View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Firenze University Press: E-Journals

Form@re
Open Journal per la formazione in rete

articoli

La Flipped Classroom funziona? Analisi critica di evidenze empiriche sulla sua efficacia per l'apprendimento

Does Flipped Classroom work? Critical analysis of empirical evidences on its effectiveness for learning

Juliana E. Raffaghelli[a]

[a] *Università degli Studi di Firenze, julianaelisa.raffaghelli@unifi.it*

Abstract

Questo articolo ha tentato di sintetizzare la ricerca di qualità più elevata per vedere in che misura i principi della Flipped Classroom (FCM) sono guidati da dati rigorosi che confermano e ampliano i benefici di apprendimento frequentemente associati al modello dai suoi sostenitori. Dopo aver analizzato 17 rassegne della letteratura sul tema, si è giunti alla conclusione che gli approcci più rigorosi della ricerca sono molto cauti sull'impatto del modello, mentre le revisioni che adottano criteri più ampi tendono ad assumere posizioni più entusiaste. Inoltre, la maggior parte delle ricerche esistenti affronta la FCM nell'istruzione superiore, lasciando lo studio del livello primario e secondario scoperti. Più di 232 studi (che comprendono 9809 studenti) portano a individuare tre cluster tematici che guidano future ricerche e pratiche: un cluster metodologico, uno pedagogico e uno organizzativo. Inoltre, l'approccio *visible learning* è stata utilizzato per supportare alcune raccomandazioni di progettazione e l'implementazione della FCM.

Parole chiave: flipped classroom; classe capovolta; meta-analisi; revisione sistematica; efficacia per l'apprendimento

Abstract

This paper attempts to synthesize the highest quality evidence to see to which extent the principles of the Flipped Classroom (FCM) are driven by rigorous data that confirm and expand the learning benefits frequently claimed by the FCM's advocates. After producing a synthesis of 17 reviews of the literature on the issue, it has been found that most rigorous research approaches are cautious about the impact of the model, while the broader review approaches embrace more enthusiastic positions. Moreover, most existing research deals with FCM in higher education, with K12's effects less covered. More than 232 studies (encompassing 9,809 students) lead to identify three thematic clusters guiding future research and practices: a methodological, a pedagogical and organizational cluster. Furthermore, a *visible learning approach* has been used to support some recommendations for FCM's learning design and deployment.

Keywords: flipped classroom; meta-analysis; systematic review; learning effectiveness

© 2017 Author(s). Open Access article distributed under CC BY-NC-ND 4.0

1. Introduction

In 2009, Sams and Bergmann (2013) proposed a systematic approach called "The flipped classroom method" (from now on FCM). It was based on well-known pedagogical principles in education research and practice (clear advance organizers for learning, active learning and peer-learning as mean to promote students' engagement and high level skills' learning). To this formula, it was added the adoption of digital technologies either to support the advance organizers (in the forms of videos and quizzes for self-assessment) and active learning (mostly through students' response systems in class). The novelty of the FCM was to emphasize the form and sequences into which these effective elements were delivered: traditionally presented through an opening lecture, the advance organizers and initial knowledge were to be presented through videos that the students should consult independently, as pre-class activities. Instead, the in-class activities would encompass more practical exercises, peer and teachers' consultation, problem solving and students' presentations. Allegedly, Sams and Bergmann claimed that this approach should make a radical difference between the traditional methods: lectures and homework, in that sequence (Sams, & Bergmann, 2013). The FCM impacted immediately on an international professional' community in search of guidelines to intervene in increasingly complex educational environments. Professional networks of teachers were suddenly created to showcase experiences and exchange tips for practice, just to mention but a few: The Flipped Learning Network, http://flippedlearning.org/ in USA, the Flip Net, http://flippedlearning.org/ in Italy, the Flipped Classroom in Spain http://www.theflippedclassroom.es/ or in Austria, http://www.flipped-classroom-austria.at/. As it was highlighted by the report on a survey conducted in 2014 by Sophia and Flipped Learning Network on 2,358 teachers, in a matter of two years, the teachers that recognized the term increased from 73% to 96%; and from 48% of teachers that had experimented the model to 78% of them. Moreover, this enthusiastic report revealed that 93% of the teachers flipping their classrooms started as their own initiative, (indicating that this is) part of a grassroots movement from classroom teachers. Observing this popularity, we formulated the following research question: to which extent the increase of interest and the possibly connected educational practices, are in connection with the research evidence?

The question is clearly connected to an underlying educational problem, namely, the frequent adoption of educational approaches and models that are scarcely connected with sound evidence came out from quality research activities that lead to valid results (Calvani, 2012; Hattie, 2008; 2015b). This problem is not new in education and was posed early by the movement called Evidence Based Education. Taking into consideration the assumption that the clinical practice in medicine is only based on empirical research, some educational researchers claimed by the early Nineties this approach to be necessary in education. Particularly, it was pointed out that governmental programmes and investements in education should be driven by research evidence, that is, *evidence-based*[1].

In this paper, the aim of the research is to search the highest quality evidence and synthesize it in order to see to which extent the principles of the FCM are driven by rigorous data that confirm and expand the learning benefits frequently claimed by the FCM's advocates: students' engagement, satisfaction, self-regulation and learning (particularly higher level

[1] The methods adopted by the EBE movement have been explained in depth in the works of Hattie (2008) and all the research centers working through this method (i.e.: the Johns Hopkins University School of Education's Center for Research and Reform in Education (CRRE), or the Society for Evidence-based Learning and Instruction (SApIE) in Italy.

skills). By analyzing the research on the overall FCM method with the studied ES of the separated components of the FCM, the approach in this paper will attempt to draw some recommendations for evidence-based practice.

2. The Flipped Classroom Model: Characteristics and claimed impacts

While the authors that launched the model (Bergmann, & Sams, 2012) focused particularly the role of technologies in the pre-class activities (i.e., videos) and of active and peer-learning during in-class activities, they did not connect specifically the FCM design to specific theoretical frameworks. It appears that there was former debate on inverting the class activities, for example in Lage, Platt and Treglia (2000), and the attention was put specifically in promoting students' activity and improving the communication with teachers and instructors (Jensen, Kummer, & Godoy, 2015). To this regard, there were two separate debates that put the basis for the FCM. The first related clearly the educational technologies, like the initial systems of computer assisted learning with simple texts and assignments, as proposed by Lage et al. (2000), but they evolved in a particular interest on digital videos as available open resources or as teachers generated content (Chung, & Khe, 2017; Zuber, Hew, Lu, Wageman, & Burke, 2016). As for the second debate, the strategy called "peer-instruction" played an important role in shaping in-class activities. Developed by Harvard's professor Eric Mazur, the approach emphasized the strength of group and couple's activities to promote effective learning (Mazur, 1997), and was adopted since the first works on flipped learning (Bergmann, & Sams, 2012; O'Flaherty, & Phillips, 2015). However, different approaches and activities could be connected to what has been generally pointed out as "active learning" in FCM practices: problem solving and project-based learning (Dodds, 2015; Njie-Carr et al., 2017), modelling with students' response systems and individual exercises followed by group activity (Chung, & Khe, 2017).

