Check for updates

A Critical Review of Research on Student Self-Assessment

Heidi L. Andrade*

Educational Psychology and Methodology, University at Albany, Albany, NY, United States

This article is a review of research on student self-assessment conducted largely between 2013 and 2018. The purpose of the review is to provide an updated overview of theory and research. The treatment of theory involves articulating a refined definition and operationalization of self-assessment. The review of 76 empirical studies offers a critical perspective on what has been investigated, including the relationship between self-assessment and achievement, consistency of self-assessment and others' assessments, student perceptions of self-assessment, and the association between self-assessment and self-regulated learning. An argument is made for less research on consistency and summative self-assessment, and more on the cognitive and affective mechanisms of formative self-assessment.

Keywords: self-assessment, self-evaluation, self-grading, formative assessment, classroom assessment, self-regulated learning (SRL)

This review of research on student self-assessment expands on a review published as a chapter in the *Cambridge Handbook of Instructional Feedback* (Andrade, 2018, reprinted with permission). The timespan for the original review was January 2013 to October 2016. A lot of research has been done on the subject since then, including at least two meta-analyses; hence this expanded review, in which I provide an updated overview of theory and research. The treatment of theory presented here involves articulating a refined definition and operationalization of self-assessment through a lens of feedback. My review of the growing body of empirical research offers a critical perspective, in the interest of provoking new investigations into neglected areas.

DEFINING AND OPERATIONALIZING STUDENT SELF-ASSESSMENT

Without exception, reviews of self-assessment (Sargeant, 2008; Brown and Harris, 2013; Panadero et al., 2016a) call for clearer definitions: What is self-assessment, and what is not? This question is surprisingly difficult to answer, as the term *self-assessment* has been used to describe a diverse range of activities, such as assigning a happy or sad face to a story just told, estimating the number of correct answers on a math test, graphing scores for dart throwing, indicating understanding (or the lack thereof) of a science concept, using a rubric to identify strengths and weaknesses in one's persuasive essay, writing reflective journal entries, and so on. Each of those activities involves some kind of assessment of one's own functioning, but they are so different that distinctions among types of self-assessment are needed. I will draw those distinctions in terms of the purposes of self-assessment which, in turn, determine its features: a classic form-fits-function analysis.

WHAT IS SELF-ASSESSMENT?

Brown and Harris (2013) defined self-assessment in the K-16 context as a "descriptive and evaluative act carried out by the student concerning his or her own work and academic abilities" (p. 368). Panadero et al. (2016a) defined it 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" (p. 804). Referring to physicians, Epstein et al. (2008) defined "concurrent self-assessment" as "ongoing moment-to-moment self-monitoring" (p. 5). Self-monitoring "refers to the ability to notice our own actions, curiosity to examine the effects of those actions, and willingness to use those observations to improve behavior and thinking in the future" (p. 5). Taken together, these definitions include self-assessment of one's *abilities*, *processes*, and *products*—everything but the kitchen sink. This very broad conception might seem unwieldy, but it works because each object of assessment—competence, process, and product—is subject to the influence of feedback from oneself.

What is missing from each of these definitions, however, is the purpose of the act of self-assessment. Their authors might rightly point out that the purpose is implied, but a formal definition requires us to make it plain: Why do we ask students to self-assess? I have long held that self-assessment is feedback (Andrade, 2010), and that the purpose of feedback is to inform adjustments to processes and products that deepen learning and enhance performance; hence the purpose of self-assessment is to generate feedback that promotes learning and improvements in performance. This learning-oriented purpose of self-assessment implies that it should be formative: if there is no opportunity for adjustment and correction, self-assessment is almost pointless.

WHY SELF-ASSESS?

