Educational Research Review 22 (2017) 74–98

Contents lists available at ScienceDirect

Educational Research Review

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

Review

Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses

Ernesto Panadero [a, *], Anders Jonsson [b], Juan Botella [c]

[a] Departamento de Psicología Evolutiva y de la Educación, Facultad de Psicología, Universidad Autónoma de Madrid, Spain
[b] Department of Science, Kristianstad University, Sweden
[c] Departamento de Psicología Social y Metodología, Facultad de Psicología, Universidad Autónoma de Madrid, Spain

ARTICLE INFO

Article history:
Received 12 November 2016
Received in revised form 8 August 2017
Accepted 10 August 2017
Available online 24 August 2017

Keywords:
Self-assessment
Self-regulated learning
Self-efficacy
Self-evaluation
Self-regulation
Learning strategies
Motivation
Gender educational differences
Self-regulated learning measurement
Emotional regulation

ABSTRACT

This meta-analytic review explores the effects of self-assessment on students' self-regulated learning (SRL) and self-efficacy. A total of 19 studies were included in the four different meta-analyses conducted with a total sample of 2305 students. The effects sizes from the three meta-analyses addressing effects on different measures of SRL were 0.23, 0.65, and 0.43. The effect size from the meta-analysis on self-efficacy was 0.73. In addition, it was found that gender (with girls benefiting more) and certain self-assessment components (such as self-monitoring) were significant moderators of the effects on self-efficacy. These results point to the importance of self-assessment interventions to promote students' use of learning strategies and its effects on motivational variables such as self-efficacy.

Contents

* Corresponding author. Despacho 109, Departamento de Psicología Evolutiva y de la Educación, Facultad de Psicología, Universidad Autónoma de Madrid, 28049, Cantoblanco, Spain.
E-mail address: ernesto.panadero@uam.es (E. Panadero).

http://dx.doi.org/10.1016/j.edurev.2017.08.004
1747-938X/

1. Introduction

Student self-assessment has been one of the main areas of research in contemporary education and educational psychology research since the seminal reviewing work by Nancy Falchikov and David Boud in the late 80's (Boud & Falchikov, 1989; Falchikov & Boud, 1989). Research on self-assessment has also transcended these fields of research, extending into fields such as social psychology (e.g. Dunning, Heath, & Suls, 2004; Kruger & Dunning, 1999) and medical education (Eva & Regehr, 2005; 2008). The educational research on self-assessment is currently going through an important phase, as shown by recent publications reviewing the accumulated empirical evidence and proposing a new agenda for self-assessment (e.g. Brown & Harris, 2013; Panadero, Brown, & Strijbos, 2016).

As will be shown in this meta-analysis, the relationship between the constructs self-assessment, self-regulated learning (SRL), and self-efficacy has been the object of empirical research for at least twenty years. This relationship is both reciprocal and intricate: self-assessment is conceptualized as a learning regulatory strategy (Nicol & McFarlane-Dick, 2006; Panadero & Alonso-Tapia, 2013; Paris & Paris, 2001); SRL is dependent on self-assessment – via self-monitoring and self-evaluation – to support student learning (Butler & Winne, 1995; Zimmerman & Moylan, 2009); and self-efficacy is thought to enhance students' activation and use of regulatory strategies, such as monitoring and evaluation (Pajares, 1996, 2008; Schunk & Ertmer, 1999). Furthermore, self-assessment might increase the perceived capability among students, which could affect students' self-efficacy (Andrade, Wang, Du, & Akawi, 2009). In the coming sections of the theoretical framework we will present self-assessment, SRL, and self-efficacy along with the moderating variables included in this meta-analytic review.

1.1. Self-assessment

In a recent state-of-the-art review, self-assessment was defined as a "… wide variety of mechanisms and techniques through which students describe (i.e., assess) and possibly assign merit or worth to (i.e., evaluate) the qualities of their own learning processes and products" (Panadero, Brown et al., 2016, Panadero, Jonsson & Strijbos, 2016, p. 2). According to this definition, self-assessment is about students assessing their own work; not about signaling their perceived understanding to the teacher through "traffic lights" or evaluating their satisfaction with the instruction. Also important to notice is the emphasis on a "wide variety of mechanisms", which acknowledges that there are different ways to implement self-

assessment in the classroom. As a matter of fact, Panadero, Brown et al. (2016) found 20 different categories of self-assessment implementations in their review of different self-assessment typologies. For example, a very simple form of self-assessment is to award a grade/mark to own work (sometimes called "self-evaluation[1]" or "self-grading"). A more complex form of self-assessment may involve a rigorous analysis of strengths and weaknesses, as well as the formulation of formative feedback, in relation to explicit criteria (Andrade, 2010).

Actually, a large number of calls have been issued recently for moving away from the simpler forms of self-assessment, where students are merely asked to score or grade themselves, as opposed to making qualitative judgments about their own performance (Andrade, 2010; Boud & Falchikov, 1989; Eva & Regehr, 2005). In particular, the formative assessment agenda has contributed to changing the focus of self-assessment research. Self-assessment is a fundamental component of formative assessment since, as stated by Sadler (1989), it is ultimately the student herself that has to "close that gap" between a current performance (as revealed by assessment) and the desired standard. A student who only follows the teachers' prescription without understanding its purpose will not learn to monitor and self-adjust her work. As emphasized by Black and Wiliam (1998), self-assessment is therefore not "an interesting option or luxury" (p. 54–55), it is essential to productive learning, and empirical research supports this idea.

1.1.1. Self-assessment effects on student learning and performance

Whether self-assessment has an impact on student learning has been explored in a number of studies. For instance, in a narrative review of research about self-assessment, Topping (2003) concluded that there is evidence that self-assessment can result in improvements in the effectiveness and quality of learning. In a more recent publication, Brown and Harris (2013) come to a similar conclusion by reviewing 23 studies, including a wide variety of operationalizations of self-assessment. The effects range from −0.04 to 1.62 (Cohen's *d*) with a median effect between 0.40 and 0.45. The authors also note that self-assessment seems to improve student performance across a range of grade levels and subject areas, but that it seems to be the implementation and the complexity of the self-assessment intervention, rather than the type, which generates the positive effects. As an example, rubric guided judgement as a type of self-assessment, has been shown to result in both very high effect sizes, as well as very low (and even negative) effect sizes. While students in the latter case (i.e. low effect sizes) participated in two self-assessment lessons, during which they used a rubric for essay writing to assess the quality of their drafts (Goodrich Andrade & Boulay, 2003), the students in the former (i.e. high effect size) also received rubrics articulating assessment criteria for essay writing. However, these students also participated in generating a list of criteria from a model paper (Andrade, Du, & Wang, 2008).

Besides rubrics, which have been suggested to support student learning if combined with self-assessment or other meta-cognitive activities (Panadero & Jonsson, 2013), feedback has also been shown to influence the relationship between self-assessment and learning. In a meta-analysis by Sitzmann, Ely, Brown, and Bauer (2010), the correlation between self-assessment and learning was stronger for courses that included feedback ($r = 0.28$) than for courses that did not include feedback ($r = 0.14$). Furthermore, when students self-assessed once, or on multiple occasions without receiving feedback, the relationship with learning was weaker (0.29 and 0.30 respectively) as compared to situations where students received external feedback on their accuracy ($r = 0.51$).

1.1.2. Self-assessment effects on self-regulated learning and self-efficacy

The abovementioned educational gains from self-assessment are suggested to be related to the enhancement of ownership of learning and the use of SRL strategies. This means that self-assessment is thought to contribute to student learning by, for instance, enhancing the clarity of the learning goals, involving students in monitoring the learning process, and facilitating reflection about the final product or learning outcome (Brown & Harris, 2013; Nicol & McFarlane-Dick, 2006; Panadero & Alonso-Tapia, 2013).

Although the relationship between self-assessment and SRL strategies has been claimed theoretically since the beginning of the formative assessment agenda (e.g. Black & Wiliam, 1998), there is a need to review what is known from empirical research. For example, Topping (2003) concluded that the evidence for self-assessment affecting students' SRL was "encouraging" and more research was needed. In a similar vein, Brown and Harris (2013) conclude that the research evidence for the connection between self-assessment and SRL is "not robust" and that it is still unclear which students benefit from training in self-assessment.

Regarding self-efficacy, according to social cognitive theory (Bandura, 1986), there are four factors increasing an individual's self-efficacy, which are own experiences of successful performance, watching others succeed, encouragement, and physiological factors. The basic presumption as to why self-assessment has an effect on students' self-efficacy is that by gaining a deeper understanding of the requirements of the task at hand, students are likely to perform better and therefore to experience successful performance. This, in turn, is thought to trigger feelings of worth and a perception of improved capability, which will finally impact on the level of self-efficacy (e.g. Paris & Paris, 2001).

In the self-assessment literature, however, the findings regarding self-efficacy are mixed. As an example, in a study by Andrade et al. (2009) generating a list of criteria from a model essay and using a rubric to self-assess drafts, was shown to

[1] Not to be confounded with the concept of self-evaluation in SRL research. Here we are referring to self-assessment studies in which students are asked to estimate their grade, also called self-scoring, with no relationship to SRL research.

increase students' self-efficacy. However, on average, all students' self-efficacy increased, including students in the control group. The increase was larger in the treatment group, but not significantly. There was also a difference between genders, where girls' average self-efficacy for writing tended to be higher as compared to the boys, especially in the beginning. Since the boys' self-efficacy increased more, as compared to the girls, there was not a statistically significant difference at the end of the intervention. Similarly, in a series of studies on the impact of rubric supported self-assessment on students' self-efficacy, Panadero (2011) found that self-efficacy was impacted by the use of rubrics, but only in one of the three studies. The mixed findings makes it interesting to explore this relationship further through meta-analytic methodology. In the coming sections, the two dependent variables are presented in more detail.

1.2. Self-regulated learning

According to one of the most widely used definitions, SRL is: "self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals" (Zimmerman, 2000, p. 14). Self-regulated learning has become one of the most prevalent educational theories to explain students' achievement as it includes a large number of variables related to learning, such as goal orientation, task specific strategies, metacognitive strategies, attribution theory, etc. (e.g. Panadero, 2017). As an example, Richardson, Abraham, and Bond (2012) performed a meta-analysis based on 11 different SRL components (such as rehearsal, effort regulation, and help seeking), showing that the use of SRL strategies was a significant predictor of academic performance. Similar results have also been reported elsewhere (e.g. Broadbent & Poon, 2015).

There are several models of SRL, but most of them include a preparatory phase, a performance phase, and an appraisal phase, each consisting of different sub-processes (Panadero, 2017). In Zimmerman's (2000) cyclical model, which is one of the most cited in the SRL literature, these phases are called forethought, performance, and self-reflection.

1.2.1. Self-regulated learning effects on student learning and performance

A number of previous meta-analyses have investigated the influence of SRL on student learning and performance (for a meta-review, see Panadero, 2017). For instance, Hattie, Biggs, and Purdie (1996) examined 51 intervention studies aiming to enhance student learning through SRL. In their meta-analysis, the overall effect size for student performance was 0.57. Dignath, Büttner, and Langfeldt (2008) analyzed the effects of SRL on primary school students' academic achievement, cognitive and metacognitive strategy application, and motivation. Results from 48 comparisons show that SRL interventions were effective, even at the primary school level (the overall mean effect size for academic performance was $d = 0.62$). For motivational aspects (including self-efficacy), the mean effect size was even higher ($d = 0.76$). In a later study (Dignath & Büttner, 2008), encompassing both primary- and secondary school (357 effect sizes in total), the mean effect size for academic performance was 0.61 for primary school and 0.54 for secondary school, suggesting that SRL can be fostered effectively at both primary and secondary school level. However, the effect for motivational aspects was considerably lower for secondary students.

In the abovementioned meta-analyses, SRL strategies are clustered into larger categories, making it impossible to disentangle the effects from individual SRL components. In a study by Sitzmann et al. (2010), however, 16 different SRL components were investigated (k = 430), reporting that goal level, persistence, effort, and self-efficacy were the constructs with the strongest effects on learning for adults.