Another important theory later connected by the same Bergmann and Sams (2012) to the FCM was Mastery Learning, particularly taking into account Bloom's taxonomy. One of the strongest assumptions of FCM was the promotion of high level skills and knowledge, overcoming the limitations of the lecture, which supposedly promoted only knowledge and information. In Bloom's terms, the traditional model the lecture (first sequence) could be associated with impacts at the level of understanding and remembering; and homework (second sequence) would promote higher level skills (applying, analyzing). Reverting this model, the FCM left the student alone in the first phase (understanding and remembering) and guides her in the second phase, namely, for applying, analyzing (Zainuddin, & Halili, 2016).

As we can observe, at a first sight the FCM appears to be a homogeneous approach, but an initial consideration of the literature let us see that there are several approaches and modes of implementing the flipping. This diversity is reinforced by the specificities of subject fields, being most of the research on FCM undertaken in Higher Education with less empirical research in F-12 (Chung, & Khe, 2017). And with a doubtless weight of STEM in the initial practices and connected research (Zainuddin, & Halili, 2016).

The several studies converge in the following impacts for the FCM (Abeysekera, & Dawson, 2015; DeRuisseau, 2016; Lai, & Hwang, 2016; O'Flaherty, & Phillips, 2015): augmented motivation, improved self-efficacy and self-regulation of learning, development of critical thinking, visibility of learning processes both for the teacher and the students, specific subject learning, among other effects. However, these results can be achieved only if considering some constrains that the same authors and other advocates of the FCM warn against (Plaisent, Dayagbil, Pogoy, & Bernard, 2016).

In fact, the original FCM introduced by Sams and Bergmann led to the Four Pillars for flipped learning, or the F-L-I-P principle, namely: (i) flexible environment; (ii) learning culture; (iii) intentional content; and (iv) professional educator. While these four pillars do not clearly introduce the kernel of the FCM (which is the students' independent activity focused on content acquisition on a topic taught and active learning guided by teachers in class instead of theoretical or modelled lectures) they spot the requirements for the FCM to work properly: technologies of good quality (flexible environment), quality and innovative pedagogical approaches and content and an educator that is skilled enough to conduct the complex integration of sequences.

In the literature, the criticalities of the FCM have been associated to the infrastructures, the quality of video resources, the digital skills of teachers and students, the culture of homework, the cognitive load that the model could introduce, and the overall learning culture including technology and pedagogical innovations' acceptance (Betihavas et al., 2016; Kaw, Besterfield-Sacre, & Clark, 2016; Logan, 2015).

3. Method

Taking into consideration the research problem and the initial exploration of the literature, our study will focus the instructional method of flipped classroom through a synthesis of evidence as defined within the EBE approach in its broader conception. Due to the low presence of empirical studies showing means' comparisons or directly ES (effect size) in the case of FCM as we will show further, performing a meta-analysis is not possible. In this case is offered hence a critical synthesis of a relevant number of reviews of the literature on FCM. The approach has been applied by the SApIE group for other instructional methods (Pellegrini, & Mensuali, 2015; Vivanet, 2015).

Hence, the basic unit of analysis composing the data for this research were research articles of a specific type, that is, reviews of the literature on FCM. Reviews of the literature can adopt several methods based on the conceptual and thematic analysis with critical focuses – Critical Reviews –, or more quantitative and extensive analysis using both coding and representations as text mining techniques – Systematic Reviews. Moreover, the reviews can synthesize empirical, quantitative research – Meta-analysis – or qualitative research – Meta-ethnographies – (Bonaiuti, Calvani, & Ranieri, 2016). Therefore, the process of data extraction was performed on scientific databases, as sources of aggregated scientific documents, and particularly literature reviews. Four frequently used repositories were explored: Google Scholar, Scopus, ERIC and WOS. Since every repository offers different features, the research strategy performed varied slightly. For WOS and Scopus the research criteria adopted were: TITLE-ABS-KEY: ((flip*) AND (classroom OR learning)); for WOS it was added the parameter AND (review OR survey), whereas for Scopus the criterion AND (LIMIT-TO (DOCTYPE, "re") was adopted through the advanced search interface. As for ERIC, the criteria were (review+AND+"flip*+classroomORlearning); whereas for Google Scholar the criteria were: allintitle: "Flip* classroom" OR "Flip* learning" AND review. The timeframe adopted to perform the above mentioned search was 2000-2017. Also Italian journals were searched, raising interesting results. One could appreciate Special Issue of Bricks Journal 2/2015, http://www.rivistabricks.it/2017/08/02/bricks-n-2-2015/ which devotes 12 articles to showcase several good practices; or the conceptual analysis made by Cecchinato e Papa (2016) and Giglio (2016). However, none of these approaches reported experimental research or systematic reviews of the research at Italian level. Overall the search strategy followed the Preferred-Reporting of Items for Systematic Reviews and Meta-Analyses (Prisma, Moher, Liberati, Tetzlaff, Altman, & Prisma Group, 2009), as it is represented in Figure 1; the results and flow of the research are there represented.

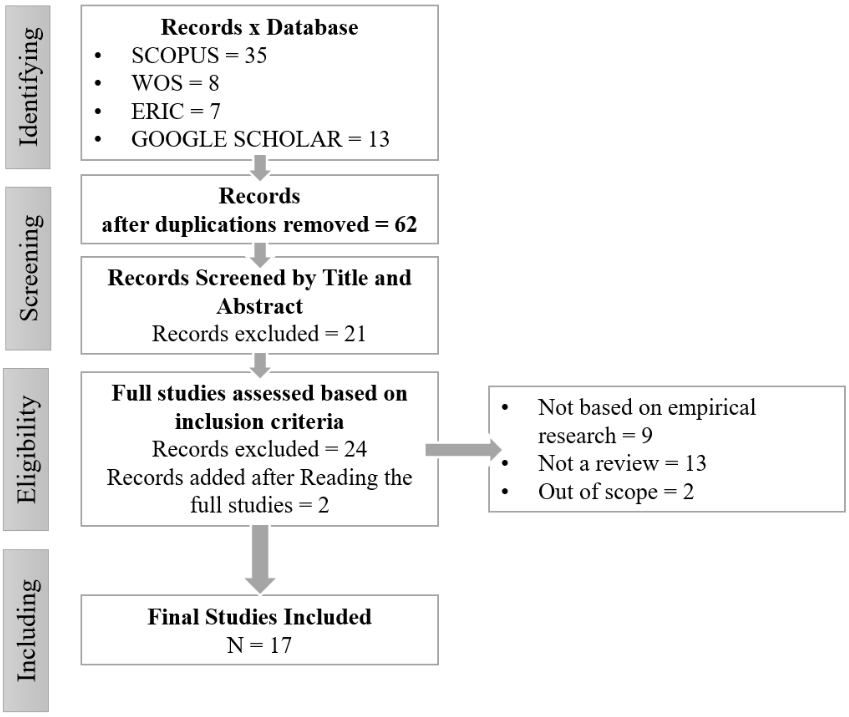

Figure 1. The Preferred-Reporting of Items for Systematic Reviews and Meta-Analyses (Prisma) flow-chart.

