (Sung et al., 2003), in the same line of inquiry-based learning, for fostering students' curiosity and motivation (Specht, Bedek, Duval, & Held, 2013). Such infrastructures can help students, mainly senior undergraduate and postgraduate, to develop their ability, not only computing-competent users, but also non-competent users, contributing to the creation of a new generation of professionals, companies and organizations related to Science and Technology (Bethel et al., 2011). In fact, Supercomputing is considered one of the three pillars, along with theory and laboratory research, on which much of the progress of science and engineering is based. For this reason, students could employ HPC laboratories that have long developed and implemented strategies for making computational tools and analytical technologies available for use by all sectors of society (Bruce et al., 1997). In general, the use of Supercomputers extends into sectors such as Electronics, Telecommunications and Computer Science, Natural Sciences, Computer Science and Engineering, Medicine, Life Sciences, Mechanics, Economics, Social Sciences and Management. Investment in education related to the use of Supercomputers is important in many fields of knowledge, as seen in those described in the UNESCO classification: Engineering (Fabricius, Freundl, Köstler, & Rüde, 2005; Smith, 2014), Economy and Linguistics (Moses & Mariasingam, 2006) and Science (Farian, Anne, & Haas, 2008; Gaziza, Salima, Guldina, & Elena, 2012). Currently, adequate preparation of students is essential to drive innovation and support scientific research. Thus, training for a more highly skilled and knowledgeable workforce, capable of using Supercomputing technology (Coveney, McGuire, Parchment, Kenway, & Parsons, 2012) is much in demand.

The goal of the present paper is to create a visual map of how Supercomputing can improve results obtained by different users. In order to take advantage of Supercomputers, users must know how to adapt their problems to this programming paradigm. Consequently, they must first be taught how to rethink the problems and subsequently be able to try them on a Supercomputer to get better results. When we talk about students we are mainly considering, but not limited to, postgraduate students starting their research career with a Master Thesis or a PhD, as well as postdoc researchers. The study of HPC is an essential part of postgraduate programmes in Computational Science (Zarza, Lugones, Franco, & Luque, 2012). This fulfills the goals of higher education: prepare postgraduate level students for their careers both in academia and industry. This consideration means that Computational Science could be considered in the core area of many postgraduate disciplines where the use of HPC will be essential to solve complex data problems (Alexandrov & Alexandrov, 2015).

After a detailed search of the literature, no review in relation to this topic was found. This demonstrates a lack of attention paid to this field despite the importance these infrastructures have for training future researchers. In order to support the study, we consider that the classification of the results across a field, where the scope has not been described, categorized or evaluated (Hammick, 2005), helps researchers to understand the existing body of knowledge, providing a theoretical foundation for an empirical study, justifying a new contribution to the accumulated knowledge (Levy & Ellist, 2006), which is the essential condition for a field of knowledge to be classified as “scientific” (Hunter, Schmidt, & Jackson, 1982). With the purpose of appreciating the relevance of this study, we have to highlight the importance of the effective literature reviews for the advance of knowledge in general and for the understanding of the breadth of the research on a topic, the empirical evidence, the development of theories or the provision of a conceptual background for a subsequent research, identifying, finally, the topics or research domains that require a deeper analysis (Cooper, 1988) (Leedy & Ormrod, 2001). It is important to promote this study, with the aim of describing how the use of Supercomputers can support learners in the improvement of their specific skills. We also aim to enhance understanding of the core assessment concepts as they occur in this type of training. For the development of the research, a collaboration of a Group of Experts on Scientific Computing was established, charged with clarifying concepts to fulfill the goals established.

The findings will likely be of interest to students, researchers, administrators, and industry leaders as they make, more often, decisions based on computational modelization and intensive calculation. Currently the use of Supercomputers is becoming a key element in the improvement of certain disciplines of higher education. In fact, it is believed that scientists cannot be productive or efficient in terms of global research standards if they are not able to integrate Supercomputing into their research process as a required factor (Emmott & Rison, 2008). In sum, this review synthesizes the relevant literature through training in Supercomputing to provide a comprehensive analysis previously lacking in studies related to this issue. Specifically, the present study poses four research questions: (RQ1): Which are the factors considered as improvements to training in Supercomputing?; (RQ2): How can Supercomputing training improve the results of the students and/or researchers in a high education organization?; (RQ3): What is currently known about the limitations of Supercomputing training? and (RQ4)**:** What are the solutions described for solving the limitations related to Supercomputing training?

The remainder of this paper is structured as follows. In section 2, we present the research method for the review of literature, describing research questions, criteria used for inclusion/exclusion, sources of studies, search strategy, data extraction and synthesis of findings. In section 3, we present the results of the proposed systematic study. Finally, in section 4, we present conclusions, limitations and discussions of future research directions.

2. Methodology

The methodology used was the specific systematic review (Kitchenham, Budgen, & Brereton, 2015), based on a taxonomy of literature reviews in education (Cooper, 1988), which is considered true hypothesis-driven research, and in which studies are selected and combined with the use of a pre-defined protocol to reduce subjectivity and the possibility of researcher bias (Anderson et al., 2011; Kitchenham, 2004).

It is necessary to clarify that currently an important part of the training in Supercomputing is taught in specialized intensive courses, mainly as part of other subjects, especially in postgraduate and master's studies. This must be taken into account when analyzing the training and educational processes jointly.

2.1. Research questions

Once the scope of study was established, the research questions were precisely stated to guide the review (Fink, 2010). We specified 4 research questions (RQs) in order to characterize the concept of Supercomputing training and education:

- (RQ1): Which are the factors considered to be improvements to training in Supercomputing?
- (RQ2): How can Supercomputing training improve the results of the students and/or researchers in a high education organization?
- (RQ3): What is currently known about the limitations of Supercomputing training?
- (RQ4): What are the solutions described for solving the limitations related to Supercomputing training?

Addressing the research questions posed above required the collection of several opinions by an Expert Group in Supercomputing formed by the staff of the Technical Department of the Supercomputing Center of Castile and Leon (Spain). Their work was focused on the provision of accurate and unbiased technical information through their experience both in technical tasks, and in the areas of training and education based on Supercomputing applied in several fields of science and technology. Experts simultaneously discussed both strengths and weaknesses of the initial list of proposed research questions, which were ranked with the aim of making the final selection of those considered as adequate for fulfilling the goals of the study.