1.2.2. Self-regulated learning and self-assessment

As mentioned above, a strong theoretical connection has been suggested between SRL and self-assessment (e.g. Andrade, 2010; Panadero & Alonso-Tapia, 2013). For example, two of the SRL subprocesses in Zimmerman's (2000) model (self-monitoring and self-evaluation) have a clear similarity to self-assessment, since they are oriented towards assessing own performance. Furthermore, self-assessment has been proposed to be key for the internalization of standards, so that students can regulate their own learning more effectively (Paris & Paris, 2001, p. 95), which involves the first phase of SRL (forethought).

Consequently, self-assessment does not only affect the self-reflection phase, but also the forethought phase (for instance when providing the students with assessment criteria, so that they are able to set realistic goals for the task) and the performance phase (since monitoring can be done with more accuracy, as there is a clearer understanding of the final product/learning outcome) (Andrade, 2010; Panadero & Alonso-Tapia, 2013). The importance of integrating planning with self-assessment related processes has been shown by Dignath et al. (2008), who used meta-analytic methodology to explore SRL interventions in primary school settings. Findings show that metacognitive interventions aiming at a combination of planning and monitoring (P&M), or planning and evaluating, were the most successful to enhance students' strategy use ($d = 1.50$ and 1.46 respectively) and had significant effects on motivational outcomes ($d = 0.58$ and 1.59), as well as academic performance (P&M $d = 0.78$) (Dignath et al., 2008 p. 115, Table 9). It has therefore been argued that interventions to promote self-assessment should be initiated before students start performing the task, for instance by providing the students with assessment criteria, so that the students can plan, monitor, and evaluate with the help of these criteria (e.g. Andrade & Valtcheva, 2009; Jonsson, 2014). One of the most direct pieces of empirical evidence for self-assessment having an effect on all phases of the SRL cycle comes from a study by Panadero and Romero (2014). In this study, the use of explicit assessment criteria was shown to have a significant impact on the forethought phase ($\eta^2 = 0.257$), the performance phase ($\eta^2 = 0.084$),

and the self-reflection phase ($\eta^2 = 0.217$). As can be seen, the preparatory/forethought phase activation of learning strategies was affected the most.

1.3. Self-efficacy

Self-efficacy is the belief about the personal capabilities to perform a task and reach the established goals (Bandura, 1997). The concept was introduced by Bandura (1977), who also made developments within the social cognitive theory (1986), making a major impact in education and educational psychology (van Dinther, Dochy, & Segers, 2011). Self-efficacy has been found to be the strongest predictor of academic performance in tertiary education in two meta-analyses (Richardson et al., 2012; Robbins et al., 2004), and a more recent study also report effect sizes of similar magnitude (Honicke & Broadbent, 2016).

The influence of self-efficacy on the conceptualization and development of SRL has been crucial (Zimmerman, 2000). Over the years, self-efficacy has become one of the most important variables not only in research on motivation, but also in research on SRL (e.g. Schunk & Usher, 2011), and self-efficacy has therefore been incorporated into SRL models (Panadero, 2017). For example, self-efficacy is an essential sub-process in the models by both Zimmerman (2000) and Pintrich (2000). As will be evident from this meta-analysis, self-assessment literature includes a significant number of studies analyzing self-efficacy in isolation from SRL (e.g. Sitzmann et al., 2010, reporting a moderate correlation between self-assessment and self-efficacy based on 32 effect sizes), and to a lesser extent, in combination.

1.4. Moderating variables

There are a number of variables that may influence the effects of self-assessment on SRL and self-efficacy explored in the literature. The four moderating variables included in this meta-analysis are presented below.

First, gender is a factor that is likely to influence the effects of self-assessment on SRL and self-efficacy, since gender differences have been reported for all three variables. However, the findings are not conclusive for all variables and the amount of research evidence also differs greatly. Starting with the effects of gender on self-assessment, this has not been extensively studied (e.g. Brown & Harris, 2013; Wright & Houck, 1995) and what is known is mostly in relation to the accuracy of self-scoring. Gender effects on self-assessment is therefore an under-researched area and several reviews have recommended future research to explore this effect (e.g. Boud & Falchikov, 1989; Brown & Harris, 2013; Panadero & Jonsson, 2013). There is also evidence supporting the existence of gender differences in SRL, which, according to Bussey (2011), do not relate to differences in SRL capabilities, but to self-efficacy and expectations (i.e. two SRL sub-processes). As opposed to self-assessment, gender differences in relation to self-efficacy have been explored in a number of reviews and meta-analyses. In a meta-analysis by Whitley (1997), it was reported that males had a self-efficacy for computers, which was 0.41 standard deviations above the average for females. Pajares conducted two narrative reviews (2003, 2005) on self-efficacy for writing and mathematics, finding that males tended to have higher self-efficacy for mathematics, while females tended to have higher self-efficacy for writing during middle school, but that this difference tend to decrease at older ages. However, these results need to be re-interpreted in the light of the meta-analysis by Huang (2013), which subsumes the work by Whitley (1997) and also uses a more rigorous meta-analytical methodology as compared to Pajares (2003, 2005). The findings from Huang (2013) were a general effect of 0.08 favoring males, therefore a "small difference" as interpreted by the author, and differences across subjects. Gender is therefore a crucial variable to be considered in the current meta-analytic review.

Second, another factor potentially influencing the effects of self-assessment on SRL and self-efficacy is age (or educational level) of the students. In relation to self-assessment, such research is more or less non-existing and no previous studies comparing self-assessment skills across different ages have been found, although there have been calls to approach self-assessment training as a skill that need practice to develop (Panadero, Brown et al., 2016, Panadero, Jonsson et al., 2016). In relation to the development of SRL strategies across age and educational level this is a growing area of research. For instance, studies show that already young students can be taught and develop SRL strategies (Perry & Rahim, 2011; Whitebread, Bingham, Grau, Pino Pasternak, & Sangster, 2007). Furthermore, meta-analyses by Dignath and Büttner (2008), Dignath et al. (2008), and Hattie, Biggs, and Purdie (1996) found differential effects of SRL interventions aiming for primary and secondary school students. First, effect sizes for academic achievement from interventions aiming to enhance SRL were larger for younger students (i.e. primary- and lower secondary school) than for older students (i.e. secondary- and higher education students) (Dignath & Büttner, 2008; Hattie et al., 1996).[2] Second, Dignath and Büttner (2008) found a larger effect on motivational outcomes for primary students, as compared to secondary students, from SRL interventions. And finally, when exploring strategy use, the effects were reversed (i.e. secondary education students benefited more, as compared to primary students). Age was also included as a variable in the meta-analysis by Huang (2013), who reported that gender differences in self-efficacy tend to increase as age increases.

Third, as reported by Panadero, Brown et al. (2016), there is a large number of different self-assessment practices, which might have differential effects. As an example, in the list of effect sizes for the relationship between self-assessment and learning, as presented by Brown and Harris (2013, p. 382), studies with a similar design can be found both at the top and the

[2] Hattie et al. (1996) found larger effects for primary and lower secondary education as compared to older students (secondary- and higher education students). Dignath and Büttner (2008) found larger effects for students in primary school as compared to students in secondary education.

bottom of the list (e.g. self-assessing writing with a rubric), but which differs in the comprehensiveness of the self-assessment intervention. As suggested by for instance Panadero and Jonsson (2013), as well as the work by Sitzmann et al. (2010), different self-assessment components may therefore have differential effects. Furthermore, from previous research, rubrics could be expected to have a larger impact on performance/avoidance SRL as compared to learning SRL. It is therefore of great need to compare the effects of self-assessment interventions with different components and intensity in relation to effects on SRL and self-efficacy.

A fourth potential moderating variable is the agent who implements the intervention (i.e. the teacher or the researcher). This moderator comes from SRL research, where it has been found that when the SRL interventions were conducted by the researcher, the effect size was higher as compared to when teachers were in charge of the implementation (Dignath et al., 2008). Here it will be explored for both SRL and self-efficacy.

1.5. Measurement of self-regulated lerning and self-efficacy

The choice of measurement instruments for SRL and self-efficacy is also a potential moderating variable, since different types of instruments have been shown to provide different results. For instance, in the study by Dignath et al. (2008), intervention studies using questionnaires reported a higher impact on SRL, as compared to other types of measurements (e.g. multiple choice test). However, which instruments to use is a controversial issue in the field. There is a critique of self-reported data (Boekaerts & Corno, 2005; Veenman, 2011), as well as a defense for other types of self-reported data (Samuelstuen & Bråten, 2007). There is also a tension between off-line and online measurements (Winne & Perry, 2000), as well as suggestions for new ways of measuring SRL (Panadero, Klug, & Järvelä, 2016). In this meta-analysis, SRL measurements have not been used as a moderating variable. Instead, due to its importance, this variable has been decompounded into three different (dependent) SRL variables.

This tension between different measurement instruments is not visible in the self-efficacy literature. As shown in the meta-analysis by Honicke and Broadbent (2016), self-report has been the primary method to measure self-efficacy. Specifically, the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich, Smith, Garcia, & McKeachie, 1991) has been the most frequently used instrument. An interesting observation is therefore that even though the MSLQ was created as an instrument to measure SRL, it is to a large extent used to measure self-efficacy. In sum, the implication for this meta-analytic review is that there was no need to make a distinction between different measurement instruments for self-efficacy, since all studies in the sample used the same measurement method (i.e. questionnaires).

1.6. Research questions

According to the research reviewed above, the empirical support for the effects of self-assessment on students' SRL strategies and self-efficacy is seen as promising, but not conclusive. There are, however, some concerns making it difficult to judge the validity of this conclusion. First, there is a number of studies about the effects of self-assessment on SRL that are not included in the abovementioned narrative reviews by Topping (2003) and Brown and Harris (2013). Second, those reviews do not make any distinction between: (a) different research designs (such as qualitative research, quasi-experimental, and experimental design); (b) different measures of SRL; or (c) SRL and self-efficacy. Thirdly, the review by Brown and Harris (2013) only covered K-12, and provided meta-analysis only for the relationship of self-assessment to academic achievement.

Given the inconclusive findings from the primary research, and the limitations from recent reviews, this study aims to use meta-analytic methodology to explore the evidence accumulated on the effects of self-assessment on SRL and self-efficacy. Specifically, this study will explore the following research questions:

RQ1: Do self-assessment interventions have an effect on students' SRL?
  RQ1a. Is there a differential effect based on different SRL instruments and constructs?
RQ2: Do self-assessment interventions have an effect on students' self-efficacy?
RQ3: Do the moderating variables gender, age/educational level, self-assessment intervention, and implementation agent influence the effects of self-assessment on students' SRL and/or self-efficacy?

2. Method

2.1. Selection of studies

The search was conducted at two different occasions. The initial search was performed in June 2015. The first author conducted an independent search using his university's access to PsycINFO, ERIC, and Google Scholar. The second author performed an independent parallel search on a common interface called "Summon", which includes all available databases his university subscribes to. The search included, but were not limited to, databases such as PsycINFO, PubMed, ScienceDirect, Web of Science, and Google Scholar. Both authors used the following combinations of keywords: Self-regulated learning + self-assessment; Self-regulated learning + self-evaluation; Self-regulated learning + monitoring; Self-regulation + self-assessment; Self-regulation + self-evaluation; Self-regulation + monitoring; Self-assessment + self-

efficacy; and Self-evaluation + self-efficacy. Reference lists of empirical and review articles were also examined for additional references. A second search was performed in April 2016 using the same databases and keywords to include any recently published publications on the topic.

The inclusion criteria used were that: (a) the study included empirical results of self-assessment interventions in relation to SRL and/or self-efficacy; (b) the study had at least one control group; (c) the study had been peer-reviewed (i.e. either journal articles or dissertations); and (d) the study was published in English.

Fig. 1 presents a flow chart of the systematic review. In total, 142 records were identified through database and manual searches. After removing duplicates, the abstracts of the 65 remaining publications were screened in order to select only empirical studies. The full texts of 36 empirical studies were read and assessed. After excluding the studies not meeting all inclusion criteria, 19 studies were included in the meta-analysis. This meta-analysis complies with PRISMA[3] guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009).