The 17 selected studies were coded by the author of the study adopting the dimensions of analysis illustrated in the Figure 2. The dimensions were sort of "queries" adopted to analyze and classify the studies, according to a conceptual scheme responding to the research question.

Data Field	Data Sub-field	Assigned values	Type of Data
1. Article Identification	Article Title	Title as published	Text/ Coded As found
	Type of article	Range of values: Journal article, conference proceedings, reports, magazine articles. The articles and conference proceedings should be peer-reviewed. All documents in English.	
	Source	Journal title	
	Author/s	Author(s) name and surname	
	Publication date	Year (2000-2017)	
	Key words	Author's keywords	
	Abstract	Full abstract as published	

2. Scientific Database	WOS SCOPUS ERIC Google Scholar	Presence [1] /Absence[2]	Value / Coded As found
3. Research Type	Research Type according to the EBE characterization	1- Meta-analysis 2- Systematic Reviews 3- Critical Reviews 4- Randomized Controlled Studies or Quasi-experiments.	Label/ Coded on the basis of methodological paragraphs within the selected articles.
4. Research Approach	Flipped-Classroom's Elements under analysis	Classification of the types of elements characterizing the flipped approach. Three main categories of analysis: 1. preparatory elements out-side class 2. in-class activities 3. assessment approach	Multiple Labels/ Coded on the basis of occurrence of elements within the selected articles.
	N of articles reviewed	N of articles finally selected for the review	Number/Coded on the basis of the nr provided in the review.
	Number of subjects	N of participants engaged through the several studies reviewed	Number/Coded on the basis of the nr provided in the review.
	Timeframe	Period covered by the publications reviewed	From-To/Coded on the basis of the reviewed articles' years of publication
	ES	Effect sizes eventually informed in the meta-analysis or RCT	Value of ES/ Coded on the basis of the value provided in meta-analysis or empirical studies.
	Level of Instruction	Level of instruction of the studies reviewed	Label/ Coded on the basis of the levels of instruction detected or mentioned.
	Subject Field	Disciplinary area of intervention: Biomedical sciences, STEM, Engineering, Social Sciences, General	Text/ Coded on the basis of the subject fields' detected or mentioned.
5. Conclusions		Conclusions of the study	Text/Coded on the basis of the conclusions as expressed in the study.
6. Study Shortcomings		Shortcomings of the study under analysis	Text/Coded on the basis of: • Methodological shortcomings (Type of primary studies, rigorous method of selection); • Other shortcomings identified by the authors of the study under analysis.
7. Further research		Instructions and implications for further research	Text/Coded on the basis authors instructions for further research.

Figure 2. Database structure showing the data type and values assigned to the articles processed.

3. Results

The 17 studies hereby considered spanned from 2013 to 2017, but most reviews (15/17) covered the timeframe 2015-17. Moreover, only 4 of 17 studies were Proceedings, with 1 Teaching Document that after being read was seemingly a review of the literature for a specific item (STEM – Science, Technology, Engineering and Mathematics areas – higher education). It is worth to notice that most proceedings can be placed amongst the first studies of the period covered (2013, 2015), while more structured research follows them (2015, 2016, 2017).

It could be inferred from this information that the FCM as instructional method is moving from an initial stage of primary studies (of several types) to a more advanced stage where some synthesis (reviews) can be found. However, one could ask about the quality of primary studies summarized, an information that actually emerged from the types of studies. In fact, the 'critical reviews' elaborated as narrative reports on the research analyzed were 8/17 (Chung, & Khe, 2017; Delgado, Wardlow, McKnight, & O'Malley, 2015; DeLozier, & Rhodes, 2017; Dodds, 2015; Logan, 2015; Rahman, Aris, Mohamed, & Zaid, 2015; Zainuddin, & Halili, 2016; Zuber et al., 2016). Six studies qualified as systematic reviews, adopting specific criteria to select and characterize the research (Betihavas et al., 2016; Bishop, & Verleger, 2013; Kerr, 2015; Njie-Carr et al., 2017; O'Flaherty & Phillips, 2015; Wen, Zaid, & Harun, 2015); and 1 study could be classified as meta-analysis (Chen, Lui, & Martinelli, 2017). In spite of the labels given by the authors (which could in fact differ from our classification) we classified papers over the basis of the method followed to review the literature. The remaining two articles were a quasi-experimental research included by the quality of the analysis that introduced comparisons between flipped, semi-flipped and blended courses and a clear conceptualization based on a sound review of the literature (Jensen et al., 2015; Kaw et al., 2016). Beyond the impossibility of performing a meta-analysis to summarize the best evidence or the 'what works and in under which conditions' effect', we envision that there the researchers and practitioners face actual difficulties to produce experimental conditions able of understanding the effects of the FCM on several dependent variables of interest (from study skills to higher level skills' learning). In addition, two articles did not expressed clearly the number of articles analyzed, and 9 out of 17 did not showed the number of subjects engaged in the primary studies (N) as it is expected and necessary to perform meta-analysis. From the studies providing this information, we learn that a total number of 232 articles were covered, and 9809 subjects participated to the several primary empirical researches. This is an irrelevant number if we just envisage the high numbers of teachers and students engaged in FCM networks and experiences.

Another important issue relates the levels of instruction and the subject field into which the several reviews could be classified. As it is showed in the Figure 3, most reviews have summarized empirical research on FCM conducted in Higher Education; notably, a synthesis of research seems to be necessary for the K12 level, with only 4/15 studies including some experiences in secondary education, and only one review entirely devoted to primary and secondary education. Moreover, if we consider the disciplines reviewed, the results is skewness over the broader are of STEM, with specific focus on Biomedical Sciences and Engineering. These results cannot be easily unfold, but they are actually communicating that rigorous research on FCM is less frequent or completely missed in certain disciplinary fields, like Social Sciences and the Humanities. We cannot tell whether this is the outcome of the lack of interest on this instructional method in these areas, it is a problem of feasibility of application, or the fact that the forms of learning in these areas are refractory to this method.