2.2. Data collection

2.2.1. Search strategy and data sources

The first filter in the search strategy of the investigation consisted of searches from 2005 to 2017. This period was considered because since 2005 relevant studies related to the integration of technology in education and the use of up-to-date hardware and software resources were conducted (Barbara & Donna, 2005; Bomsdorf, 2005; Lim, 2005; Mustafa, 2005; Volman, 2005). This fact could be considered as a key feature for the diffusion of technology in education, and its effectiveness and efficiency for teaching and learning purposes (Gülbahar, 2007). The second filter was the selection of articles written in English (Hammick, Dornan, & Steinert, 2010). The strategy was completed as follow: (1) look for the main search terms; (2) check the keywords in known relevant papers and (3) look for alternate forms of the terms, such as synonyms and relevant keywords. After that, we used Boolean operators OR and AND to add them into the search.

The final decision about the terms used in the search was taken in collaboration with experts in the field of Supercomputing. The following concepts were considered best for searching the responses to every RQ: “High Performance Computing course”, “High Performance Computing training”, “Quality in education” + “High Performance Computing”, “Quality in education” + “Supercomputing”, “Quality training” + “High Performance Computing”, “Quality training” + “Supercomputing”, “Supercomputing course” and “Supercomputing training”. The search of these concepts was done by consulting these Data Bases, which are considered to be the best repositories for searches about computational science: IEEExplore Digital Library, ACM Digital Library, Elsevier ScienceDirect, Scholar Google and Web of Science. In the end, 1911 papers were retrieved.

2.2.2. Inclusion/exclusion criteria

The second filter was the inclusion and exclusion criteria. It comprises the selection of the most important primary studies from literature.

Initially the selection was 1911 retrieved papers. Later, 1783 articles were excluded, and a third data filter, selecting only those papers included because they addressed educational approaches and/or teaching methods, narrowed the data pool analyzed here to 128 manuscripts, adding another 8 manually searched studies, totaling 136 manuscripts, as the most relevant for answering the research questions (See Fig. 1).

The following list shows the inclusion and exclusion criteria adopted.

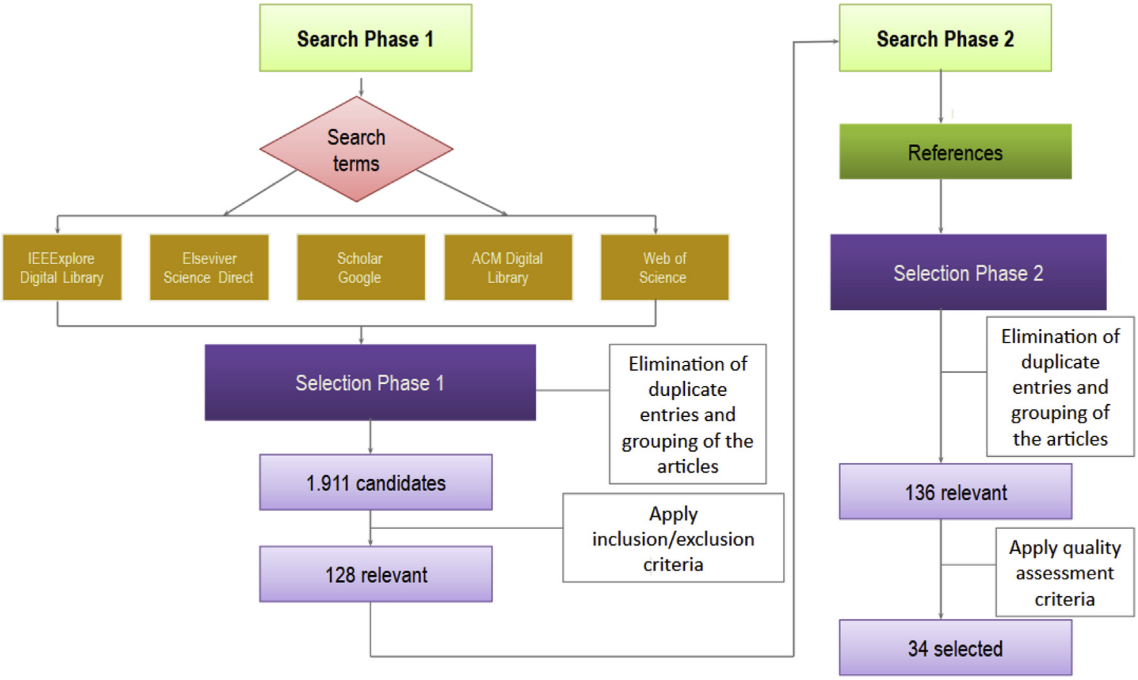

Fig. 1. Search and selection process.

Inclusion criteria:

- Studies in which the training is analyzed in depth, in the area of Supercomputing through real case studies and analysis of how it is integrated in the learning environment.
- Studies covering uses of Supercomputing in relation to the training acquired.
- Inclusion of surveys and/or questionnaires on satisfaction of students in their Supercomputing courses.
- Written in English.

Exclusion criteria:

- Scant relationship with Supercomputing (related more to other technologies such as cloud computing or similar).
- Documents that are mere descriptions of the course's content, such as information on timetable, calendar, subjects, etc.
- Documents with superficial descriptions of Supercomputing courses.
- Documents lacking references to training and Supercomputing in the title or in the abstract.
- Papers whose purpose is the development of programs or codes regarding Supercomputing.
- Studies that are repeated in more than one database.
- Duplicate studies with the same results. In this case the most complete was considered.

2.2.3. Search Phases and study selection

As seen in Fig. 1 below, the search is divided into two phases. The search terms are described as:

- Search Phase 1: involves individual searching in the electronic databases previously selected for constituting a set of candidate documents. The identification of potentially suitable articles is based on:
  (a) Determination of the set of restrictions relating to the language of publication.
  (b) Definition of sources of primary studies to be obtained.
  (c) Gathering potentially useful titles and abstracts of primary studies obtained.
- Search Phase 2: the most relevant articles are gathered in the references lists previously selected, considering other significant documents, for the goal of the study, following these steps:
  (a) Applying criteria for inclusion and exclusion to the titles and abstracts of the articles selected, to identify relevant articles that would provide the data for answering the research questions.
  (b) Extracting articles that could be potentially suitable on the basis of eligible titles and abstracts.
  (c) Assessing the quality of the selection of relevant documents to reach the final selection.