The following information was collected from the selected articles: general information (authors' names, year of publication, country, aim, research questions, and hypotheses), sample characteristics (sample size and gender proportions), study design (including the variables used), procedure (including self-assessment intervention), results, and conclusions. Table 1 present a summary with the more important information from the included publications.

The size of the studies in the sample range from 30 to over 300 participants (mean 115.25). In total, the studies encompass 2,305[4] students from primary school to higher education. Studies involving students from primary school are few (2), while studies from secondary school and higher education are more evenly represented in the sample (10 and 8 studies respectively). The overall mean age in the sample is 17.5 years (range 10.5–26 years) and approximately 70 percent of the sample is female.

In studies not reporting student characteristics, or not reporting specifically enough, first authors were contacted and asked to provide the required information. However, not all authors responded to this query.

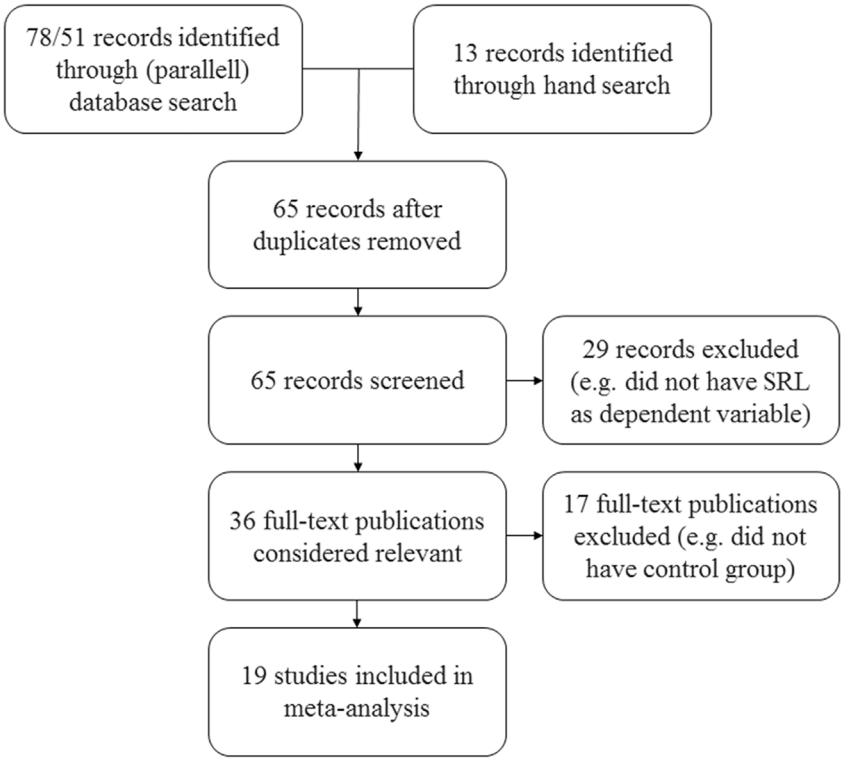

Fig. 1. Flowchart of included studies handling.

[3] PRISMA stands for Preferred Reporting Items for Systematic reviews and Meta-Analyses. More information in the above reference or at the PRISMA website: http://prisma-statement.org/.

[4] By adding the number of participants in the four different meta-analyses, there would be a total of 2987 participants. However, since some of the participants are included in more than one meta-analysis, the more conservative number is reported in the text.

Table 1
Summary of included studies.

First Author (Year)	Sample Size	% Female	Subject	Educational level	SRL variable	SRL instrument	Design	SSA intensity	Average Effect Size
Schunk and Ertmer (1999)	22	91	Computer education	Higher education	Learning SRL	Questionnaire	Experimental	Low-medium	−0.15
		88							+0.70
									+0.57
Panadero et al. (2012)	20	53	Geography	Secondary school	Learning SRL	Questionnaire	Experimental	High	+0.73
									−0.17
									+0.17
									−0.94
Panadero et al. (2013)	49	84	Computer education	Higher education	Learning SRL	Questionnaire	Experimental	High	+0.13
DiGiacomo (2014)	30	50	Mathematics	Secondary school	Learning SRL	Questionnaire	Experimental	Medium	+0.24
Kahrizi et al. (2014)	40	73	EFL	Higher education	Learning SRL	Questionnaire	Experimental	High	+0.25
Panadero and Romero (2014)	218	87	Learning and Development	Higher education	Learning SRL	Questionnaire	Quasi-experimental	Medium	+0.20
Mahlberg (2015)	186	55	Various	Higher education	Learning SRL	Questionnaire	Quasi-experimental	Low-medium	+0.39
Goodrich (1996)	40	58	Biology	Secondary school	SRL measured qualitative	Think-aloud protocols	Quasi-experimental	Medium-high	0.00
Panadero et al. (2012)	80	53	Geography	Secondary school	SRL measured qualitative	Think-aloud protocols	Experimental	High	+1.96
Panadero and Romero (2014)	218	87	Learning and Development	Higher education	SRL measured qualitative	Open questions	Quasi-experimental	Medium	+0.93
Panadero et al. (2012)	20	53	Geography	Secondary school	Negative SRL	Questionnaire	Experimental	High	−1.59
									+0.56
									−2.58
									−1.42
Panadero et al. (2013)	49	84	Computer education	Higher education	Negative SRL	Questionnaire	Experimental	High	−0.20
Panadero and Romero (2014)	218	87	Learning and Development	Higher education	Negative SRL	Questionnaire	Quasi-experimental	Medium	+0.53
Zimmerman and Kitsantas (1996)	20	100	Dart	Secondary school	Self-efficacy	Questionnaire	Experimental	Medium	+1.00
									+0.51
Kitsantas and Zimmerman (1998)	20	100	Dart	Secondary school	Self-efficacy	Questionnaire	Experimental	Low	+2.10
									+4.89
									+2.70
									+4.97
Schunk and Ertmer (1999)	22	91	Computer education	Higher education	Self-efficacy	Questionnaire	Experimental	Low-medium	+1.37
									+0.20
									+0.36
Olina and Sullivan (2002)	120		Psychology	Secondary school	Self-efficacy	Questionnaire	Experimental	Low-medium	+0.05
Kitsantas, Reiser, and Doster (2004)	24	57	Computer education	Secondary school	Self-efficacy	Questionnaire	Experimental	Low-medium	+0.05
									+1.19
									−0.31
									+0.64
Olina and Sullivan (2004)	170		Psychology	Secondary school	Self-efficacy	Questionnaire	Experimental	Low-medium	+0.71
Coronado-Aliegro (2007)	104	54	Spanish as a Foreign Language	Higher education	Self-efficacy	Questionnaire	Quasi-experimental	Medium	+0.10
Ramdass and Zimmerman (2008)	42	52	Mathematics	Primary school	Self-efficacy	Questionnaire	Experimental	Medium	+0.19
Andrade et al. (2009)	268	63	Essay writing	Primary and secondary school	Self-efficacy	Questionnaire	Quasi-experimental	High	+0.22
Nbina and Viko (2010)	192	53	Chemistry	Secondary school	Self-efficacy	Questionnaire	Quasi-experimental	Low-medium	+3.02
Panadero et al. (2012)	20	53	Geography	Secondary school	Self-efficacy	Questionnaire	Experimental	High	+0.22
									−0.21
									−0.04
									+0.53
Alishah and Dolmaci (2013)	54	–	EFL/ESL	Higher education	Self-efficacy	Questionnaire	Quasi-experimental	Low-medium	+0.07
Panadero et al. (2013)	49	84	Computer education	Higher education	Self-efficacy	Questionnaire	Experimental	High	+0.50
Baleghizadeh and Masoun (2014)	57	100	EFL	Higher education	Self-efficacy	Questionnaire	Quasi-experimental	Low-medium	+0.53
Panadero and Romero (2014)	218	87	Learning and Development	Higher education	Self-efficacy	Questionnaire	Quasi-experimental	Medium	+0.13

Although subjects such as computers and language occur slightly more often in the sample, there is large variation in the content domains focused, such as geography, mathematics, and psychology. There are also a couple of studies focusing on psychomotor skills, such as throwing dart. Eight studies in the sample used a quasi-experimental design, while the remaining studies used experimental conditions.

2.2. Dependent and moderating variables

2.2.1. Self-regulated learning variables

Three different types of SRL measurements have been used in the studies included in this meta-analytic review. These variables are defined by a combination of SRL constructs and measurement instruments.

The first variable is "Learning SRL", which encompasses measurements of SRL via questionnaires, representing actions associated with learning strategies serving positive self-regulatory functions for students' learning, such as metacognitive strategies, motivation, and emotions. A typical example of an instrument is the MSLQ (Pintrich et al., 1991), which includes scales such as rehearsal, elaboration, or managing time and study environment. This is also the instrument that is most frequently (Roth, Ogrin, & Schmitz, 2016). The higher the value on these scales, the more the students are thought to use regulatory actions that will enhance their learning.

The second SRL variable is "Negative SRL", which has its theoretical foundations in the work by Boekaerts ans Corno (2005) and Kuhl (2000). Like Learning SRL, in self-assessment research Negative SRL has also been measured by questionnaires, but associated with negative emotions and stress, encompassing regulatory actions directed by anxiety, external pressure to perform, and task avoidance. An increase in Negative SRL is therefore thought to be detrimental for students' learning. The main instrument for measuring Negative SRL is the Emotion and Motivation Self-Regulation Questionnaire (EMSR-Q), initially used in Panadero (2011), where one of the scales was termed "Performance/Avoidance self-regulation". The EMSR-Q was later validated in a study (Alonso-Tapia, Panadero, & Ruiz, 2014) with a significantly higher sample (664 participants). In this validation study, it was decided to confirm the distinctness of this SRL variable in relation to the Learning SRL-Q, because the effects were completely opposite: Higher values on the Avoidance SRL-Q meant more negative self-regulatory actions that are detrimental for learning with a negative correlation to Learning SRL scores. For this reason, it was more informative to run separate meta-analyses on Learning SRL and Negative SRL in this study. Three of the studies in the sample explored Negative SRL.

Finally, the third SRL variable is called "SRL measured qualitatively" and includes qualitative data of SRL, such as think-aloud protocols and open questions. Only three studies in the sample have used this type of measurement, but due to the current debate on validity in SRL measurements (Boekaerts & Corno, 2005; Panadero, Klug et al., 2016; Winne & Perry, 2000) it was decided do explore "SRL measured qualitatively" as a separate variable.

2.2.2. Self-efficacy

All the studies in the sample used self-reported questionnaires or scales to measure self-efficacy, which is also the general tendency in self-efficacy research (Honicke & Broadbent, 2016). As the construct and measurement is similar across studies, only one category of self-efficacy was employed in this review.

2.2.3. Gender

Student gender was used as a moderating variable in the meta-analyses. The percentage of females was computed to run the analyses. As can be seen from Table 1, females were in majority in most of the studies and in some studies the entire sample was female.

2.2.4. Age and educational level

Age/educational level were used as moderating variables due to their potential influence on the development of self-assessment, SRL, and self-efficacy. As can be seen in Table 2, a range of educational level courses was reported in some studies. In those cases, an average of students' age was used. Regarding educational level, there were three categories: primary, secondary, and higher education.

2.2.5. Self-assessment intervention categories

In order to be categorized as a self-assessment intervention, students have to assess the quality of their own performance. This means that several concepts and measures, which may be referred to as self-assessment elsewhere, were not included. Examples of concepts and measures that were not included are:

—Evaluations of the quality of, or the satisfaction with, instruction. The focus of the (self-) assessment has to be student performance — not instruction.
—Estimations of own general competence, self-concept, or understanding. The (self-) assessment has to address the quality of performance on specific tasks/assignments — not general (latent) constructs.

Table 2
Sample age.