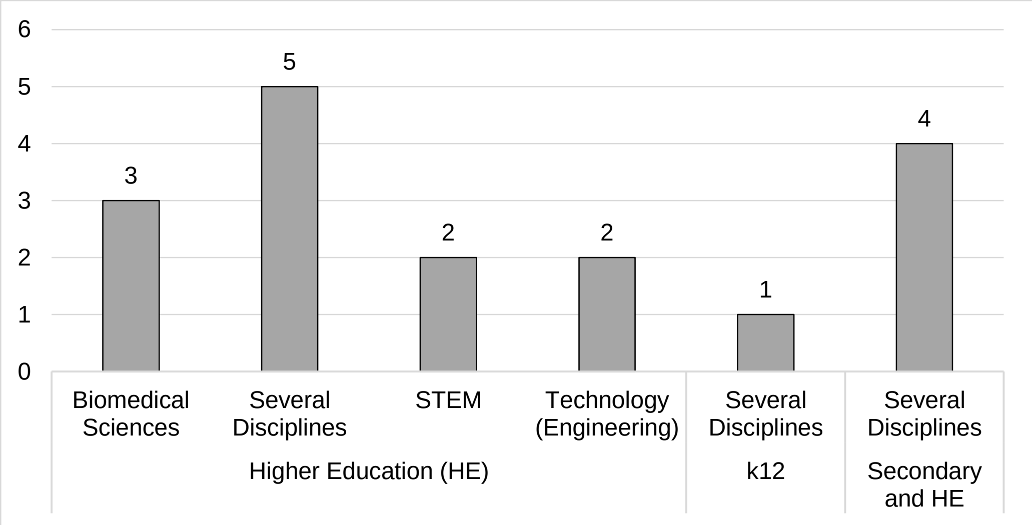

Figure 3. Classification of articles by discipline and instruction level.

The rest of the information relating the 17 articles processed were analyzed in terms of the pedagogical sequences and the impact of the FCM, introduced in the following paragraphs.

As for the pedagogical sequences, there is general consensus on the importance of videos, while less studies have focused to which extent the videos have been elaborated by the teachers or are external, mostly Open Educational Resources (OER). It is also interesting to see that the key term OER does not appear, in spite of most videos (particularly the frequently adopted Khan Academy videos) fall within this category. Moreover, in only few cases the reviews have considered the videos to be only or the main source of the pre-class activities.

The second sequence "in-class activities" is even more heterogeneous. Overall, there is agreement on the "active learning" component, but the several forms of characterizing and hence synthesizing the studies show different levels of granularity to explain the pedagogies adopted in in-class activities. In a good number of studies (9/17) the "active learning" key word encompass higher granularity strategies like problem-based, peer-learning, cooperative learning and so on. In the rest of the studies, activities in class are described at lower levels of granularity (quizzes, micro-lectures, short exercises).

Only two of the studies adopt a theoretical framework to characterize this sequence (Bloom's taxonomy and Knowledge construction theory). However, the sequence that shows the biggest gaps is the closing one, devoted to assessment. Several reviews (10/17) do not report any focus on assessment processes as part of the FCM, assuming the final exam traditionally adopted, as well as students' surveys on their opinions on the FCM may account for the its effects.

The analysis of the studies in which assessment is a matter of analysis highlight the importance of focus according to the type of learning processes under analysis, particularly regarding the higher level skills associated to active learning approaches. This is the case for 5 out of the 17 studies: the Bloom's taxonomy connected to specific assessment activities where these type of skills are to be put into action; as well as other frameworks to analyze specific effects of the FCM are considered (i.e. ICT literacy skills). In two cases, the third sequence is considered but only traditional forms of assessment (mostly connected with low level skills) are spotted.

The results in Figure 4 and 5 show a critical picture. An oversight on the column “Conclusions” let us grasp that the FCM is highly dependent on a number of conditions: teachers’ training (studies 15, 7), the technological infrastructures (3, 5, 10, 15), students’ support during independent activity or homework (3, 10, 11). There are also a number of FCM results that the reviews recommend to consider alongside the positive effects on students’ learning and engagement: reported low students’ satisfaction with the innovations proposed by FCM (3, 5, 10) or by the entangled students’ overload (3, 5, 10, 11, 15, 17). Finally, a relevant number of reviews showed concern on methodological issues such the comparability of FCM research studies as well as the lack of accurate quasi-experimental (pre-post test) and experimental studies (1, 14, 13, 17).

Moreover, even if only 3 of 17 studies that were labeled as “Critical” showing more concerns on the method that being enthusiastic about it, of the 14 remaining studies only 5 are completely “Positive”, 6 are “Positive but cautious” and 3 are “Positive but highly cautious”. Putting these results against the shortcomings observed in the review/research approach (“Shortcomings” column), we notice that most studies labeled as Positive are based on primary studies that are methodologically diversified (including action research, experiences and conceptual studies) or explain poorly the review method. Instead, the opposite is true for the studies labeled as “Critical”, “Highly cautious” and “Cautious”. With the exception of one “Cautious” study which review method is not clearly identifiable, the remaining 13 studies have adopted rigorous review methods (basing on meta-analysis, systematic reviews or critical review methods on empirical, peer-reviewed research). In order to show clearly this relationship, the Figure 4 brings a mosaic plot where the text-labels have been converted to a ordinal scale, the frequencies count and compared.

As for the column “Further Research”, as one could expect the claims go in the direction of clearer descriptions of FCM encompassing better empirical research where the sequences can be compared with a non-flipped instructional method, and the effects of every sequence isolated in their contribution to the overall method. In fact, one issue is to understand whether the effect of the peer-learning and the higher interaction with the teacher, or the quality of video resources accounts for the FCM’s impact on learning effectiveness more than the actual sequence in which the resources are presented (1, 10, 12, 13, 15, 17).

A number of studies focus more specific topics, going beyond the enthusiastic embracement of the whole method, in order to understand the specific contribution to the FCM to learning and the whole class environment’ enhancement. For example, the effects of FCM on drop-outs (3), high order knowledge and skills like self-regulation(5), metacognition (7, 11), critical thinking (9), independent study (14) and long term benefits and transfer to other learning activities in a lifelong-learning perspective (11, 14). Furthermore, the studies analyzed suggest further control of variables that could encompass extraneous effects all else held constant. This is the case of the socio-economic status of students or their difficulties in handling technologies (4), the novelty effect of the FCM technologies and pedagogical approach (4), the teachers and students workload (4, 15), the teachers’ training (4, 6, 7, 12) the quality of the videos (5,12), the class size and the effective possibilities of scalability (5).