2.2.4. Assessment of item quality

Quality assessment can be performed in several ways considering that the choice of this research was based on a combination of methods due to the scope of the study, both for its technical and educational content. The works used as a basis for the final selection of the items are those specified by Wen, Li, Lin, Hu, and Huang (2012) and the categories established by Kirkpatrick (1967), later revised by education scholars (Harden, Grant, Buckley, & Hart, 1999; Belfield, Bullock, Eynon, & Wall, 2001; Tochel et al., 2009 y; Hooper, King, Wood, Bilics, & Gupta, 2013) in order to help the assessment of the quality of the articles related to educational interventions.

The analysis of the quality of works assesses the rigor and credibility of the relevant studies, in relation to their capacity and suitability for answering the research questions and their possible impact on the study's conclusions. In order to get the optimal selection of items, 10 quality assessment questions have been defined in the line of the works previously mentioned. The quality assessment criteria are summarized as follows:

- Q1: Is the paper based on research regarding training in supercomputing or is it simply a description of uses of supercomputing as a tool in a field of knowledge?
- Q2: Is there a clear statement of the goals of the research, covering participants' views on the learning experience, its organization, presentation, content, teaching methods, etc.?
- Q3: Was the context of the study well-defined in order to know how the research was being carried out?
- Q4: Was the research design appropriate to address the goals of the research, describing changes in attitudes, beliefs or perceptions as a result of the training in Supercomputing?
- Q5: Is there an adequate description of the transfer of learning to the workplace or willingness of learners to apply new knowledge and skills?
- Q6: Is there a change, due to supercomputing training, in organizational practice?
- Q7: Are the conclusions of the study clear and useful for the object of the research?
- Q8: Are the limitations of the study clearly shown?
- Q9: Is there a clear description of the acquisition of concepts, procedures and principles?
- Q10: Is the study of value for analyzing the benefit for clients/users as a direct result of the educational program?

Each question has only three possible answers: "Yes", "Partly", or "No". These three answers are scored as follows: "Yes" = 1, "Partly" = 0.5, and "No" = 0. For a given study, its quality score is computed by summing up the scores of the answers to the QA questions.

Finally, to guarantee that only the most relevant items had an acceptable quality, a punctuation of every study was conducted, in order to analyze the reliability of the results of the review. The requirement was to receive a rating higher than 5.5 (medium level - Kirkpatrick (1967) and later works based on this one). As a result, 34 of the total 136 articles were accepted (24.98%) and 102 (75.02%) were rejected, as having a quality score below 5 in the second selection phase (see Table 1).

3. Results

In the present study, 34 manuscripts published between 2005 and 2017 (see Fig. 2) have been identified as relevant to answering the questions of the research.

The sources of literature on this research topic are widely spread over a range of many publications with different characteristics (as seen in Table 2), picked out in 27 different publications. The analysis of the information mentioned shows the multidisciplinary nature of the study.

The answers to each of the research questions raised in previous lines are described below:

3.1. RQ1: which are the factors considered that improve supercomputing training?

The factors considered to improve Supercomputing training, extracted from the review, are the following:

a. Student explores real scientific problems through hands-on experience (S2, S4, S7, S9, S10, S12, S13, S15, S19, S25, S29), with practical examples of parallel systems and programs (S5, S16, S25), learning the most advanced functionalities of computers,

Table 1
Quality punctuation of relevant studies.

Quality level	Nº of studies	Percent
VERY HIGH (8.5 ≤ score ≤ 10)	3	2.20%
HIGH (7 ≤ score ≤ 8)	11	8.08%
MEDIUM (5.5 ≤ score ≤ 6.5)	20	14.70%
LOW (3 ≤ score ≤ 5)	12	8.82%
VERY LOW (0 ≤ score ≤ 2.5)	90	66.20%
TOTAL	**136**	**100.00%**

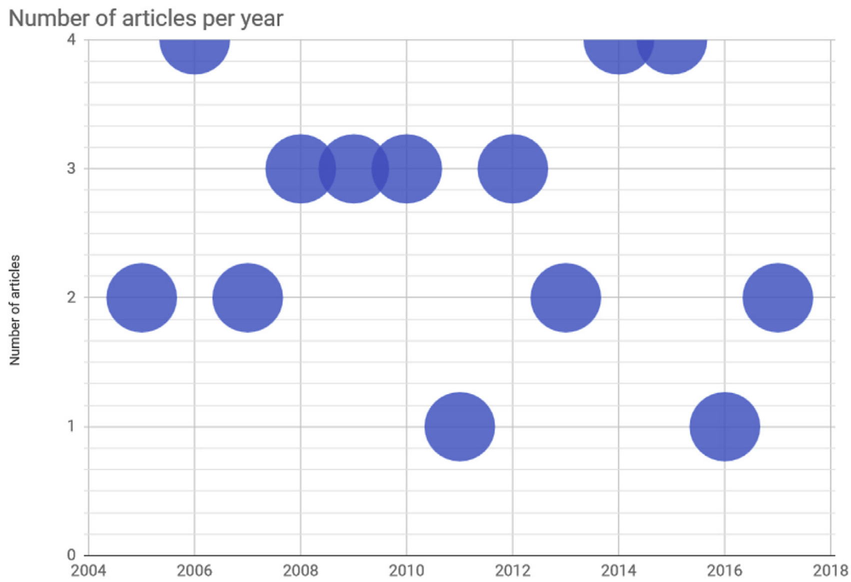

Fig. 2. Number of publications per year.

Table 2
Selected studies by publication type.