Study	Age			
	Age extracted from the publication (SD)	Grade extracted and the corresponding age (Country)	Calculated for this review	Offered by the author
Goodrich (1996)		7th (USA) 12-13	12.5	
Zimmerman and Kitsantas (1996)	15.8			
Kitsantas and Zimmerman (1998)	15.16			
Schunk and Ertmer (1999)	21			
Olina and Sullivan (2002)				The author provided the following info: "majority of the students in the study were 16 year olds"
Olina and Sullivan (2004)		10-11th (Latvia) 16-17	16.5	
Kitsantas et al. (2004)	15.2			
Coronado-Aliegro (2007)	23			
Ramdass and Zimmerman (2008)		5-6th (USA) 10-12	11	
Andrade et al. (2009)		3-7th (USA) 8-13	10.5	
Nbina and Viko (2010)				
Panadero et al. (2012)	15.9 (0.92)			
Alishah and Dolmaci (2013)				
Panadero et al. (2013)	20.6 (2.1)			
Panadero and Romero (2014)	22.17 (3.92)			
Baleghizadeh and Masoun (2014)	26 (3.28)			
DiGiacomo (2014)		6-7th (USA) 11-13	12	
Kahrizi et al. (2014)	20.5			
Mahlberg (2015)	22.96 (7.79)			

In a most rudimentary form, students can be asked to self-assess without any aids. Self-assessment interventions can also be more wide-ranging and supporting, for instance by including instruments (such as criteria or rubrics), training, and/or feedback on self-assessment performance. The classification of self-assessment interventions in this study is based on the following different categories that were identified from the included publications (Table 3 contains information on categories for each study):

1. Monitoring (M). In studies like Zimmerman and Kitsantas (1996), students are provided a protocol or log to record their performance.
2. Self-evaluation (SEV). In studies like Schunk and Ertmer (1999), Alishah and Dolmaci (2013), and Baleghizadeh and Masoun (2014), students fill in self-assessment or self-efficacy questionnaires.
3. Self-assessment instruments (SAI). In studies like Panadero, Alonso-Tapia, and Huertas (2012), 2013), and Kahrizi, Farahian, and Rajabi (2014), students are provided with instruments (such as rubrics, scripts, or checklists) that can be used to plan, monitor, and/or self-assess own task performance. In all but two studies (Goodrich, 1996; Mahlberg, 2015), students are also provided with feedback.

Table 3
Self-assessment intervention categories.

Study	Monitoring	Self-evaluation	Self-assessment instrument	Feedback and revision	Self-assessment training	Self-assessment intensity
Goodrich (1996)	0	0	1	0	0	**1**
Zimmerman and Kitsantas (1996)	1	0	0	0	0	**1**
Kitsantas and Zimmerman (1998)	1	0	0	0	0	**1**
Schunk and Ertmer (1999)	0	1	0	0	0	**1**
Olina and Sullivan (2002)	0	0	0	1	0	**1**
Kitsantas et al. (2004)	1	0	0	0	0	**1**
Olina and Sullivan (2004)	0	0	0	1	0	**1**
Coronado-Aliegro (2007)	0	1	0	0	0	**1**
Ramdass and Zimmerman (2008)	1	1	0	0	0	**2**
Andrade et al. (2009)	0	0	1	1	0	**2**
Nbina and Viko (2010)	1	0	0	1	1	**3**
Panadero et al. (2012)	0	0	1	1	0	**2**
Panadero et al. (2013)	0	0	1	1	0	**2**
Alishah and Dolmaci (2013)	0	1	0	0	0	**1**
DiGiacomo (2014)	1	1	0	1	0	**3**
Kahrizi et al. (2014)	0	0	1	1	1	**3**
Panadero and Romero (2014)	0	0	1	1	0	**2**
Baleghizadeh and Masoun (2014)	0	1	0	0	0	**1**
Mahlberg (2015)	0	0	1	0	0	**1**

4. Feedback and revision (FR). In studies like Panadero et al. (2013), DiGiacomo (2014), and Kahrizi et al. (2014), students are provided with teacher feedback and are then allowed to revise their work. In studies like Panadero et al. (2012) and DiGiacomo (2014), students are provided feedback and are then allowed to do other (similar) tasks. This condition is in most cases associated with some self-assessment instrument, but there are exceptions.
5. Self-assessment training (SAT). In Kahrizi et al. (2014) and Nbina and Viko (2010) students were exposed to training in self-assessment during an extended period of time. Because there are only two studies using this type of self-assessment, this category will not be included in the meta-analyses.

All studies included in this review fit within this categorization, but most of them use different combinations, such as SAI + FR (which is the most common). Other combinations are: (M + SEV + FR), (M + SEV), and (M + FR). For that reason, a general self-assessment intervention score variable was calculated to explore whether more comprehensive interventions have larger effects.

The initial coding of the selected articles was done by the second author. Then the first author independently coded eight articles. Perfect agreement was reached at all occasions (Cohen K = 1) as the specificity of the categories is high.

2.2.6. Implementation agent

Current SRL research suggests that who is in charge of implementing the self-assessment intervention may influence the outcome. Due to the low number of studies in which the teacher was the implementation agent, this variable could only be used in relation to self-efficacy.

2.3. Statistical methods

To estimate the effects of the self-assessment interventions Cohen's *d* was used, or standardized mean difference (Cohen, 1988), dividing the difference between the means by the pooled standard deviation. Effect sizes were coded such that positive values indicate better performance for the groups with self-assessment, as compared to the groups without self-assessment. As the standard deviation was not reported in two of the studies (Olina & Sullivan, 2002, 2004), it was estimated from other statistics in the papers. In addition, in one of these studies (Olina & Sullivan, 2004) the sample size of one of the four conditions was not reported. Equal size for the four groups was therefore assumed.

Meta-analyses were conducted assuming a random-effects model. A random (rather than a fixed) effects model was chosen for two main reasons (Borenstein, Hedges, Higgins, & Rothstein, 2010). First, a random effects model allows for generalizing the results of the effect of self-assessment beyond the specific set of studies included here. Second, random effects models are more conservative concerning statistical inferences as compared to fixed effects models. Additionally, the inter-study, or specific, variance was estimated by the maximum likelihood method.

Statistical analyses of the heterogeneity were performed with Review Manager (2012), the SPSS macros by Lipsey and Wilson (2001), and the R package METAFOR (Wichtbauer, 2010). The forest plots were obtained with Review Manager (2012). Combined estimates of the effect size indices weighting were obtained by the inverse of variance method. The procedures provided by Hedges and Olkin (1985) were employed for analyzing the sample of effect size values: Q test for homogeneity, Q_b for categorical moderators, and Q_R for continuous moderators.

2.4. Publication bias

Initially, the threat of publication bias was intended to be explored for all four meta-analyses of the main dependent variables. However, the small number of studies in two of them (Negative SRL k = 6, and SRL measured qualitatively k = 3) hindered any reasonable use of the tests for asymmetry. Therefore, only the two other meta-analyses were included in the publication bias analysis. Publication bias was explored analyzing deviations from symmetry in the funnel plot (Light & Pillemer, 1984). We also calculated the Orwin's (1983) fail-safe number.

Regarding the meta-analysis of Learning SRL (k = 12) the Orwin fail-safe number is $N_{fs} = 31$. Therefore, with a number of 31 estimates of effect sizes not published, and with an average zero effect size added to our meta-analysis, the combined effect would statistically be zero. According to the criterion proposed by Rosenthal (1979), the value for tolerance is $12 \cdot 5 + 10 = 70$. As this value is higher than N_{fs} it cannot be excluded that the observed effect is due to publication bias. However, both the visual inspection of the funnel plot (Fig. 2) and the tests applied suggest that there is no publication bias in this set of studies (Kendall tau test, $p = .534$; Egger test, $p = 0.44$), as symmetry could be assumed based on the funnel plot and the Kendall and Egger tests.

Regarding the self-efficacy meta-analysis (k = 27) the Orwin fail-safe number is $N_{fs} = 488$. The Rosenthal's value for tolerance is $27 \cdot 5 + 10 = 145$. As this value is lower than N_{fs} we can exclude that the whole observed effect is due to publication bias. However, both the visual inspection of the funnel plot (Fig. 3) and the tests applied, may indicate some publication bias (Kendall tau test, $p = 0.0001$; Egger test, $p = 0.0001$). Visual inspection of the funnel plot suggests that two of the *d* values could be considered as outliers (Kitsantas & Zimmerman, 1998; b & d effect sizes). A sensibility test, excluding those two estimates, was performed, showing that Kendall's test is still significant ($p = 0.0021$), but not the Egger's test ($p = 0.3298$). In

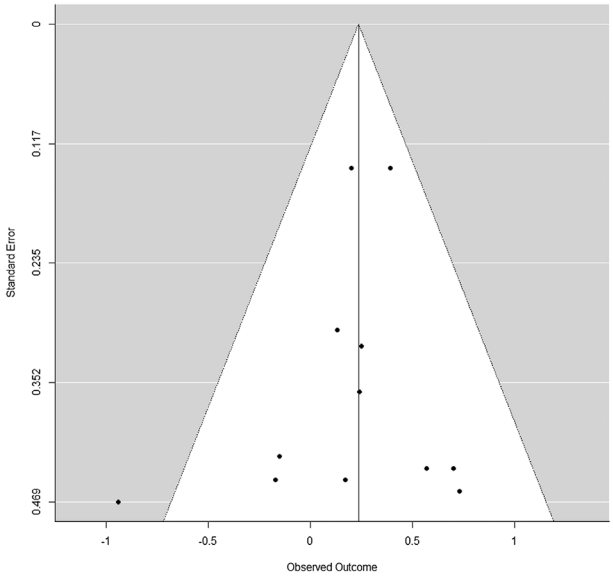

Fig. 2. Funnel plot meta-analysis of self-assessment effects on Learning SRL.

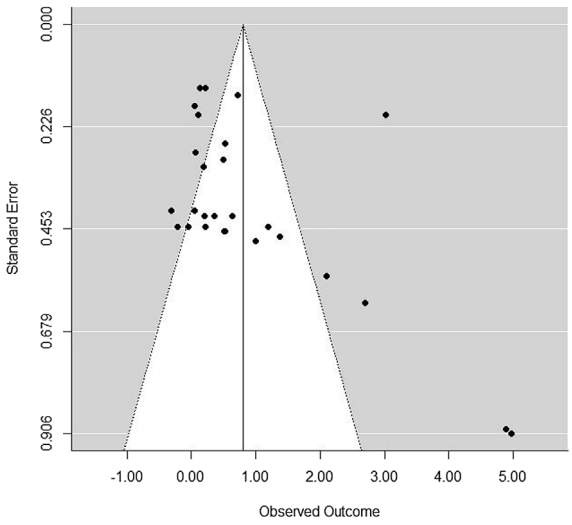

Fig. 3. Funnel plot meta-analysis of self-assessment effects on self-efficacy.

short, although there is no threat of publication bias for the existence of a true effect for self-efficacy, there might exist some over-estimation of the magnitude of the effect.

3. Results

This section is organized around the four dependent variables as outlined above: Learning SRL, Negative SRL, SRL measured qualitatively, and self-efficacy. For each of the variables the influence of self-assessment is presented, then the effects from the moderating variables (i.e. gender, age/educational level, self-assessment intervention, and implementation agent). Since a minimum of $k = 3$ was considered necessary for sufficiently consistent results from the meta-analyses, groups with $k = 1$ were not run. Furthermore, groups with $k = 2$ are reported (highlighted in grey in the tables), but not discussed in the text.

3.1. Effects of self-assessment on learning SRL

Fig. 4 presents the forest plot for effect sizes from self-assessment interventions in relation to Learning SRL. Data comes from 12 effect sizes and 369 participants, as well as from studies with different designs (i.e., experimental and quasi-experimental). The combined effect size is $d = 0.23$, 95% CI [0.08–0.39]. These results support the existence of a positive effect of self-assessment on SRL. It is important, however, to notice that the effect size is relatively small.

3.1.1. Moderating variables

As can be seen in Table 4, none of the effects of the moderating variables were statistically significant. Therefore, according to the current empirical evidence Learning SRL seems to be unaffected by these moderating variables.

3.2. Effects of self-assessment on negative SRL

Fig. 5 presents the forest plot for effect sizes from self-assessment interventions in relation to Negative SRL. Data comes from 6 effect sizes and 347 participants, as well as from studies with mostly experimental design (one study with quasi-experimental design). The combined effect size is $d = -0.65$, 95% CI [-1.52–0.22]. It should be kept in mind that Negative SRL represents self-regulatory actions thought to be detrimental for learning. In other words, lower values in relation to this variable should be interpreted as positive for students' learning. Consequently, these results support the positive effect of self-assessment on SRL. As compared to the effects of self-assessment on Learning SRL, this effect size is somewhat larger (i.e. medium effect size).