Author	Conclusions	Shortcomings	Further research
[1] (Bishop & Verleger, 2013)	Positive but highly cautious.	• Rigorous approach; • Narrow selection of articles.	• Clearer descriptions of FCM; • More empirical research.
[2] (Delgado et al., 2015)	Positive.	• Partial focus on FCM; • Primary studies are methodologically diversified.	• Not specified.
[3] (Dodds, 2015)	Positive but cautious.	• Rigorous approach; • Reduced number and sectorial primary studies.	• Impact of FCM on drop-outs. • The impact of the “novelty effect”; • Effects of teachers and students’ workload.
[4] (Jensen et al., 2015)	Critical.	• Rigorous quasi-experimental approach.	• Effects of socio-economic status and the digital divide to adopt technologies for out-of-class activities; • Students’ workload; • Instructors’ ability to conduct active learning.
[5] (Kerr, 2015)	Positive but cautious.	• Rigorous approach.	• Design and scaffolding for self-regulation in FCM; • Study the length and format of videos for the first sequence. • Analysis of the class size to understand the scalability of FCM. • Study the differences across subject fields, particularly those encompassing more skills development.
[6] (Logan, 2015)	Positive.	• The review method is not specified.	• Lack of studies on student perceptions of the FCM; • Need for a guide teachers to best practices; • Technologies for FCM.
[7] (O’Flaherty & Phillips, 2015)	Positive but highly cautious.	• Rigorous approach; • Reduced number and sectorial primary studies.	• Faculty development to implement FCM.
[8] (Rahman et al., 2015)	Positive.	• The review method is poorly specified.	• Explore “students’ learning styles”; • Focus on in-class activities beyond the whole FCM effectiveness.
[9] (Wen et al., 2015)	Positive.	• The review method is too specific.	• Tools to measure the ICT skills and the trend of learning strategies used; • Measures of either critical thinking skill or problem solving skill; • Ways to modify the learning strategies to enhance students’ ICT skills through FCM

[10] (Betihavas et al., 2016)	Positive but highly cautious.	• Rigorous approach; • Reduced number and sectorial primary studies.	• Relationship between student satisfaction and academic performance; • FCM implementation process; • Value of pre- and within-class active learning strategies and the outcomes of the flipped classroom using alternative measures.
[11] (Kaw et al., 2016)	Positive but cautious.	• A rigorous quasi-experimental design.	• Adopt research designs that better focus higher order knowledge and skills development and metacognition. Longer term benefits of FC.
[12] (Zainuddin & Halili, 2016)	Positive.	• Primary studies are methodologically diversified.	• Suitability of FCM with poor video quality or low-trained instructors; • Focus on quality in-class activities; • More studies in Humanities and Social fields; • More empirical studies including DBR (Design-based Research).
[13] (Zuber et al., 2016)	Critical.	• Rigorous approach; • Reduced number and sectorial primary studies could be included.	• Further research into the FCM is required to determine consistent theoretical frameworks and methods.
[14] (Chen et al., 2017)	Positive but highly cautious.	• Rigorous approach • Reduced number and sectorial primary studies could be included.	• To evaluate the higher levels of Kirkpatrick's framework (three of classification measures: perception, attitude and change in knowledge and skill sets); • To focus on change in knowledge integrated with differential effects on levels of cognitive processes.
[15] (Chung & Khe, 2017)	Positive but cautious.	• Rigorous approach based diversified research; • quality research scale.	• Importance of including pre post-test comparisons; • Broader timeframe of experimental activities; • Teacher and student's workload.
[16] (DeLozier & Rhodes, 2017)	Positive but cautious:	• The review method is not specified.	• Empirical research with focus on cognitive processes (students cognitive load, individual processes); • Students' approach to learning: detected changes; • Students' engagement with independent activities/study.
[17] (Njie-Carr et al., 2017)	Critical.	• Rigorous approach; • Reduced number and sectorial primary studies.	• Better control of pre- and in-class activities, type of content and level of students; • Comparative studies across other biomedical areas (pharmacy, medical education, etc.).

Figure 4. Research outcomes, with focus on FCM's impact as instructional method.

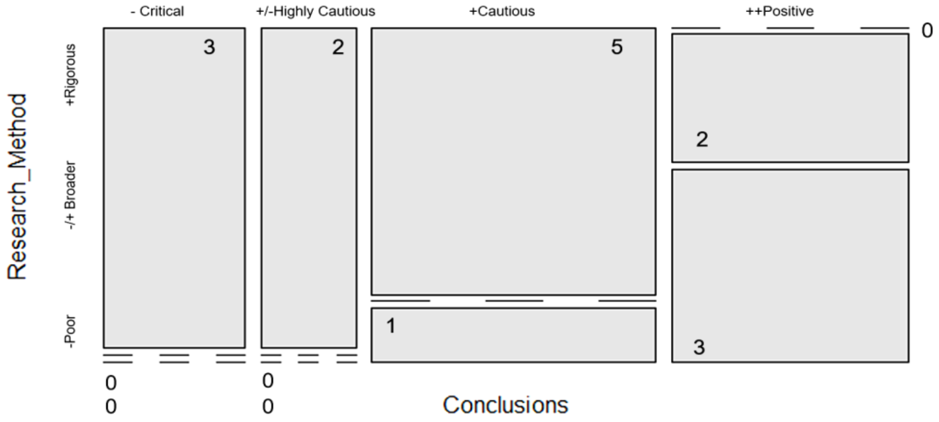

Figure 5. Mosaicplot showing the relationship between Research_Method and Conclusions in the 17 studies analyzed. The textual values have been converted to a scale and frequencies calculated.

4. Discussion

Along the analysis of the studies hereby considered, three clusters of critical issues emerged: methodological, pedagogical and organizational.

As for the methodological cluster, most rigorous studies claimed the need of more accurate empirical research studying the overall methods of FCM with traditional or not-flipped but still active face-to-face and blended learning methods. The design of FCM is still a matter of debate and due to the lack of solid constructs (higher level of granularity as pedagogical strategies instead of the type of technologies adopts – i.e., clickers) the embedded instructional methods in FCM are still obscure. As methodological consequence for research it is not easy to compare studies to make quality synthesis of research. Supporting this problem, the prevalence of studies in the STEM disciplinary field as well as at the level of higher education do not inform properly K12 practices.