Type	Nº of studies	Percent
Journals	14	41.18%
Age (Omaha)	1	2.94%
Biotechnology progress	1	2.94%
International Journal of Engineering Education	1	2.94%
Journal of Computing Sciences in Colleges	2	5.88%
Journal of Parallel and Distributed Computing	1	2.94%
Lecture Notes in Computer Science	1	2.94%
Procedia Computer Science	6	17.65%
Procedia-Social and Behavioral Sciences	1	2.94%
Conferences	**15**	**44.12%**
ACM Transactions on Computing Education (TOCE)	1	2.94%
ACM annual conference on Innovation and technology in computer science education	1	2.94%
Computational Science–ICCS 2009	1	2.94%
Computing in Science and Engineering	1	2.94%
DoD HPCMP Users Group Conference	1	2.94%
5th Global Congress on Engineering Education	1	2.94%
IEEE Transactions on Education	2	5.88%
IEEE 26th International Parallel and Distributed Processing Symposium	1	2.94%
IEEE High Performance Extreme Computing Conference (HPEC ‘16)	1	2.94%
International Symposium Computer Science and Computational Technology	1	2.94%
Second International Conference on Education Technology and Training	1	2.94%
TeraGrid Conference: Extreme Digital Discovery	1	2.94%
TG08-TeraGrid Conference	1	2.94%
Workshop on Education for High-Performance Computing	1	2.94%
Other	**5**	**14.71%**
Web resource	5	14.71%

linking theory with practical problem solving.

b. The content of the courses is normally adapted by including, as a core component, parallelism modules with an easy-to-use high performance cluster system in teaching-oriented computer science curriculum (S1,S27).
c. Supercomputing encourages the exploration of the quantitative characterization of a program's performance on a variety of platforms (S15, S18).
d. Supercomputing teachers have positive pedagogic characteristics and specific professional competences as scientists and researchers (S11), and are qualified to help the students with complex projects (S17).
e. It is not necessary for all users to have skills in the use of a Supercomputer. In fact, visualization tools help students to understand in depth the nature of particular scientific problems (S21) through the learning of parallel computing (S31), thus approaching the solving of these problems (S16, S21) by manipulating large datasets easily (S20).

f. Supercomputing training aims to prepare people, to a high standard, across a continuum process to achieve an adequate level of knowledge and proficiency to deal with the largest computing systems available (S2).
g. Training in Supercomputing builds a model-learning methodology that engages students in different types of problem solving (S34).

3.2. RQ2: how can supercomputing training improve the results of the students and/or researchers in a high education organization?

Currently, high performance computing (HPC) is used in higher education to train people on how to solve an array of problems, making it a critical technology in fields such as science, engineering, finance and research and development. Supercomputing may be viewed as the intersection between the frontiers of science and computing, and introduces students to core concepts such as large scale simulation.

In the review, we identified some characteristics of the Supercomputing training, that are directly related to the improvement of the research results in higher education organizations. They are the following:

a. Supercomputing enables users to become competent to work deeper on different problems, closely related with their research fields (S4, S15) through reproducible experiments (S5) which use codes of increasing difficulty to illustrate and experiment through the main programming models (S30).
b. The work with a Supercomputer is cutting-edge in research (S8), and increases the interest in performing independent research work (S29).
c. The availability of network infrastructure, based on a Supercomputing facility (S26, S33), enables the collaboration among research groups distributed geographically (S33), and allows participation in projects of multidisciplinary areas (S8, S19).
d. Supercomputing training leads to a deeper understanding of math and science concepts (S34) that help the researchers in the use of mathematical models (S3, S26, S34) to describe a problem in a specific research field, extract conclusions and evaluate performance on the proposed problem.
e. Courses based on how to run an HPC-related project are an important way of training in Supercomputing (S18), by incorporating experimental components such as models for the applications (S22), simulations and visualizations (S23).
f. Implanting research training using a Supercomputer can strengthen students' enthusiasm for discovery and creativity, broadening their knowledge (S13).
g. As computational capabilities increase, users perform their work faster, finding solutions to more challenging problems, and implementing more research models in the same amount of time (S20).
h. High quality visualization can lead to a new level of understanding of how the data can be partitioned and processed parallelly to simplify the interpretation of the research results (S21). It also represents a mechanism to strengthen the experimental-computational synergy (S3).
i. Numerical simulation afforded by a Supercomputer enables the study of complex systems and natural phenomena that would be too expensive, or even impossible to study by direct experimentation (S14).
j. Simulations and data design practice generate large volumes of data (S34) and prepare students for future research in the area of very large database systems (S1), covering a variety of applications (S22).
k. Supercomputing contributes to the invention of new tools and new philosophies of work which enhance learning about computing (S9).
l. Knowledge in Supercomputing encourages the designing of efficient programs that provide students with a crucial understanding regarding the importance of underlying algorithms in the implementation of larger programs (S9).
m. Feedback on the performance of the students in a Supercomputing course helps guide the training, maximizing the time that they use for their research, which is highly desirable on a research project due to its importance (S17).

3.3. RQ3: what is currently known about the limitations in supercomputing training?

Considering the response of the research questions of this study, we identified some limitations that must be mentioned. The findings obtained in RQ1 and RQ2 show a wide description of the different roles of Supercomputer uses (students, teachers, experts, researchers, etc.), access to these infrastructures and the way of developing courses. With these in mind as well as better organization and understanding of the information, the limitations were grouped into four types (see Table 3), depending on the dimension involved: Students, Teachers, HPC infrastructures and Methodology and development of the training.

3.4. RQ4: what are the described solutions for solving the described limitations in relation to the training in supercomputing?

Following the same structure of the previous question, for a better organization and understanding of the information, the solutions of the limitations were grouped into four types (see Table 4), depending on the dimension involved: Students, Teachers, HPC infrastructures and Methodology and development of the training. Considering the response of the research questions of this study, we identified some limitations that must be mentioned in every type.

Table 3
Limitations in supercomputing training.