3.2.1. Moderating variables

As can be seen in Table 5, none of the effects of the moderating variables were statistically significant. However, it is noteworthy that two of them approached statistical significance, despite the low number of studies reporting on this SRL variable.

3.3. Effects of self-assessment on SRL measured qualitatively

Fig. 6 presents the forest plot for effect sizes from self-assessment interventions in relation to the SRL measured qualitatively variable. Data comes from 3 effect sizes and 338 participants, as well as from studies with either experimental or quasi-experimental design. The combined effect size is $d = 0.43$, 95% CI [-0.17–1.03]. Similar to the effects of self-assessment on Negative SRL, this is a medium effect size. These results are promising in further supporting the positive effect of self-assessment on SRL. However, this result needs more statistical power as only three studies were included in the current meta-analysis. Further research using this type of SRL measurement will be key for the field, as the results are encouraging and even a few additional studies would make a significant contribution.

3.3.1. Moderating variables

As can be seen in Table 6, none of the effects of the moderating variables were statistically significant. It is vital to remember that the number of studies exploring this SRL variable was particularly low ($k = 3$), which affects the power of the tests.

3.4. Effects of self-assessment on self-efficacy

Fig. 7 presents the forest plot for the 27 effect sizes of self-assessment interventions for self-efficacy. The combined effect size is $d = 0.73$, 95% CI [0.39–1.08]. These results support the positive effect of self-assessment on self-efficacy. As compared to the effects of self-assessment on the SRL variables, this effect size is larger.

3.4.1. Moderating variables

As can be seen in Table 7, the moderating effect of gender [$Q(1) = 5.271$; B = ; $p = 0.026$] and two of the self-assessment intervention categories (monitoring and self-assessment instrument) were statistically significant. The first finding suggests that self-assessment interventions may have a larger effect on girls' self-efficacy, as compared to boys'. The second finding suggests that monitoring is a powerful component of self-assessment in terms of promoting students' self-efficacy ($d = 1.456$) and that not using a self-assessment instrument might increase self-efficacy more.

4. Discussion

This meta-analytic review has explored the effects of self-assessment interventions on students' SRL strategies and self-efficacy, along with moderating variables assumed to have an impact on such effects. The interventions to promote self-assessment were shown to have a positive effect on students' SRL and, to a higher extent, on students' self-efficacy. Furthermore, two of the moderating variables, gender and self-assessment components, were shown to have differential effects on students' self-efficacy.

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1999a Schunk & Ertmer	5.3	0.8	11	4.7	0.9	11	3.3%	0.68 [-0.19, 1.54]
1999b Schunk & Ertmer	5.6	0.6	11	5.7	0.7	11	3.5%	-0.15 [-0.98, 0.69]
1999c Schunk & Ertmer	5.9	0.7	11	5.5	0.7	11	3.3%	0.55 [-0.31, 1.40]
2012a Panadero et al	52	2.33	10	50.3	2.33	10	3.0%	0.70 [-0.21, 1.61]
2012b Panadero et al	51.7	2.33	10	52.1	2.33	10	3.2%	-0.16 [-1.04, 0.71]
2012c Panadero et al	51.4	2.33	10	51	2.33	10	3.2%	0.16 [-0.71, 1.04]
2012d Panadero et al	51.7	2.33	10	53.9	2.33	10	2.8%	-0.90 [-1.83, 0.03]
2013 Panadero et al	43.15	6.73	20	42.24	6.78	29	7.5%	0.13 [-0.44, 0.70]
2014 Di Giacomo	3.45	0.66	15	3.3	0.61	15	4.7%	0.23 [-0.49, 0.95]
2014 Kahrizi et al	213.5	22.57	20	208.7	15	20	6.3%	0.25 [-0.38, 0.87]
2014 Panadero & Romero	31.97	7.8	111	30.44	7.14	107	33.8%	0.20 [-0.06, 0.47]
2015 Malhberg	4.63	1	124	4.25	0.92	62	25.5%	0.39 [0.08, 0.70]
Total (95% CI)			**363**			**306**	**100.0%**	**0.23 [0.08, 0.39]**

Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 11.04$, df = 11 (P = 0.44); $I^2 = 0\%$
Test for overall effect: Z = 2.94 (P = 0.003)

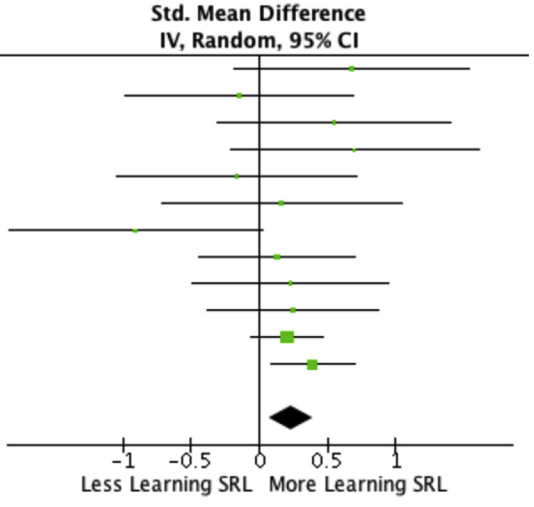

Fig. 4. Forest plot for the meta-analysis self-assessment effects on Learning SRL.

Table 4
Moderating variables effects on learning SRL.

Quantitative moderators						
Moderating variable	k	QR (1)	B			95% CI
Gender	11	0.131 (p = 0.717)	0.003			(-0.011–0.016)
Age	12	1.021 (p = 0.312)	0.029			(-0.027–0.085)
SA intervention (Total)	12	0.872 (p = 0.351)	-0.124			(-0.383–0.136)
Categorical moderators						
Moderating variable		QB	Group	k	*d*	95% CI
Ed level[a]		1.628 ($p = 0.443$)	Secondary	4	-0.042	(-0.501–0.417)
			University	7	0.284	(0.084–0.484)
Implementation agent		0.507 ($p = 0.476$)	Researcher	10	0.185	(-0.037–0.406)
			Teacher	2	0.324	(0.012–0.635)
SA intervention categories						
Self-evaluation		0.246 ($p = 0.620$)	Self-evaluation	4	0.325	(-0.097–0.747)
			No self-evaluation	8	0.206	(0.000–0.412)
SA instrument		0.246 ($p = 0.620$)	SA instrument	8	0.206	(0.000–0.412)
			No SA instrument	4	0.325	(-0.097–0.747)
Feedback and revision		2.054 ($p = 0.152$)	F&R	8	0.152	(-0.049–0.352)
			No F&R	4	0.383	(0.138–0.629)

Note: Highlighted in grey moderating variables that will not be discussed in the paper because one of the groups is below k = 2.
Values in Self-evaluation and SA instrument are the same because the group of studies is the same but reversed.
SA = self-assessment.
[a] The category with primary students was excluded from this analysis as it was k = 1.

4.1. The effect of self-assessment interventions on students' self-regulated learning

The results suggest that self-assessment interventions have a positive effect on students' SRL strategies. These effects range from small to medium in the three meta-analyses (0.23, 0.65, and 0.43). The findings differ from previous self-assessment reviews by Topping (2003) and Brown and Harris (2013), which concluded that the evidence supporting the relationship between self-assessment and SRL were not robust at that time. However, the findings from the present study are based on a more solid meta-analytic methodology, as well as a larger sample of studies specifically targeting effects on SRL strategies.

Besides the general effect, there are two other findings that warrant further attention. First, the SRL measurements provide differential effects. Second, there are no statistically significant effects of the moderating variables investigated. These findings are discussed further below.

4.1.1. Differences between self-regulated measurements

A key contribution of this study is the distinction made between different SRL measurements. A similar approach has been reported by Dignath et al. (2008), who used different measurements as moderators (p. 188 Table 15). It is, however, difficult to compare the results from these two meta-analyses since: (1) The studies included in this review mainly used questionnaires, while Dignath et al. (2008) included studies that used, for example, multiple-choice tests or simulation tasks; (2) Dignath et al. (2008) did not include studies investigating Negative SRL; and (3) this review only includes three studies measuring SRL qualitatively, which is not sufficient to make any comparisons with the "other" category in Dignath et al. (2008).

As shown in the results section, although the different SRL measurements point in the same direction, there are differences in the magnitude of the effect sizes. In the case of SRL measured qualitatively, with only three effect sizes included, more research is needed in order to draw any firm conclusions. However, in relation to Learning SRL the findings suggest that self-assessment interventions may increase positive self-regulatory actions oriented towards learning goals. This effect seems to be a consequence of self-assessment helping students to deepen their knowledge about the learning goals, how to monitor their progress, and to evaluate and correct themselves when needed (Panadero & Alonso-Tapia, 2013). Furthermore, self-assessment is seen to decrease negative self-regulatory actions, as found in relation to Negative SRL. This indicates that, by being aware of the learning goals, as well as how to monitor and evaluate own performance in relation to them, students might for instance experience less fear of failure.

It is important to explore the reasons for the larger effect of self-assessment when SRL is measured via qualitative data (0.43), as compared to questionnaires (0.24), in more detail. The use of self-reported questionnaires for measuring SRL has been contested based on the limitations of asking the students to accurately report their actions when: (a) they have been concentrating on a task and were more or less unable to record their actions; (b) they do not have any standard to compare their SRL strategies against, especially when it comes to Likert-scales (i.e. how are students supposed to know how much "very much" or "very often" is without comparing to a standard or to other students?) (Sitzmann et al., 2010); and (c) there is a general tendency of inaccuracy in such measurement situations (Eva & Regehr, 2005). In addition, some of the studies included in this meta-analytic review used general context SRL questionnaires, which do not fully capture the effect of specific interventions (Alonso-Tapia et al., 2014; Boekaerts & Corno, 2005).

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
2012a Panadero et al	20.5	2.33	10	24.2	2.33	10	15.6%	-1.52 [-2.54, -0.50]
2012b Panadero et al	19.8	2.33	10	18.5	2.33	10	16.4%	0.53 [-0.36, 1.43]
2012c Panadero et al	17.9	2.33	10	23.9	2.33	10	14.3%	-2.47 [-3.69, -1.24]
2012d Panadero et al	17.4	2.33	10	20.7	2.33	10	15.8%	-1.36 [-2.35, -0.36]
2013 Panadero et al	46.15	11.94	20	48.21	8.82	29	18.3%	-0.20 [-0.77, 0.37]
2014 Panadero & Romero	24.64	7.83	111	20.86	6.4	107	19.5%	0.53 [0.26, 0.80]
Total (95% CI)			**171**			**176**	**100.0%**	**-0.65 [-1.52, 0.22]**

Heterogeneity: $Tau^2 = 0.98$; $Chi^2 = 46.22$, df = 5 ($P < 0.00001$); $I^2 = 89\%$

Test for overall effect: Z = 1.47 (P = 0.14)

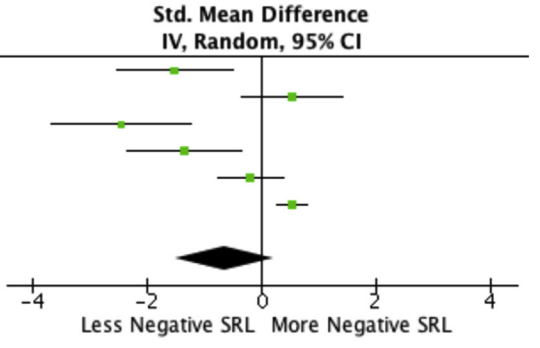

Fig. 5. Forest plot for the meta-analysis self-assessment effects on Negative SRL.

Table 5
Moderating variables effects on negative SRL.