With regard to the pedagogical cluster, it highlighted the need to know how the several instructional methods contribute to the learning effectiveness. A relevant number of reviewers directly found or indirectly adverted that the overall FCM could not be the discriminant factor for effective learning, but the good adoption of other proven instructional methods (interactive videos, advanced organizers, peer-instruction, mastery learning, teacher and peers feed-back). These methods are connected to well-known research traditions and pedagogical theories that should be clear to the FCM's implementers. This problem could be at least conceptually connected to the lack of attention to the appropriate design of assessment. Indeed, the formative and summative assessment in FCM still receives little attention as we could observe in the results section. If the FCM is supposedly connected with impacts on higher level skills and knowledge, the final assessment should not be the same applied for traditional lectures. Moreover, it should identify the progress in those type of skills, not only memorizing and understanding, along relevant cycles of instruction including the transfer of skills (both for higher education and K12). Finally, within this cluster the teachers should carefully consider at the time of designing for learning two sources of interference: the cognitive load and the novelty effect. The first type of interference could explain the often referred students' negative perceptions on FCM interventions. The technologies needed to implement the FCM, as well as a different pedagogy to which the students have not been socialized in their prior experience could encompass extraneous cognitive load blocking students' engagement or producing high workload. The second type of interference induces positive effects that are not associated to the instructional method but to the enthusiasm generated by the technological novelty. Accurate learning design at higher granularity levels (pedagogical theories and instructional methods) at a low granularity level (media design for videos, instructional design for other content as well as distance and in class activities) remains a crucial issue for effective FCM.

Finally, for the organizational cluster it should be considered that the complexity of the method requires acknowledgement at institutional level to obtain appropriate support with regard to the technological infrastructures and the educational technologies to be used. Moreover, the pedagogical complexity intertwined with the technological ones require particular attention to teachers' training, as well as students support to 'get access' to a learning culture of innovation. Issues like lack of acceptance of a Bring Your One Device (BYOD) approach; or the lack of access to the internet at home, or parental permission in the case of K12, could easily undermine the preparatory sequence and the independent students' practice. But also poor self-regulation; or low acceptance of homework by the students (particularly K12) could intervene in the effectiveness of the pre-class activities that are deemed crucial in the model.

4.1. Discussing the FCM through the lens of "Visible Learning": recommendations for learning design and practice.

The above critical issues lead us to search for advice in formulating an effective FCM. Building on the affirmation hereby explored that the FCM is a method composed by other effective instructional methods, we will examine the contributions of the 'Visible Learning' approach for K12 (Hattie, 2008) and higher education (Hattie, 2015a) relating these specific methods to the FCM sequences. We will embrace Hattie's recommendation of 'intentionally work to make visible the teachers' method and the connected students' results'. At each component of the sequence, we will display the ES (effect size) associated as quality meta-analytic evidence on a specific method/activity[2].

Pre-class activities. The first crucial phase of FCM consists on the independent activities aimed at delivering the content as it could be done in a traditional lecture (Understanding and Remembering, in Bloom's taxonomy), as follows:

1. Introduce clearly the learning goals and the related methods, exploring together with the students the previous knowledge. This important step prepares the student to self-regulate learning along the innovations introduced by the FCM. The advanced organizers have shown an ES of d = 0.41, and models based on clear goals to be reached an ES of d = 0.60. Moreover, Metacognitive strategies to self-regulate personal study have also shown a medium ES of d = 0.60;
2. Adopt appropriate digital environments, educational technologies and particularly videos to deliver the content for independent learning. These tools and resources should be carefully designed in advance in accordance to the principles of media

[2] For this elaboration, the website http://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement/ has also been consulted. Oversimplifying, we remind the reader that the ES is a statistic measure adopted to understand the "strength of the experimental effect". In Hattie's work, the "hingpoint" for effective instructional methods is d = 0.40.

design to control the cognitive load level (see Bonaiuti, Calvani, Fini, & Landriscina, 2011 for Italian readers; Clark, & Mayer, 2011). Amongst the educational technologies, quality interactive videos have shown a medium ES of d = 0.52, while distance education has been connected to a very low ES of d=.09. Therefore, the teachers have to bear in mind that technologies don't do the work alone. Moreover, the teachers should consider to adopt short videos reproducing sequences of direct instruction, since this last can still be considered very effective (d = 0.60);

3. Provide sources of feed-back on the independent pre-class activities. The feed-back has shown a medium-high ES (d = 0.73). To this regard, in this phase the teacher should adopt the available technologies for immediate feed-back, which should be divided into two type: formative quizzes and simple learning analytics to self-check progress in independent activities; with explanations associated to 'what happened' in the online activities (particularly taking into consideration a very simple synthesis of learning analytics). This would combine the effects of feed-back with metacognitive strategies. Moreover, at organizational level, in cultures where homework could fail as strategy, integrating the scores obtained during independent, technology-enhanced activities, to final scores could ensure that the students do the programmed work;
4. In the case of K12, provide information to families, if FCM will be adopted by the first time. Parental involvement has shown an ES of d= .49, which is a medium value and could reinforce the engagement of parents in providing operational support and permission to adopt technological devices and internet at home.

In-class activities. The second phase of FCM is based on the guided activities in class aimed supporting the achievement of higher level skills and knowledge (analyzing, evaluating and creating learning goals in Bloom's terms). To accomplish this task, the elements are:

1. Take into careful consideration the results of the independent phase and provide feed-back prior to pass to the in-class sequence;
2. Select the 'learning architecture' (Clark, 2000) that is most appropriated for the own discipline as well as for the target group. Less mature students require more guidance and could benefit from classroom discussions on the content delivered (d = 0.82), using concept maps (d = 0.64); whereas more mature or academically skilled students can smoothly go into cooperative learning (d = 0.40). For all groups problem solving teaching through progressive demonstrations has proven to be effective (d = 0.60);
3. The strategies of peer-tutoring to solve exercises: in spite of the demonstration and training required for a correct implementation, they can be considered effective (d = 0.55);
4. Collaborative learning (d = 0.29) inquiry-based activities (d = 0.35) and worked examples for individual practice (d = 0.37), and problem-based learning (d = 0.15) should be carefully conducted by the teacher, that has to balance all the resources and sequences of an accurate learning design, and consider her own efficacy in orchestrating learning as well as controlling the climate class;
5. Technological mediators of learning activities should be also carefully considered. The ES studies have shown medium-low ES for activities like Intelligent Tutoring systems (d = 0.43), Gaming/Simulations (d = 0.37), Computer Assisted Instruction in reading (d = 0.26), math (d = 0.30) and science (d = 0.23);
6. All over the design of the in-class sequence, the teacher should carefully consider the complexity of the whole. To more complex independent, digital activities (pre-

class) it could follow a simpler sequence of in-class that maintains the focus on the intrinsic cognitive load and controls extraneous cognitive load. Or vice versa, complex activities in class should be accompanied by simple digital environments and resources.