Groups	Limitations	References
Students	Insufficient educational background, with different degrees and/or specialization, mainly in programming languages.	S4, S9,S17,S23,S26,S28
	Constraints of the typical high school curriculum, in relation with Supercomputing.	S12,S17,S25
	Lack of basic skills for designing, implementing and operating in Supercomputing environments, especially for developing parallel software applications to achieve best efficiency, performance and scalability.	S1,S15,S29
	Need of mathematical sophistication to complete an advanced Supercomputing course to use mathematical models properly.	S32
Teachers	Substantial effort and dedication for a successful deployment of the tools in order for the course to be properly prepared, which means less time to prepare the content.	S13
	Need of complementing the training of the teachers, especially in parallel computing, through the engagement of experts.	S2,S11
	Lack of experience in the field of HPC.	S1
	Difficulty for the recruitment of enough tutors to maintain a reasonable tutor/trainee ratio.	S2
HPC Infrastructure	Rapid changes in the systems and software currently and in the coming years, with a continuous growth in the volume of data managed.	S2,S17,S18,S20,S27,S29
	Need to predict the utility and relevance of forthcoming technologies in both the short and the long term.	S28
	Low availability of Supercomputing facilities due to the high costs.	S8
	Lack of funds to purchase expensive parallel computers.	S29, S31
	Users are not adequately served by their local Supercomputing Center.	S1,S28
	Difficulties in coordinating the Supercomputing sites of the diverse training facilities to handle a large number of inexperienced students.	S2,S30
	Problems with the use of the interface of the operating system.	S13
Methodology and development of the training	Lack of use of on-line learning technologies.	S2
	Existing training material is inadequate for training delivery.	S2,S20,S28
	Need to focus the undergraduate courses on the skills for improving training in Supercomputing technology in nearly all areas.	S1,S17,S28
	Time constraint, both for students and teachers, mainly for the implementation of numerical algorithms.	S13,S20,S25,S28,S30
	Lack of training of writing programs for Supercomputing machines to see clearly the effects of parallel programming.	S24,S25
	Necessity to effectively employ massively parallel high-performance computing machines in scientific computations.	S29

4. Discussion

Our findings suggest that including Supercomputers in learning processes has provided a wide range of new opportunities for learners, mainly due to the improvement in the quality of training obtained, which provides better results in the practical cases through real-life situations. At the same time, the study identifies the learning needs and devises strategies of how to meet those needs. However, as seen in the previous section, there is a lack in the literature concerning the degree in which the use of a Supercomputer impacts the final result of a research project. In order to find out how the characteristics of this infrastructure can improve the learning process pedagogically, a systematic effort and more empirical studies are needed, as seen in other analyses in relation with the use of technology in learning (Mikropoulos & Natsis, 2011). Although research pointed out the need for better previous training of the students, this review provides other actions to be deployed, such as new methods of teaching or the need of an easier access to the Supercomputing infrastructures.

In order to deepen the research, the study implemented several steps to support the quality of the results of the analysis done. The interpretation of the final results can be generalized as goals, actions and strategies for teaching and guiding learners in their learning process, taking into consideration that without a critical and rational discussion about the opportunities of new technologies in education, little progress can be made towards a real debate about the impact of these infrastructures.

4.1. Outcomes

The main outcomes which respond to the RQ raised at the beginning of this study are summarized below.

- (RQ1) The factors considered for improving training in Supercomputing show that the most relevant one is the way in which the course is organized and the adaptations that are currently being applied in the curricula related to the students of these techniques, using examples with a problem-based model, showing how to solve real problems with access to a Supercomputer. Another important factor is the qualification of teachers, essential for a better performance, and finally the use of tools, as visualization facilities that provide an easy understanding for all the students.
- (RQ2) As a conclusion to this question, it is found that the possibility of working in a network of multidisciplinary research groups clearly enhances the performance of the researchers. The contacts provided by these groups open the possibility of participating in

Table 4
Solutions to limitations in supercomputing training.

Groups	Solutions to Limitations	References
Students	Implementation of a wide number of introductory courses of Supercomputing for students without a computational background.	S9,S19,S32
	Reformation of the curriculum system (including parallel computing topics at early state), teaching methods, type of examination and the model of theoretical courses.	S13,S14,S29
	Attracting students in several ways such as summer workshops, using robots, etc.	S10
Teachers	Increase in the number of teaching assistant contact hours.	S19
	Adequate teaching environment for providing outlines of the basic vocabulary and concepts related to Supercomputing.	S20
	Use of face-to-face training sessions delivered by experts in every field, which means widespread acceptance, and the improvement of the transfer of knowledge.	S2,S28
	Team work environment in Supercomputing training, for a better exchange of ideas with the coordination of the tutor.	S7,S8,S25,S27, S31
	Transmission of the skills to perform repeatability experiments.	S5
HPC Infrastructure	Supercomputing systems tend to be shared by multiple users, interacting remotely, representing a solution of the high cost of using a Supercomputer.	S20,S29
	Implementation of open source software and operating systems to reduce the cost and for an easy use, providing a user-friendly interface.	S1,S6,S12,S30
	New and easy-to-use software makes computer simulation accessible to a wider group of people.	S34
	Use of optimal and feasible high computing power with relatively low cost, through alternatives to Supercomputing platforms, such as the cluster-based computing system for its flexibility.	S1
	Use of grid technology in the educational environment helps the students in the use of complex models and simulations.	S22,S30
	The progress of networks technologies has permitted the use of local networks of enterprises and training rooms for parallel computing.	S11
Methodology and development of the training	Development of a timely and quality "study pack" of materials, such as tutorials, textbooks via web, user guides, electronic slides, interactive computer-based courses, books and journals, that teach Supercomputing in a rapidly changing environment.	S2,S6,S7,S8,S9,S12,S13,S14,S17,S18,S25,S28,S30
	Use of pedagogical tools which enhance trust in students' work.	S5
	Provision of an easy access to a real Supercomputing system inspires and motivates students, giving the opportunity to implement algorithms using different architectures and parallel-programming models, which enable students to make objective performance comparisons.	S12,S20,S30
	Collaboration of Supercomputing Centers in the provision of a platform to access the resources easily, and with the support from different institutions, which ensure that training needs and expectations are met for all users, without interference to the research activities of the supercomputing center.	S2,S7,S10,S14S28,S30
	Wide variety of training programs in Supercomputing with different sets of subjects offered to complement knowledge, according to the expected skills at the conclusion of the training program.	S14,S26
	Benchmarking to improve training, by describing a taxonomy for general benchmark experiments, and accessing (usually for benchmarking comparisons) an even wider class of machines.	S5,S8,S14
	Development of research-oriented teaching method, to achieve a superior teaching quality, and to involve students in the research oriented learning, based on a high-performance computing research-oriented learning environment.	S13, S31

research and development projects, using techniques of numerical simulation, which contribute to solving the processes on which research is based in a short period of time, and allowing a better interpretation of the results.