Quantitative moderators						
Moderating variable	k	QR(1)	B			95% CI
Gender	6	3.322 (p = 0.068)	0.042			(-0.003–0.088)
Age	6	3.673 (p = 0.055)	0.256			(-0.006–0.518)
Categorical moderators						
Moderating variable		Q_B	Group	k	*d*	95% CI
Ed level		3.111 ($p = 0.078$)	Secondary	4	−1.198	(-2.158 to −0.238)
			University	2	0.180	(-1.013–1.272)

Note: Highlighted in grey moderating variables that will not be discussed in the paper because one of the groups is below k = 3. The variables self-assessment intervention and all its subcategories are not reported because all the effect sizes shared the same scores and there was not variability. Regarding the variable implementation was excluded from this analysis as one of the groups was k = 1.

Due to the abovementioned limitations of self-reported measurements, together with the tendency for students to overestimate their SRL (Boekaerts & Corno, 2005), the findings from the Learning SRL variable might actually be lower than expected, providing an explanation for the much larger combined effect size for SRL measured qualitatively. The data coming from the three SRL measured qualitative studies may in some respects be considered more valid, as compared to questionnaire data, since it is more direct (i.e., provided by the students' actions without being "filtered" through the questionnaire) and also more sensitive to the specific context. A similar pattern can be found if comparing the two studies that included a combination of qualitative and self-reported questionnaire data. Panadero and Romero (2014) used an open question providing self-reported qualitative data and Panadero et al. (2012) used thinking aloud protocols. Results from questionnaire data did not reach statistical significance in any of the studies, while results from the qualitative data did.

As mentioned before, this means that there might be problems with the validity and accuracy when using questionnaires to measure specific SRL interventions and students' inability to accurately report their SRL strategies via questionnaires (Alonso-Tapia et al., 2014; Boekaerts & Corno, 2005; Roth et al., 2016). Qualitative data may therefore have provided a more valid picture of students' actual use of SRL strategies in those two studies. Consequently, forthcoming studies using data other than Learning SRL questionnaires, could be expected to report higher effect sizes.

4.1.2. Moderating effects on self-regulated learning

The influence of the following moderating variables was explored: Gender, age/educational level, self-assessment interventions, and implementation agent. As seen by the results, none of these variables had a statistically significant impact on students' SRL strategies. However, the limited number of studies included in these meta-analyses makes it difficult to definitely rule out the possible influence of these variables. Notably, for Negative SRL the effects of gender and age were close to statistical significance. Here, however, we will make a conservative interpretation of the results and therefore only mention that there seems to be a tendency for boys and younger students to experience less Negative SRL as a consequence of self-assessment interventions.

Previous meta-analyses on SRL (Dignath & Büttner, 2008; Dignath et al., 2008; Hattie et al., 1996; Sitzmann et al., 2010) have not addressed gender effects, which means that the findings cannot be compared to any similar research. However, gender differences in SRL could have been expected since such differences have been reported in individual studies on SRL. The existence of gender differences has also been proposed in reviews on self-assessment, but primarily at a theoretical level, due to the lack of empirical support (Boud & Falchikov, 1989; Brown & Harris, 2013; Panadero & Jonsson, 2013).

Regarding age and educational level, previous meta-analyses have found differential effects of educational level, where for instance secondary education students benefited more from SRL interventions in terms of an increased use of SRL strategies, as compared to students in primary school (Dignath & Büttner, 2008). Since age and educational level did not produce statistically significant effects in this meta-analysis, the results do not align with the findings reported by Dignath and Büttner (2008). Furthermore, since Dignath and Büttner (2008) did not investigate Negative SRL, it is not possible to say whether the tendency for younger students to experience less Negative SRL as a consequence of self-assessment interventions is in conflict with their findings or not.

The third moderator was type of self-assessment intervention, which is also an aspect not previously explored in SRL meta-analyses. As suggested by Brown and Harris (2013), the effects of self-assessment interventions on student learning might depend more on the comprehensiveness of the interventions than on the type of self-assessment. However, the results show no effect of the type of self-assessment intervention on students' SRL. This is not to say that all self-assessment interventions are equally effective, just that the ones explored here had similar effects for SRL strategies. It should also be noted that these results only refer to Learning SRL, since there were too few studies in the other SRL categories.

This finding is of particular importance for research on the use of rubrics as an instrument for facilitating self-assessment. It has been suggested that the use of rubrics might decrease negative SRL strategies, while supporting student self-regulation (Brookhart, 2014; Panadero & Jonsson, 2013). According to Panadero and Jonsson (2013), as well as Jonsson (2014), the transparency provided by explicit criteria can support students' self-regulation by helping them to set realistic goals, as well as to monitor and self-evaluate their performance. However, the present meta-analytic review does not support the

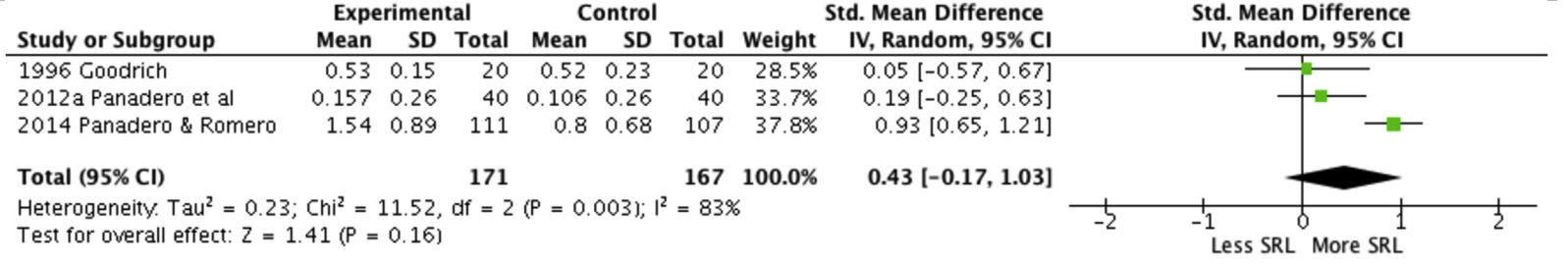

Fig. 6. Forest plot for the meta-analysis self-assessment effects on SRL measured qualitatively.

Table 6
Moderating variables effects on SRL measured qualitatively.

Quantitative moderators				
Moderating variable	k	$Q_R(1)$	B	95% CI
Gender	3	0.035 ($p = 0.852$)	-0.009	(-0.108–0.089)
Age	3	1.202 ($p = 0.273$)	0.136	(-0.107–0.378)
SA intervention (Total score)	3	2.210 ($p = 0.137$)	1.371	(-0.436–3.178)

Note: none of the categorical moderators was calculated because all the groups had the same score or one of the groups was k = 1.
SA = self-assessment.

assumption that rubrics would have a larger impact on Learning SRL as compared to other self-assessment interventions. It should be noted, however, that the effects of rubrics on Negative SRL could not be explored due to the limited number of empirical studies on this topic.

There are two possible explanation for the unexpected finding that rubrics do not have a larger impact on Learning SRL as compared to other self-assessment interventions. Firstly, the most specific explanation is provided by Panadero and Romero (2014), who discuss the negative effect of the rubric used in their study. Students had a limited time to complete the task (1 h) and the rubric might have increased students' feelings of working under pressure by making the expectations explicit (including the highest standard). Additionally, the task counted for the final grade, which was also made explicit and more evident by the rubric. According to the authors, rubrics might therefore need longer interventions in order to avoid inducing feelings of stress among the students under such pressing conditions. Secondly, research on the effects of using rubrics for formative purposes differs greatly regarding the quality of the interventions, as well as the quality of the research (Panadero & Jonsson, 2013), both of which may affect the outcome of this meta-analytic review.

The last moderator was implementation agent. However, no conclusions can be drawn, due to the low number studies in which the teachers implemented the self-assessment interventions. Therefore, the findings from previous meta-analyses, where the SRL interventions benefited from being implemented by the researcher (Dignath & Büttner, 2008; Hattie et al., 1996), should be taken into consideration for future self-assessment research.

4.2. The effect of self-assessment interventions on students' self-efficacy

As suggested by Schunk (1996), having students assess their progress makes it clear to them that they have become more competent, which in turn strengthens their self-efficacy. This assertion is supported by the results from our self-efficacy meta-analysis, where self-assessment interventions have a substantial impact on students' self-efficacy ($d = 0.73$). The magnitude of this effect size is comparable to the effect of feedback on students' achievement, as reported by Hattie (2009). The results also come from a larger set of effect sizes (k = 27), as compared to our SRL meta-analyses. The findings from the self-efficacy analyses therefore imply that self-assessment interventions can be a powerful strategy to increase students' confidence in their own capabilities. This, in turn, is important since self-efficacy is one of the major predictors of student performance (Honicke & Broadbent, 2016; Richardson et al., 2012; Robbins et al., 2004; Sitzmann et al., 2010).

The influence from self-assessment interventions on students' self-efficacy is two-folded. As self-assessment provides the student with information about the learning goals and how to progress towards them (Andrade, 2010; Panadero & Alonso-Tapia, 2013), training in self-assessment may make students feel more confident (Schunk, 1996). Furthermore, as task performance is likely to improve as a result of self-assessment training (Brown & Harris, 2013), confidence may increase even further.

4.2.1. Comparing self-efficacy and self-regulated learning effect sizes

An interesting observation is that there is a considerable difference between the size of the observed effects from the three SRL type of measurement and self-efficacy. One possible explanation for this difference is offered by Sitzmann et al. (2010). From their meta-analysis on different outcomes of self-assessment, they concluded that self-assessment may be best characterized as an affective evaluation outcome. If the results from the current study are organized from the largest combined effect to the lowest, they come in the following order: self-efficacy (0.73), Negative SRL (0.65), SRL measured qualitatively (0.43), and Learning SRL (0.24). By going one step further and analyzing the underlying constructs of, and data for, the different variables, the following pattern emerges: (a) self-efficacy is a personal belief mediated by cognitive, but also motivational and affective, processes (Bandura, 1997, chapter 4); (b) Negative SRL is largely a measure of how emotions (e.g. anxiety) influence SRL strategies; (c) of the three studies providing data for the SRL measured qualitatively variable, only one used cognitive measures (Goodrich, 1996, p. 49), while emotion regulation was included in one of the others (Panadero et al., 2012); and (d) data for Learning SRL is a diverse combination of multiple instruments (such as the MSLQ and the EMSR-Q) including cognitive, motivational, and emotional aspects. A tentative explanation for the observed pattern of combined effect sizes could therefore be, as suggested by Sitzmann et al. (2010), that affective measures are influenced more by self-assessment as compared to cognitively oriented measures.