Assessment activities. The third phase of FCM accompanies the development of the first and second phase and ends up with a specific moment in the designed learning strategy. Its elements are:

1. Every assessment strategy will have to base its design in tight connection with the learning goals set and properly illustrated to the students;
2. All along the first and second phase the teacher will have collected scores and students' outputs that will be accommodated in a final scheme of integrated assessment. We already considered feed-back informing the results of every step highly effective (d = 0.73), but an overall formative evaluation has also good effects on students' reflection and learning (d = 0.68);
3. Collect students' feed-back over their own reflection processes along the activities and discuss on them by the end of the FCM process in order to inform the final formative assessment;
4. Discuss with the students the overall class performance in terms of formative and summative assessments. The students' evaluations on the impact of teaching have are effective not only for the teacher (making teaching *visible*, d = 0.47);
5. Develop instruments that better inform the final score. A rubric with a breakdown of goals and activities as well as a clear scale to score students' performance could be the case, when the class-size is large and the teacher workload for formative and summative feed-back is considerable;
6. Within the rubric, split low and high level skills' assessment, in connection to the goals and activities undertaken along the FCM.

5. Conclusions

This article has brought some evidence on the fact that the FCM, in spite of its great popularity, has not definitely proven its efficacy as instructional method. After producing a synthesis of 17 reviews of the literature on the issue, it has been found that most rigorous approaches in research are highly cautious about the impact of the model, while the broader review approaches embrace more enthusiastic positions. Having considered these critical issues, we attempted to enrich the picture through an analytical exercise: since the FCM is a sort of complex hybrid, integrated by embedded instructional methods and theories, we split the three main sequences and analyzed every one of them as separate instructional method/strategy on the light of the Visible Learning approach, as part of the EBE approach. This study presents a number of limitations that should be carefully considered at the time of adopting some of the conclusions as principles here embraced. Firstly, the synthesis was produced by a single researcher, with no availability of control for the articles' coding process. The strategy to face this problem was twofold: from one hand the articles were coded two times in a spaced term of time, from the other hand the codes were built on the basis of objective information gather from the article leaving little space for interpretation. However, the final phases of interpretation could be fact of discussion and disagreement by other researchers. The extensive tables were built in an attempt to make –as far as possible- all information to be available in order to allow other researchers to revise the primary sources of information (the 17 articles consulted). Secondly, the quality of the primary research (on which the reviews analyzed worked), was classified on the basis of the information provided by the reviewers; the primary studies were not analyzed directly. Thirdly and finally, our effort to conceptualize the embedded instructional methods of FCM, in connection with the Visible Learning approach has only hypothetical value and its real combined effectiveness should be studied through further empirical research that in time can be summarized. To this regard, there are ongoing projects like the (Rudd, 2013) that will bring light in the near future.

Doubtless, FCM is an interesting pedagogical innovation, that builds on the basis of excellent practices and professional knowledge, as well as on well disseminated theories of instruction that are already an intellectual heritage of most teachers. However, we advocate here for more cautious approaches and more dialogue between educational research and educational practices to justify the investments and professional efforts for teaching and learning: paraphrasing Hattie (2015b) moving beyond the politics of distraction.

References

Abeysekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: definition, rationale and a call for research. *Higher Education Research & Development*, *34*(1), 1–14. http://doi.org/10.1080/07294360.2014.934336 (ver. 15.12.2017).

Bishop, J., & Verleger, M. A. (2013). The Flipped Classroom: A Survey of the Research. In *2013 ASEE Annual Conference* (p. 6219). Atlanta: American Society for Engineering Education. http://doi.org/10.1109/FIE.2013.6684807 (ver. 15.12.2017).

Bergmann, J., & Sams, A. (2012). *Flip your classroom : reach every student in every class every day*. Arlington: International Society for Technology in Education.

Bonaiuti, G., Calvani, A., Fini, A., & Landriscina, F. (2011). *Principi di comunicazione visiva e multimediale: fare didattica con le immagini*. Roma: Carocci.

Bonaiuti, G., Calvani, A., & Ranieri, M. (2016). *Fondamenti di didattica : teoria e prassi dei dispositivi formativi*. Roma: Carocci.

Bricks Journal. Special Issue, 2, 2015. http://www.rivistabricks.it/2017/08/02/bricks-n-2-2015/ (ver. 15.12.2017).

Calvani, A. (2012). *Per un'istruzione evidence based. Analisi teorico-metodologica internazionale sulle didattiche efficaci e inclusive*. Trento: Erickson.

Cecchinato, G., & Papa, R. (2016). *Flipped classroom: un nuovo modo di insegnare e apprendere*. Torino: UTET.

Chen, F., Lui, A. M., & Martinelli, S. M. (2017). A systematic review of the effectiveness of flipped classrooms in medical education. *Medical Education*, *51*(6), 585–597. http://doi.org/10.1111/medu.13272 (ver. 15.12.2017).

Chung, K.L., & Khe, F.H. (2017). A critical review of flipped classroom challenges in K-12 education: possible solutions and recommendations for future research. *Research and Practice in Technology Enhanced Learning*, *12*(1), 4. http://doi.org/10.1186/s41039-016-0044-2 (ver. 15.12.2017).

Clark, R.C. (2000). of Instruction. *Performance Improvement*, *39*(10), 31–38.

http://doi.org/10.1002/pfi.4140391011 (ver. 15.12.2017).

Clark, R.C., & Mayer, R.E. (2011). *E-learning and the science of instruction : proven guidelines for consumers and designers of multimedia learning*. San Francisco, CA: Pfeiffer.

Delgado, A.J., Wardlow, L., McKnight, K., & O'Malley, K. (2015). Educational technology: a review of the integration, resources, and effectiveness of technology in K-12 Classrooms. *Journal of Information Technology Education*, *14*(14), 397–416.

DeLozier, S.J., & Rhodes, M.G. (2017). Flipped classrooms: a review of key ideas and recommendations for practice. *Educational Psychology Review*, *29*(1), 141–151. http://doi.org/10.1007/s10648-015-9356-9 (ver. 15.12.2017).

DeRuisseau, L.R. (2016). The flipped classroom allows for more class time devoted to critical thinking. *Advances in Physiology Education*, *40*(4). http://advan.physiology.org/content/40/4/522 (ver. 15.12.2017).

Dodds, M. (2015). *Evidence for the Flipped Classroom in STEM*. http://www-users.cs.york.ac.uk/~miked/publications/flipped_classroom.dodds.pdf (ver. 15.12.2017).