- (RQ3) The limitations of the study are mainly focused on the insufficient background of the students in certain specialties. This is caused by many factors such as constraints in the curriculum, rapid changes in this field, sophistication of the mathematical models, and even other questions such as a general lack of experience in the field of Supercomputing, which affects questions such as recruitment of teachers, use of Supercomputing facilities, and the time constraints of the courses. Another important limitation is Supercomputing infrastructure cost, which causes difficulties in accessing Supercomputing resources.
- (RQ4) The solutions proposed for the described limitations in the study focus on the improvement of training of students previous to a Supercomputing course in order to have the basic knowledge to attend the course, with complete and updated material and with adequate access to a Supercomputer.

4.2. Limitations

During the course of the literature review, the greatest difficulty encountered was the heterogeneity in the type of articles, such as in content related with the typology of training. As seen in the answer to the RQs and also in the study of the courses, it is necessary to improve some basic aspects for an optimal training, as provision of previous basic training for those students who, despite being mainly but not limited to postgraduates, have an insufficient background in Supercomputing techniques, the adequate qualification of teachers and a wider use of Supercomputing facilities. Our hope is also that the present study will encourage researchers and institutions in general to contribute to the development of agreements which address the challenges of using a Supercomputer as part of the fundamental knowledge for scientific uses through the use of a real HPC infrastructure.

5. Conclusions and future work

This study provides yields two main outcomes in relation to the state-of-art of Supercomputing training. First, it provides a summary of the most relevant factors in improving training, as well as the factors that improve the results through the use of a Supercomputer. The second outcome of the review is the analysis of the limitations found of a better performance of learners and the solutions for these limitations. The intention of this instrument is to allow the design of activities and the deployment of strategies which gradually foster the different components of training through the use of these infrastructures. Moreover, it helps to raise awareness of the potential of Supercomputers to enable learners' a best performance.

From a practical perspective the results of this study will allow detecting the main problems or gaps in Supercomputing learning initiatives by taking into account our research questions. With this information it is possible to define learning plans better and more accurately, thus helping to increase both researchers' and advanced students' performance.

Finally, future work should be focused, comprehensively and simultaneously, on analyzing new methods and tools to evaluate the quality of Supercomputing training, and on exploring a full range of fields where the use of Supercomputers will be helpful in enhancing the results of research projects. As seen in the study, personalized introductory lessons according to the previous knowledge of every student, the possibility of working online through new scientific networks of communications, the use of real Supercomputers for optimizing workflows, and more practical sessions should be the key factors of courses related to the success of Supercomputing training in the future.

Funding

The work described in this article has been partially supported by Spanish Ministry of Economy and Competitiveness under research project ECO2015-63880-R and by Junta de Castilla y Leon, under grant LE028P17.

Appendix A. Bibliographic details of the papers that were assessed using quality assessment criteria.

S1	Abuzaghleh, O., Goldschmidt, K., Elleithy, Y., & Lee, J. (2013). Implementing an affordable high-performance computing for teaching-oriented computer science curriculum. *ACM Transactions on Computing Education (TOCE)*, vol. 13, no. 1.
S2	Armosky, B., Brown, S., Drummond, T., Ferguson, J., Gerber, R., Hacker, T. J., & Traxler, K. (2007). HPC University. In *TG08-TeraGrid Conference, Madison, WI*.
S3	Bergeron, K., Cummings, R., Decker, C. R., Freeman, M. J., Hoke, C. C., Seidel, J., & MCDaniel, D. M. (2008). An Undergraduate Computational Aerodynamics Curriculum. In *DoD HPCMP Users Group Conference, 2008. DOD HPCMP UGC* (pp. 18–24). IEEE.
S4	Bernabé, G., Cuenca, J., García, L. P., Giménez, D., & Rivas-Gómez, S. (2014). A High Performance Computing Course Guided by the LU Factorization. *Procedia Computer Science*, *29*, 1446–1457.
S5	Burkhart, H., Guerrera, D., & Maffia, A. (2014). Trusted high-performance computing in the classroom. In *Proceedings of the Workshop on Education for High-Performance Computing* (pp. 27–33). IEEE Press.
S6	Cuenca, J., & Giménez, D. (2015). *TCPP Curriculum in Parallel Programming courses of the Degree in Computer Science at the University of Murcia, Spain*. Retrieved from: http://dis.um.es/~domingo/14/EarlyAdopter/paper.pdf.
S7	Czarnul, P. (2014). Teaching high performance computing using beesycluster and relevant usage statistics. In *Procedia Computer Science*,vol. 29.
S8	Fabricius, U., Freundl, C., Köstler, H., & Rüde, U. (2005). High performance computing education for students in Computational Engineering. In *Lecture Notes in Computer Science*, vol. 3515, no. II, pp. 27–35.
S9	Farian, H., Anne, K. M., & Haas, M. (2008). Teaching high-performance computing in the undergraduate college CS curriculum. *Journal of Computing Sciences in Colleges*, *23*(3), 135–142.
S10	Fernández Slezak, D., Turjanski, P. G., Montaldo, D., & Mocskos, E. E. (2010). Hands-On Experience in HPC With Secondary School Students. *IEEE Transactions on Education*, *53*(1), 128–135.
S11	Gaziza, B., Salima, N., Guldina, K., & Elena, K. (2012). Parallel Computing in Training of Informatics Teachers. *Procedia-Social and Behavioral Sciences*, *51*, 883–887.