Another possible explanation for the observed pattern among effect sizes, is that it might be an artefact produced by instrument design. In relation to self-efficacy, students are asked whether they feel capable of performing a particular

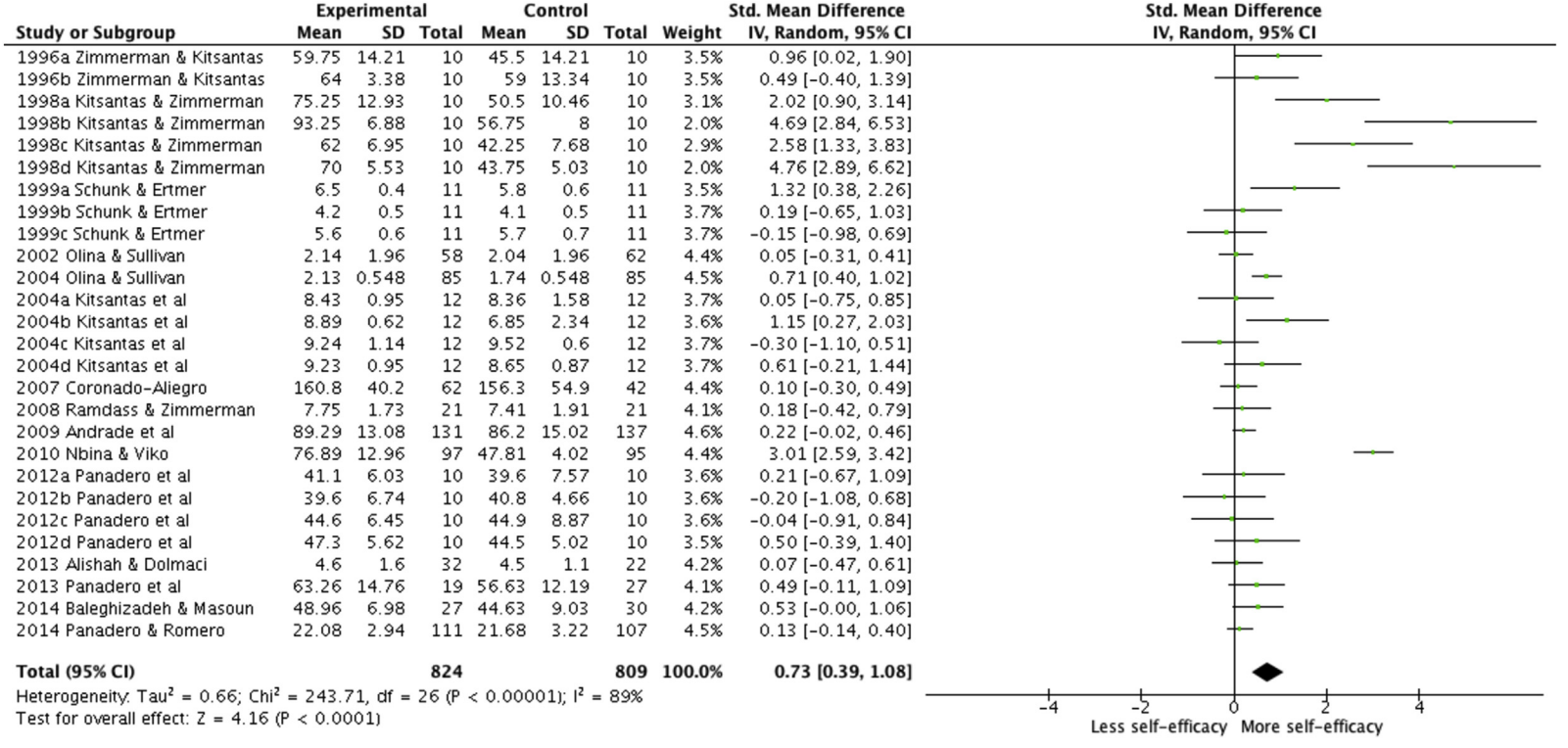

Fig. 7. Forest plot for the meta-analysis self-assessment effects on Self-efficacy.

Table 7
Moderating variables effects on self-efficacy.

Quantitative moderators						
Moderating variable	k	$Q_R(1)$	B			95% CI
Gender	24	5.271 ($p = 0.022$)	0.026			(0.004–0.049)
Age	25	0.562 ($p = 0.453$)	-0.028			(-0.099–0.045)
SA intervention (Total)	27	0.003 ($p = 0.935$)	0.019			(-0.604–0.641)
Categorical moderators						
Moderating variable		Q_B	Group	k	*d*	95% CI
Ed level[a]		3.839 ($p = 0.147$)	Primary	2	0.206	(-1.084–1.496)
			Secondary	17	1.106	(0.624–1.589)
			University	8	0.390	(-0.273–1.052)
Implementation agent		1.930 ($p = 0.165$)	Researcher	21	0.957	(0.525–1.389)
			Teacher	6	0.348	(-0.394–1.091)
SA intervention categories						
Monitoring		12.962 ($p = 0.000$)	Monitoring	12	1.456	(0.970–1.941)
			No monitoring	15	0.310	(-0.081–0.701)
Self-evaluation		1.818 ($p = 0.178$)	Self-evaluation	7	0.386	(-0.325–1.096)
			No self-evaluation	20	0.959	(0.523–1.396)
SA instrument		3.912 ($p = 0.048$)	SA instrument	7	0.197	(-0.503–0.896)
			No SA instrument	20	1.023	(0.596–1.451)
Feedback and revision		1.284 ($p = 0.257$)	F&R	10	0.538	(-0.059–1.135)
			No F&R	17	0.984	(0.496–1.471)

Note: Highlighted in grey moderating variables that will not be discussed in the paper because one of the groups is k = 2.
SA = self-assessment.
[a] The study with primary students was included because there were two other categories and k was bigger than 1 but caution needs to be taken when interpreting this results regarding primary education students studies.

task. Self-efficacy instruments typically make use of projective scales or items asking for personal opinions or perceptions that are future-oriented. The MSLQ, which is the most frequently used tool for measuring both SRL (Orwin 1983) and self-efficacy (Honicke & Broadbent, 2016), includes self-efficacy items such as: “I believe I will receive an excellent grade in this class” or “I'm certain I can master the skills being taught in this class”.[5] Items that measure (meta-)cognitive aspects of SRL, on the other hand, are usually more precise and oriented towards frequency of behavior, often in present or past tense. Going back to the MSLQ, examples of items are: “I make simple charts, diagrams, or tables to help me organize course material” or “I work hard to do well in this class even if I don't like what we are doing”. These differences between instruments for affective and cognitive measures could therefore create a differential effect, based on instrument design, where the level of uncertainty in future-oriented items could be higher (“How capable do you think you are to perform this in the future?”), as compared to more precise past-oriented questions (“What was your level of concentration in the task you just performed?”). In sum, this calls for future research to analyze the design of SRL and self-efficacy instruments, in order to explore whether this construct irrelevant variance exists and might contribute to the observed differences.

4.2.2. Moderating effects on self-efficacy

Two of the four moderators analyzed have a statistically significant impact on students' self-efficacy: gender and self-assessment intervention type. Next, we analyze all the moderating variables in detail. First, the findings suggest that girls' self-efficacy increases more as a result of self-assessment interventions, as compared to boys'. Although most studies included in this meta-analysis had a sample of both boys and girls, only a few explored gender differences. Consequently, the selected studies offer very little in terms of hypotheses about why boys and girls are affected differently. An exception is Andrade et al. (2009), who suggest that by using rubrics for self-assessment, girls may feel more successful since there are no risk of debilitating external feedback or social comparisons.

In Huang's meta-analysis (2013) a small difference favoring males' self-efficacy ($d = 0.08$) was found. This difference was explained by moderator analysis, demonstrating that content domain contributed significantly to the variation in effect sizes. While females displayed higher self-efficacy in relation to language arts, males' self-efficacy was higher in mathematics, computer, and social sciences. The influence of content domain has not been investigated in this meta-analysis, but as can be seen in our Table 1, there is an even distribution of subjects that might favor either gender.

Second, no significant effects were found in this meta-analytic review regarding age/educational level, and no previous meta-analyses have been found exploring this aspect in relation to self-efficacy. An important implication of these findings is that self-assessment interventions might have an equally powerful effect on students' self-efficacy across educational levels.

[5] There are two other MSLQ scales using this type of items that are also related to affective outcomes: Task value and Expectancy components.

Third, regarding self-assessment interventions, our findings do not support the idea that more comprehensive self-assessment interventions, or intervention using self-evaluation or feedback and revision components result in larger effects on students' self-efficacy. However, it was found that both monitoring and self-assessment instrument components have differential effects. Regarding monitoring, it was found that studies including monitoring reported an increase in self-efficacy. A viable hypothesis for this finding might lie in the research design of such monitoring studies as they tend to focus on performances that provide immediate feedback that is easy to interpret. For example, in Zimmerman and Kitsantas (1996) the performance investigated is dart throwing, which means that the participants can easily see their scores on the board immediately after releasing the dart. Similarly, in Ramdass and Zimmerman (2008) the students were taught how to check their answers in division tasks in mathematics, which means that they also received immediate feedback on their performance. In line with this hypothesis, Sitzmann et al. (2010) make a distinction between interpersonal and psychomotor tasks on the one hand and cognitive tasks on the other, because in the former case students can observe whether their actions were successful, but not in the latter. The students could therefore easily see whether their answers were correct or not, without making any interpretations or assessments according to quality criteria. This situation can be compared to other self-assessment interventions, involving more complex performance, such as classifying arthropods (Goodrich, 1996) or performing a landscape analysis (Panadero et al., 2012). In these latter cases, it is not immediately obvious for the students to what extent they have succeeded on the task, since any feedback (internal or external) must be grounded in an assessment of the performance according to criteria, which is more complex than monitoring performances that can be categorized as either right or wrong. The second self-assessment intervention that shown significant effects was the use of self-assessment instruments (e.g. rubrics, scripts). Participants using such instruments reported lower self-efficacy after the intervention than participants not using them. A viable hypothesis is that, as these instruments make the assessment criteria and standards explicit, students become aware of the complexity of high quality performance and, therefore, report lower self-efficacy. Future research should clarify whether students using a self-assessment instrument have a more realistic perception of their capacity, as compared to students who do not know the criteria or standards in such detail. Importantly, this finding contradicts the proposition that the use of rubrics improves students' self-efficacy (Panadero & Jonsson, 2013). The main explanations for these conflicting findings are that: a) in 2013, a lower number of studies exploring this effect was found, and b) in the present review, using meta-analytic methodology, the effect, as well as the direction of the effect, can be determined with greater accuracy.

Fourth, regarding the implementation agent, no significant effects were found and no previous meta-analyses have been found exploring this aspect.

4.3. Future lines of research

While the results of this meta-analytic review are important for the fields of self-assessment, SRL, and self-efficacy, the selected studies do not clarify the exact mechanisms for how self-assessment interventions influence students' SRL strategies and self-efficacy. Consequently, research is needed, which investigate the actual processes of self-assessment and SRL more closely. As was shown in the introduction, self-assessment may influence all phases of the SRL cycle, including the planning phase. Still, since there are different forms of self-assessment, as well as self-assessment interventions with different components, future research needs to be more specific with regards to self-assessment typologies, in order to advance our understanding of how self-assessment interventions impact on students' SRL (Panadero, Jonsson et al., 2016).

A second line of future research relates to the discussion emerging from the study by Sitzmann et al. (2010), which could be supported by the findings from this review. It needs to be clarified whether self-assessment interventions benefit affective outcomes more than cognitive ones, while at the same time ruling out the hypothesis that this finding is an artifact created by the measurement instrument design. Also, even if self-assessment would primarily influence affective outcomes, such as self-efficacy, it is still likely to have an effect on cognitive aspects.

A third suggestion for future research, which relates to the discussion above, comes from the fact that there has not yet been any sufficiently fine-grained measurements separating the different SRL components. In the light of the findings from this review, that the effect of self-assessment is different for SRL and self-efficacy variables, it might be hypothesized that self-assessment does not have the same level of impact for all SRL processes.

An obvious recommendation, based on the findings, is that upcoming self-assessment research exploring the effects on students' SRL strategies should move away from using only self-reported data from questionnaires (see also Samuelstuen & Bråten, 2007). Instead, it would be advisable to include more direct SRL measurements (e.g. thinking-aloud protocols, observations). A broader base of different data, including a combination of intervention and measurement (Panadero, Klug et al., 2016), is likely to enhance our understanding of the SRL processes.

Finally, more research is needed to explain the possible causes for the larger impact of self-assessment on girls' self-efficacy.

4.4. Educational implications

The major educational implication from this meta-analytic review is that, as pointed out by Sadler (1989), Boud (1995), and Black and Wiliam (1998), self-assessment is a necessity for productive learning. Self-assessment has been shown to have a positive influence on academic performance (Brown & Harris, 2013) and according to the results in our review, self-

assessment also has a positive influence on students' SRL strategies and self-efficacy. We therefore re-iterate that self-assessment needs to be a major instructional aim in every classroom, but not only for the sake of improving performance. Since self-assessment interventions has an impact on students' SRL and self-efficacy, self-assessment needs to be implemented for the sake of students' empowerment and self-sustained learning.

Unfortunately, it is not possible from the results of this meta-analytic review to identify which self-assessment components, or combinations of components, are the most effective in terms of fostering SRL strategies or self-efficacy. Even though findings suggest that monitoring is an effective self-assessment component for increasing students' self-efficacy, most studies in the sample using monitoring rely on performances that provide immediate feedback that is easy to interpret. It is therefore not necessarily the process of monitoring that is effective in fostering increasing students' self-efficacy, but the immediate and easily interpretable feedback that affects students' perceptions of their capability.