Flip Net. http://flippedlearning.org/ (ver. 15.12.2017).

Flipped Classroom. http://www.theflippedclassroom.es/ (ver. 15.12.2017).

Flipped Classroom Austria. http://www.flipped-classroom-austria.at/ (ver. 15.12.2017).

Giglio, A. (2016). The flipped classroom. *Italian Journal of Educational Technology*, *24*(1), 65–66. http://doi.org/10.17471/2499-4324/880 (ver. 15.12.2017).

Hattie, J. (2008). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hattie, J. (2015a). The applicability of Visible Learning to Higher Education. *Scholarship of Teaching and Learning in Psychology*, *1*(1), 79–91. http://doi.org/10.1037/stl0000021 (ver. 15.12.2017).

Hattie, J. (2015b). *What doesn't work in education: the politics of distraction*. Pearson Open. https://www.pearson.com/content/dam/corporate/global/pearson-dot-com/files/hattie/150602_DistractionWEB_V2.pdf (ver. 15.12.2017)..

Jensen, J.L., Kummer, T.A., & Godoy, P.D. (2015). Improvements from a flipped classroom may simply be the fruits of active learning. *CBE Life Sciences Education*, *14*(1), ar5. http://doi.org/10.1187/cbe.14-08-0129 (ver. 15.12.2017).

Kaw, A., Besterfield-Sacre, M., & Clark, R.M. (2016). Comparing the effectiveness of blended, semi-flipped, and flipped formats in an engineering numerical methods course. *Advances in Engineering Education*, *5*(3). http://advances.asee.org/publication/comparing-the-effectiveness-of-blended-semi-flipped-and-flipped-formats-in-an-engineering-numerical-methods-course/ (ver. 15.12.2017).

Kerr, B. (2015). The flipped classroom in engineering education: a survey of the research. In *2015 International Conference on Interactive Collaborative Learning (ICL)* (pp. 815–818). IEEE. http://doi.org/10.1109/ICL.2015.7318133 (ver. 15.12.2017).

Lage, M.J., Platt, G.J., & Treglia, M. (2000). Inverting the Classroom: A Gateway to

Creating an Inclusive Learning Environment. *The Journal of Economic Education*, *31*(1), 30–43. http://doi.org/10.1080/00220480009596759 (ver. 15.12.2017).

Lai, C.L., & Hwang, G.J. (2016). A self-regulated flipped classroom approach to improving students' learning performance in a mathematics course. *Computers & Education*, *100*, 126–140. http://doi.org/10.1016/j.compedu.2016.05.006 (ver. 15.12.2017).

Logan, B. (2015). Deep Exploration of the flipped classroom before implementing. *Journal of Instructional Pedagogies*, *16*(July). Retrieved from https://files.eric.ed.gov/fulltext/EJ1106741.pdf (ver. 15.12.2017).

Mazur, E. (1997). *Peer instruction: a user's manual*. New York, NY: Prentice Hall.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., & Prisma Group. (2009). Preferred eporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, *6*(7), e1000097. http://doi.org/10.1371/journal.pmed.1000097 (ver. 15.12.2017).

Njie-Carr, V.P.S., Ludeman, E., Lee, M.C., Dordunoo, D., Trocky, N.M., & Jenkins, L. S. (2017). An integrative review of flipped classroom teaching models in nursing education. *Journal of Professional Nursing*, *33*(2), 133–144. http://doi.org/10.1016/j.profnurs.2016.07.001 (ver. 15.12.2017).

O'Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: a scoping review. *The Internet and Higher Education*, *25*, 85–95. http://doi.org/10.1016/j.iheduc.2015.02.002 (ver. 15.12.2017).

Pellegrini, M., & Mensuali, A. (2015). L'efficacia delle mappe concettuali per l'apprendimento: analisi critica di evidenze empiriche. *Form@re - Open Journal per La Formazione in Rete*, *15*(3), 129–141. http://doi.org/10.13128/formare-17154 (ver. 15.12.2017).

Plaisent, M., Dayagbil, F., Pogoy, A. M., & Bernard, P. (2016). Is flipped classroom a tendency or a fad? In *Blended Learning: Concepts, Methodologies, Tools, and Applications* (pp. 2150–2168). New York: IGI Global. http://doi.org/10.4018/978-1-5225-0783-3.ch104 (ver. 15.12.2017).

Rahman, A.A., Aris, B., Mohamed, H., & Zaid, N.M. (2015). The influences of flipped classroom: a meta analysis. In *ICEED 2014 - 2014 IEEE 6th Conference on Engineering Education* (pp. 24–28). http://doi.org/10.1109/ICEED.2014.7194682 (ver. 15.12.2017).

Rudd, P. (2013). Flipped learning evaluation., 1–6. https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Flipped_Learning_Evaluation.pdf (ver. 15.12.2017).

Sams, A., & Bergmann, J. (2013). Flip your students' learning. *Educational Leadership*, *70*(6), 16–20. http://www.ascd.org/publications/educational-leadership/mar13/vol70/num06/Flip-Your-Students'-Learning.aspx (ver. 15.12.2017).

Sophia, & Flipped Learning Network. (2014). *Flipped Learning Survey*. Retrieved from https://www.sophia.org/flipped-classroom-survey (ver. 15.12.2017).

The Flipped Learning Network. http://flippedlearning.org/ (ver. 15.12.2017).

Visible Learning. Hattie Ranking: 195 Influences And Effect Sizes Related To Student Achievement. http://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement/ (ver. 15.12.2017).

Vivanet, G. (2015). Che cosa sappiamo sull'efficacia delle tecnologie didattiche con soggetti con disturbo dello spettro autistico? *Form@re - Open Journal per La Formazione in Rete*, *14*(4), 77–92. http://doi.org/10.13128/formare-15798 (ver. 15.12.2017).

Wen, A. S., Zaid, N. M., & Harun, J. (2015). A meta-analysis on students' social collaborative knowledge construction using flipped classroom model. In *2015 IEEE Conference on e-Learning, e-Management and e-Services (IC3e)* (pp. 58–63). IEEE. http://doi.org/10.1109/IC3e.2015.7403487 (ver. 15.12.2017).

Zainuddin, Z., & Halili, S. H. (2016). Flipped classroom research and trends from different fields of study. *The International Review of Research in Open and Distributed Learning*, *17*(3). http://doi.org/10.19173/irrodl.v17i3.2274 (ver. 15.12.2017).

Zuber, W.J., Hew, K.F., Lu, Y., Wageman, J., & Burke, P. (2016). The flipped classroom, a review of the literature. *Industrial and Commercial Training*, *48*(2), 97–103. http://doi.org/10.1108/ICT-05-2015-0039 (ver. 15.12.2017).