S12 Holmes, V., & Kureshi, I. (2015). Developing High Performance Computing Resources for Teaching Cluster and Grid Computing courses. *Procedia Computer Science*, *51,* 1714–1723.
S13 Liu, X., Guo, H. M., & Meng, C. G. (2009). Design and Implementation of HPC-Based Research-Oriented Learning Environment for Structural Chemistry. In *Education Technology and Training, 2009. ETT'09. Second International Conference on* (pp. 267–270). IEEE.
S14 Moses, G. A., & Mariasingam, M. A. (2006). The development of a high quality Internet-based degree programme in computational science. 5th *Global Congress on Engineering Education*, 146–150, Brooklyn, New York, USA, 17–21.
S15 Mullen, J. S., Arcand, W., Bestor, D., Bergeron, W., Byun, C., Gadepally, V., … & Michaleas, P. (2016). Designing a New High Performance Computing Education Strategy for Professional Scientists and Engineers. IEEE High Performance Extreme Computing Conference (HPEC '16) Twentieth Annual HPEC Conference
S16 Mullen, J., Byun, C., Gadepally, V., Samsi, S., Reuther, A., & Kepner, J. (2017). Learning by doing, High Performance Computing education in the MOOC era. *Journal of Parallel and Distributed Computing*, *105*, 105–115.
S17 Munshi, R., Coalson, R. D., Ermentrout, G. B., Madura, J. D., Meirovitch, H., Stiles, J. R., & Bahar, I. (2006). An introduction to simulation and visualization of biological systems at multiple scales: a summer training program for interdisciplinary research. *Biotechnology progress*, *22*(1), 179–185.
S18 Radenski, A. (2012). Integrating data-intensive cloud computing with multicores and clusters in an HPC course. In *Proceedings of the* 17th *ACM annual conference on Innovation and technology in computer science education - ITiCSE '12*.
S19 Ravindran, A. Tolley, P., & Mukherjee, A. (2009). Introducing reconfigurable computing in the undergraduate computer engineering curriculum. *Age (Omaha)*, vol. 14, p. 1, 2009.
S20 Richards, M., & Lathrop, S. (2011). A training roadmap for new HPC users. In *Proceedings of the 2011 TeraGrid Conference: Extreme Digital Discovery* (p. 56). ACM.
S21 Ryabinin, K., & Chuprina, S. (2015). Using Scientific Visualization Tools to Bridge the Talent Gap. *Procedia Computer Science*, *51*, 1734–1741.
S22 Shiflet, A. B., & Shiflet, G. W. (2010). Simulating the formation of biofilms in an undergraduate modeling course. *Procedia Computer Science*, *1*(1), 895–901.
S23 Shiflet, A. B., Shiflet, G. W., & Sanders, W. E. (2013). Undergraduate module on computational modeling: Introducing modeling the cane toad invasion. *Procedia Computer Science*, *18*, 1429–1435.
S24 Smith, J. (2014). *High Performance Computing and Mathematics*. On-line: http://www.math.iastate.edu/pdfs/MM2014.pdf math.iastate.edu
S25 Smith, J., & Wolffe, G. (2007). Introducing AP computer science students to high-performance computing. *Journal of Computing Sciences in Colleges*, *23*(1), 70–76.
S26 stainsby, H., Muresano, R., Fialho, L., González, J. C., Rexachs, D., & Luque, E. (2009). Teaching model for computational science and engineering programme. In *Computational Science–ICCS 2009* (pp. 34–43). Springer Berlin Heidelberg.
S27 Stewart, K. A. (2006). Template for a Group-Work Paradigm in an Undergraduate Supercomputing Course. On-line: https://www.researchgate.net/publication/2606153_A_Template_for_a_Group-Work_Paradigm_in_an_Undergraduate_Supercomputing_Course.
S28 Stitt, T., & Robindon, T. (2008). A survey on training and education needs for petascale computing. On-line: http://www.prace-project.eu/IMG/pdf/D3-3-1_document_final.pdf.
S29 Stojanovic, N., & Milovanovic, E. (2015). Teaching Introductory Parallel Computing Course with Hands-On Experience. *International Journal of Engineering Education*, *31*(5), 1343–1351.
S30 Touriño, J. Martín, M. J. Tarrío, J., & Arenaz, M. (2005). A grid portal for an undergraduate parallel programming course. *Education, IEEE Transactions on*, *48*(3), 391–399.
S31 Varma, A., Keswani, Y., Bhatnagar, Y., & Chaudhury, B. (2017). Let's HPC: A web-based interactive platform to aid High Performance Computing education. arXiv preprint arXiv:1701.06356. https://arxiv.org/abs/1701.06356
S32 Vuduc, R., Czechowki, K., Chandramowlishwaran, A., & Choi, J. W. (2012). Courses in high-performance computing for scientists and engineers. In *Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2012 IEEE* 26th *International* (pp. 1335–1340). IEEE.
S33 Xiaotao, Y., Aili, L., & Lin, Z. (2010). Research of high performance computing with clouds. In *Proc. International Symposium Computer Science and Computational Technology*, pp. 289–293.
S34 Yasar, O., Maliekal, J., Little, L. J., & Jones, D. (2006). A computational technology approach to education. *Computing in Science and Engineering*, *8*(3), 76.

References

Alexandrov, N., & Alexandrov, V. (2015). Computational science research methods for science education at PG level. *Procedia Computer Science, 51*, 1685–1693.
Anderson, L. M., Petticrew, M., Rehfuess, E., Armstrong, R., Ueffing, E., Baker, P., et al. (2011). Using logic models to capture Complexity in systematic reviews. *Research Synthesis Methods, 2*, 33–42.
Barbara, M., & Donna, V. (2005). Learner-centered framework for E-learning". *Teachers College Record, 107*(8), 1582–1600.
Belfield, H. T., Bullock, A., Eynon, R., & Wall, D. (2001). Measuring effectiveness for best evidence medical education: A discussion. *Medical Teacher, 23*(2), 164–170.
Bethel, E. W., Van Rosendale, J., Southard, D., Gaither, K., Childs, H., Brugger, E., et al. (2011). Visualization at supercomputing centers: The tale of little big iron and