5. Conclusion

The findings from this meta-analytic review suggest that self-assessment interventions have a positive influence on students' SRL strategies and self-efficacy. Importantly, the magnitude of these positive effects differ between SRL measurement types, implying that the role of SRL measurement needs to be carefully considered in upcoming research. The present review also shows that some moderating variables, such as gender and certain self-assessment components, influence the effects on students' self-efficacy. Specifically, self-assessment interventions have a larger impact on girls' self-efficacy, as compared to boys', and self-monitoring has a larger impact on students' self-efficacy as compared to the other self-assessment components investigated in this study. All in all, this review supports the idea that self-assessment interventions may have a positive influence on students' SRL and self-efficacy, although we still need to increase our corpus of knowledge about the actual mechanisms for these effects.

Acknowledgements

First author funded by the Ministerio de Economía y Competitividad via Spanish Ramón y Cajal programme (Referencia RYC-2013-13469).

References*

* Alishah, A. R., & Dolmaci, M. (2013). The interface between self-efficacy concerning the self-assessment on students studying English as a foreign language. In *Paper presented at the akdeniz language studies conference 2012* http://www.sciencedirect.com/science/article/pii/S1877042813001341.

Alonso-Tapia, J., Panadero, E., & Ruiz, M. A. (2014). Development and validity of the Emotion and Motivation Self-regulation Questionnaire (EMSR-Q). *Spanish Journal of Psychology, 17*(e55), 1–15. http://dx.doi.org/10.1017/sjp.2014.41.

Andrade, H. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. J. Andrade, & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 90–105). New York: Routledge.

Andrade, H., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practices, 27*(2), 3–13.

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice, 48*(1), 12–19. http://dx.doi.org/10.1080/00405840802577544.

* Andrade, H., Wang, X. L., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *Journal of Educational Research, 102*(4), 287–301.

* Baleghizadeh, S., & Masoun, A. (2014). The effect of self-assessment on EFL learners' self-efficacy. *TESL Canada Journal, 31*(1), 42.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. http://dx.doi.org/10.1037/0033-295X.84.2.191.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman and Company.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*(1), 7–73. http://dx.doi.org/10.1080/0969595980050102.

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology, 54*(2), 199–231. http://dx.doi.org/10.1111/j.1464-0597.2005.00205.x.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effects and random-effects models for meta-analysis. *Research Synthesis Methods, 1*, 97–111. http://dx.doi.org/10.1002/jrsm.12.

Boud, D. (1995). *Enhancing learning through self-assessment*. New York: RoutledgeFalmer.

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher-education: A critical analysis of findings. *Higher Education, 18*(5), 529–549. http://dx.doi.org/10.1007/BF00138746.

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education, 27*, 1–13. http://dx.doi.org/10.1016/j.iheduc.2015.04.007.

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367–393). Thousand Oaks, CA: SAGE.

Bussey, K. (2011). The influence of gender on students' self-regulated learning and performance. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 426–441). New York: Routledge.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281. http://dx.doi.org/10.3102/00346543065003245.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). New York: Academic Press.

* Coronado-Aliegro, J. (2007). *The effect of self-assessment in the self-efficacy of students studying Spanish as a foreign language*. PhD. University of Pittsburgh.

* (Marked with an * the studies included in the meta-analyses).

* DiGiacomo, G. (2014). *Enhancing self-monitoring and self-reflection through a self-regulatory skills intervention embedded in a middle school mathematics curriculum*. PhD. City University of New York.
Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3*, 231–264. http://dx.doi.org/10.1007/s11409-008-9029-x.
Dignath, C., Büttner, G., & Langfeldt, H. (2008). How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes. *Educational Research Review, 3*(2), 101–129. http://dx.doi.org/10.1016/j.edurev.2008.02.003.
Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*(3), 69–106. http://dx.doi.org/10.1111/j.1529-1006.2004.00018.x.
Eva, K. W., & Regehr, G. (2005). Self-assessment in the health professions: A reformulation and research agenda. *Academic Medicine, 80*(10), S46–S54.
* Goodrich, H. W. (1996). *Student self-assessment: At the intersection of metacognition and authentic assessment*. PhD. US: Harvard University.
Goodrich Andrade, H., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *The Journal of Educational Research, 97*(1), 21–34. http://dx.doi.org/10.1080/00220670309596625.
Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66*(2), 99–136. http://dx.doi.org/10.3102/00346543066002099.
Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
Honicke, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review, 17*, 63–84. http://dx.doi.org/10.1016/j.edurev.2015.11.002.
Huang, C. J. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education, 28*(1), 1–35. http://dx.doi.org/10.1007/s10212-011-0097-y.
Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment & Evaluation In Higher Education*, 1–13. http://dx.doi.org/10.1080/02602938.2013.875117.
* Kahrizi, P., Farahian, M., & Rajabi, S. (2014). The impact of self-assessment on self-regulation and critical thinking of EFL learners. *Modern Journal of Language Teaching Methods, 4*(1), 353.
* Kitsantas, A., Reiser, R. A., & Doster, J. (2004). Developing self-regulated learners: Goal setting, self-evaluation, and organizational signals during acquisition of procedural skills. *The Journal of Experimental Education, 72*(4), 269–287.
* Kitsantas, A., & Zimmerman, B. J. (1998). Self-regulation of motoric learning: A strategic cycle view. *Journal of Applied Sport Psychology, 10*(2), 220–239. http://dx.doi.org/10.1080/10413209808406390.
Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. http://dx.doi.org/10.1037/0022-3514.77.6.1121.
Kuhl, J. (2000). A functional-design approach to motivation and self-regulation. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 111–169). San Diego, CA: Academic Press.
Light, R. J., & Pillemer, D. B. (1984). *Summing up. The science of reviewing research*. Cambridge, MA: Harvard University Press.
Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
* Mahlberg, J. (2015). Formative self-assessment college classes improves self-regulation and retention in first/second year community college students. *Community College Journal of Research and Practice*, 1–12. http://dx.doi.org/10.1080/10668926.2014.922134.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*(4), 264–269. http://dx.doi.org/10.1371/journal.pmed1000097.
* Nbina, J. B., & Viko, B. (2010). Effect of instruction in metacognitive self-assessment strategy on chemistry students' self-efficacy and achievement. *Academia Arena, 2*(1), 1–10.
Nicol, D., & McFarlane-Dick, D. (2006). Formative assessment and self-regulated learning, a model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218. http://dx.doi.org/10.1080/03075070600572090.
* Olina, Z., & Sullivan, H. J. (2002). Effects of classroom evaluation strategies on student achievement and attitudes. *Educational Technology, Research and Development, 50*(3), 61–75. http://dx.doi.org/10.1007/BF02505025.
* Olina, Z., & Sullivan, H. J. (2004). Student self-evaluation, teacher evaluation, and learner performance. *Educational Technology Research and Development, 52*(3), 5–22. http://dx.doi.org/10.1007/BF02504672.
Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, 157–159.
Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543–578. http://dx.doi.org/10.3102/00346543066004543.
Pajares, F. (2008). Motivational role of self-efficacy beliefs in self-regulated learning. In D. H. Schunk, & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning. Theory, research and applications* (pp. 111–168). New York: Lawrence Erlbaum Associates.
Panadero, E. (2011). *Instructional help for self-assessment and self-regulation: Evaluation of the efficacy of self-assessment scripts vs. rubrics*. Ph.D. Universidad Autónoma de Madrid (Spain)
Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*(422). http://dx.doi.org/10.3389/fpsyg.2017.00422.
Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology, 11*(2), 551–576. http://dx.doi.org/10.14204/ejrep.30.12200.
* Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences, 22*(6), 806–813. http://dx.doi.org/10.1016/j.lindif.2012.04.007.
* Panadero, E., Alonso-Tapia, J., & Reche, E. (2013). Rubrics vs. self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers. *Studies In Educational Evaluation, 39*(3), 125–132. http://dx.doi.org/10.1016/j.stueduc.2013.04.001.
Panadero, E., Brown, G. T. L., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*. http://dx.doi.org/10.1007/s10648-015-9350-2.
Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review, 9*(0), 129–144. http://dx.doi.org/10.1016/j.edurev.2013.01.002.
Panadero, E., Jonsson, A., & Strijbos, J. W. (2016). Scaffolding self-regulated learning through self-assessment and peer assessment: Guidelines for classroom implementation. In D. Laveault, & L. Allal (Eds.), *Assessment for Learning: Meeting the challenge of implementation* (pp. 311–326). New York: Springer.
Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: When measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research, 60*(6), 723–735. http://dx.doi.org/10.1080/00313831.2015.1066436.
* Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice, 21*(2), 133–148. http://dx.doi.org/10.1080/0969594X.2013.877872.
Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36*(2), 89–101. http://dx.doi.org/10.1207/S15326985EP3602_4.
Perry, N. E., & Rahim, A. (2011). Studying self-regulated learning in classrooms. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 122–136). New York: Routledge.
Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 452–502). San Diego, CA: Academic Press.
Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*.
* Ramdass, D., & Zimmerman, B. J. (2008). Effects of self-correction strategy training on middle school students' self-efficacy, self-evaluation, and mathematics division learning. *Journal of Advanced Academics, 20*(1), 18–41. http://dx.doi.org/10.4219/jaa-2008-869.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353–387. http://dx.doi.org/10.1037/a0026838.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*(2), 261–288. http://dx.doi.org/10.1037/0033-2909.130.2.261.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638.

Roth, A., Ogrin, S., & Schmitz, B. (2016). Assessing self-regulated learning in higher education: A systematic literature review of self-report instruments. *Educational Assessment, Evaluation and Accountability, 28*(3), 225–250. http://dx.doi.org/10.1007/s11092-015-9229-2.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144. http://dx.doi.org/10.1007/bf00117714.

Samuelstuen, M. S., & Bråten, I. (2007). Examining the validity of self-reports on scales measuring students' strategic processing. *British Journal of Educational Psychology, 77*(2), 351–378. http://dx.doi.org/10.1348/000709906x106147.

Schunk, D. H. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Research Journal, 33*(2), 359–382. http://dx.doi.org/10.2307/1163289.

* Schunk, D. H., & Ertmer, P. A. (1999). Self-regulatory processes during computer skill acquisition: Goal and self-evaluative influences. *Journal of Educational Psychology, 91*(2), 251–260. http://dx.doi.org/10.1037/0022-0663.91.2.251.

Schunk, D. H., & Usher, E. L. (2011). Assessing self-efficacy for self-regulated learning. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 282–297). New York: Routledge.

Sitzmann, T., Ely, K., Brown, K. G., & Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education, 9*(2), 169–191.

Topping, K. J. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (Vol 1, pp. 55–87). Springer Netherlands.

van Dinther, M., Dochy, F., & Segers, M. (2011). Factors affecting students' self-efficacy in higher education. *Educational Research Review, 6*(2), 95–108. http://dx.doi.org/10.1016/j.edurev.2010.10.003.

Veenman, M. (2011). Alternative assessment of strategy use with self-report instruments: A discussion. *Metacognition and Learning, 6*(2), 205–211. http://dx.doi.org/10.1007/s11409-011-9080-x.

Whitebread, D., Bingham, S., Grau, V., Pino Pasternak, D., & Sangster, C. (2007). Development of metacognition and self-regulated learning in young children: Role of collaborative and peer-assisted learning. *Journal of Cognitive Education and Psychology, 6*(3), 433–455. http://dx.doi.org/10.1891/194589507787382043.

Whitley, B. E. (1997). Gender differences in computer-related attitudes and behavior: A meta-analysis. *Computers in Human Behavior, 13*(1), 1–22. http://dx.doi.org/10.1016/S0747-5632(96)00026-X.

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). Orlando, FL: Academic Press.

Wright, C. R., & Houck, J. W. (1995). Gender differences among self-assessments, teacher ratings, grades, and aptitude test scores for a sample of students attending rural secondary schools. *Educational and Psychological Measurement, 55*(5), 743–752. http://dx.doi.org/10.1177/0013164495055005005.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–40). San Diego, California: Academic Press.

* Zimmerman, B. J., & Kitsantas, A. (1996). Self-regulated learning of a motoric skill: The role of goal setting and self-monitoring. *Journal of Applied Sport Psychology, 8*(1), 60–75. http://dx.doi.org/10.1080/10413209608406308.

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299–315). New York: Routledge.