the three skinny guys. *IEEE Computer Graphics and Applications, 31*(1), 90–95.
Bomsdorf, B. (2005). Adaptation of learning spaces: Supporting ubiquitous learning in higher distance education. *Dagstuhl seminar proceedings. Schloss dagstuhl-leibniz-zentrum fr informatik*.
Bote-Lorenzo, M. L., Gómez-Sánchez, E., Vega-Gorgojo, G., Dimitriadis, Y. A., Asensio-Pérez, J. I., & Jorrín-Abellán, I. M. (2008). Gridcole: A tailorable grid service based system that supports scripted collaborative learning. *Computers & Education, 51*(1), 155–172.
Bruce, B. C., Carragher, B. O., Damon, B. M., Dawson, M. J., Eurell, J. A., Gregory, C. D., ... Potter, C. S. (1997). Chickscope: An interactive MRI classroom curriculum innovation for K-12. *Computers & Education, 29*(2–3), 73–87.
Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society, 1*(1), 104–126. https://doi.org/10.1007/BF03177550.
Coveney, P., McGuire, S., Parchment, O., Kenway, R., & Parsons, M. (2012). *UK participation in high performance computing (HPC) at the european level - publications*. https://www.gov.uk/government/publications/uk-participation-in-high-performance-computing-hpc-at-the-european-level.
Emmott, S., & Rison, S. (2008). Towards 2020 science. *Science in Parliament, 65*(4), 31–33.
Fabricius, U., Freundl, C., Köstler, H., & Rüde, U. (2005). High performance computing education for students in Computational Engineering. *Lecture notes in computer science: Vol. 3515*, (pp. 27–35). II.
Farian, H., Anne, K. M., & Haas, M. (2008). Teaching high-performance computing in the undergraduate college CS curriculum. *Journal of Computing Sciences in Colleges, 23*(3), 135–142.
Fink, A. (2010). *Conducting research literature reviews: From the internet to paper* (3nd ed.). Sage Publications.
Gaziza, B., Salima, N., Guldina, K., & Elena, K. (2012). Parallel computing in training of informatics teachers. *Procedia-Social and Behavioral Sciences, 51*, 883–887.
González, Á. F., Rosillo, R., Miguel-Dávila, J.Á., & Matellán, V. (2015). Historical review and future challenges in supercomputing and networks of scientific communication. *The Journal of Supercomputing, 71*(12), 4476–4503.
Gülbahar, Y. (2007). Technology planning: A roadmap to successful technology integration in schools. *Computers & Education, 49*(4), 943–956.
Hammick, M. (2005). A BEME review: A little illumination. *Medical Teacher, 27*(1), 1–3.
Hammick, M., Dornan, T., & Steinert, Y. (2010). Conducting a best evidence systematic review. Part 1: From idea to data coding. BEME Guide No. 13. *Medical Teacher, 32*(1), 3–15.
Harden, M., Grant, J., Buckley, G., & Hart, I. R. (1999). BEME Guide No. 1: Best evidence medical education. *Medical Teacher, 21*(6), 553–562.
Hooper, B., King, R., Wood, W., Bilics, A., & Gupta, J. (2013). An international systematic mapping review of educational approaches and teaching methods in occupational therapy. *British Journal of Occupational Therapy, 76*(1), 9–22.
Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies.* Beverly Hills: Sage.
Infante, E. F. (1986). *An agenda for improved evaluation of supercomputer performance: A report.* National Academies Press.
Jonassen, D. H. (1999). *Computers as mindtools for schools: Engaging critical thinking.* NJ: Prentice Hall, Inc.
Jones, J. P., & Nitzberg, B. (1999). Scheduling for parallel supercomputing: A historical perspective of achievable utilization. *Workshop on job scheduling strategies for parallel processing* (pp. 1–16). Berlin, Heidelberg: Springer.
Kirkpatrick, D. I. (1967). Evaluation of training. In R. Craig, & I. Mittel (Eds.). *Training and development handbook*. New York: McGraw Hill.
Kitchenham, B. (2004). *Procedures for performing systematic reviews. Joint technical report.* Software Engineering Group, Keele University, and Empirical Software Eng., Nat'l ICT Australia.
Kitchenham, B. A., Budgen, D., & Brereton, P. (2015). *Evidence-based software engineering and systematic reviews*, *Vol. 4*. CRC Press.
Leedy, P. D., & Ormrod, J. E. (2001). *Practical research: Planning and design* (7th ed.). Upper Saddle River, NJ: Merrill Prentice Hall.
Levy, Y., & Ellist, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Information Sciences, 9*, 181–212.
Lim, C. P. (2005). Online learning in higher education: Necessary and sufficient conditions. *International Journal of Instructional Media, 32*(4), 323.
Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education, 56*(3), 769–780.
Moses, G. A., & Mariasingam, M. A. (2006). *The development of a high quality Internet-based degree programme in computational science. 5th Global Congress on Engineering Education*, *Vols. 146–150*, New York, USA: Brooklyn17–21.
Mustafa, K. O.Ç. (2005). Implications of learning theories for effective technology integration and pre-service teacher training: A critical literature review. *Journal of Turkish Science Education, 2*(1), 2.
Resch, M. M. (2005). Remarks on supercomputing in Germany. *Praxis der Informationsverarbeitung und Kommunikation, 28*(4), 233–237.
Seagall, R. S. (2013). Computational dimensionalities of global supercomputing. *Journal of Systemics, 11*(9), 75–86.
Smeets, E. (2005). Does ICT contribute to powerful learning environments in primary education? *Computers & Education, 44*(3), 343–355.
Smith, J. (2014). *High performance computing and mathematics.* On-line: http://www.math.iastate.edu/pdfs/MM2014.pdf math.iastate.edu.
Specht, M., Bedek, M., Duval, E., & Held, P. (2013). *WESPOT: Inquiry based learning meets learning analytics.* Retrieved from: http://bg-openaire.eu/handle/10867/130.
Sung, N. S., Gordon, J. I., Rose, G. D., Getzoff, E. D., Kron, S. J., Mumford, D., et al. (2003). Educating future scientist. *Science, 301*, 1485.
Tochel, C., Haig, A., Hesketh, A., Cadzow, A., Beggs, K., Colthart, I., et al. (2009). The effectiveness of portfolios for post-graduate assessment and education: BEME guide No. 12. *Medical Teacher, 31*(4), 299–318.
Verschelde, J. (2012). *Introduction to supercomputing. Computing memory coalescing techniques*.
Volman, M. (2005). A variety of roles for a new type of teacher educational technology and the teaching profession. *Teaching and Teacher Education, 21*(1), 15–31.
Wen, J., Li, S., Lin, Z., Hu, Y., & Huang, C. (2012). Systematic literature review of machine learning based software development effort estimation models. *Information and Software Technology, 54*(1), 41–59. http://www.sciencedirect.com/science/article/pii/S0950584911001832.
Zarza, G., Lugones, D., Franco, D., & Luque, E. (2012). An innovative teaching strategy to understand high-performance systems through performance evaluation. *Procedia Computer Science, 9*, 1733–1742.