Clarity about the purpose of self-assessment allows us to interpret what otherwise appear to be discordant findings from research, which has produced mixed results in terms of both the accuracy of students' self-assessments and their influence on learning and/or performance. I believe the source of the discord can be traced to the different ways in which self-assessment is carried out, such as whether it is summative and formative. This issue will be taken up again in the review of current research that follows this overview. For now, consider a study of the accuracy and validity of summative self-assessment in teacher education conducted by Tejeiro et al. (2012), which showed that students' self-assigned marks tended to be higher than marks given by professors. All 122 students in the study assigned themselves a grade at the end of their course, but half of the students were told that their self-assigned grade would count toward 5% of their final grade. In both groups, students' self-assessments were higher than grades given by professors, especially for students with "poorer results" (p. 791) and those for whom self-assessment counted toward the final grade. In the group that was told their self-assessments would count toward their final grade, no relationship was found between the professor's and the students' assessments. Tejeiro et al. concluded

that, although students' and professor's assessments tend to be highly similar when self-assessment did not count toward final grades, overestimations increased dramatically when students' self-assessments did count. Interviews of students who self-assigned highly discrepant grades revealed (as you might guess) that they were motivated by the desire to obtain the highest possible grades.

Studies like Tejeiro et al's. (2012) are interesting in terms of the information they provide about the relationship between consistency and honesty, but the purpose of the self-assessment, beyond addressing interesting research questions, is unclear. There is no feedback purpose. This is also true for another example of a study of summative self-assessment of competence, during which elementary-school children took the Test of Narrative Language and then were asked to self-evaluate "how you did in making up stories today" by pointing to one of five pictures, from a "very happy face" (rating of five) to a "very sad face" (rating of one) (Kaderavek et al., 2004. p. 37). The usual results were reported: Older children and good narrators were more accurate than younger children and poor narrators, and males tended to more frequently overestimate their ability.

Typical of clinical studies of accuracy in self-evaluation, this study rests on a definition and operationalization of self-assessment with no value in terms of instructional feedback. If those children were asked to rate their stories and then revise or, better yet, if they assessed their stories according to clear, developmentally appropriate criteria before revising, the valence of their self-assessments in terms of instructional feedback would skyrocket. I speculate that their accuracy would too. In contrast, studies of formative self-assessment suggest that when the act of self-assessing is given a learning-oriented purpose, students' self-assessments are relatively consistent with those of external evaluators, including professors (Lopez and Kossack, 2007; Barney et al., 2012; Leach, 2012), teachers (Bol et al., 2012; Chang et al., 2012, 2013), researchers (Panadero and Romero, 2014; Fitzpatrick and Schulz, 2016), and expert medical assessors (Hawkins et al., 2012).

My commitment to keeping self-assessment formative is firm. However, Gavin Brown (personal communication, April 2011) reminded me that summative self-assessment exists and we cannot ignore it; any definition of self-assessment must acknowledge and distinguish between formative and summative forms of it. Thus, the taxonomy in **Table 1**, which depicts self-assessment as serving formative and/or summative purposes, and focuses on competence, processes, and/or products.

Fortunately, a formative view of self-assessment seems to be taking hold in various educational contexts. For instance, Sargeant (2008) noted that all seven authors in a special issue of the *Journal of Continuing Education in the Health Professions* "conceptualize self-assessment within a formative, educational perspective, and see it as an activity that draws upon both external and internal data, standards, and resources to inform and make decisions about one's performance" (p. 1). Sargeant also stresses the point that self-assessment should be guided by evaluative criteria: "Multiple external sources can and should inform self-assessment, perhaps most important among them performance standards" (p. 1). Now we are talking about the *how*

TABLE 1 | A taxonomy of self-assessment.

Competence		Processes		Products	
		Standards		Standards	
		Yes	No	Yes	No
Formative	Task-specific self-efficacy ratings	Judgments of progress toward specific targets	• Traffic lights • Comprehension checks • Self-monitoring; metacognition • Reflective journal writing	• Rubric- or checklist-referenced self-assessment • Self-testing	Open-ended critique of one's own work or understanding
Summative	Post-task judgments of ability based on performance		Post-task judgments of effectiveness of procedures	• Self-grading	• Self-grading

of self-assessment, which demands an operationalization of self-assessment practice. Let us examine each object of self-assessment (competence, processes, and/or products) with an eye for what is assessed and why.

WHAT IS SELF-ASSESSED?

Monitoring and self-assessing *processes* are practically synonymous with self-regulated learning (SRL), or at least central components of it such as goal-setting and monitoring, or metacognition. Research on SRL has clearly shown that self-generated feedback on one's approach to learning is associated with academic gains (Zimmerman and Schunk, 2011). Self-assessment of the *products,* such as papers and presentations, are the easiest to defend as feedback, especially when those self-assessments are grounded in explicit, relevant, evaluative criteria and followed by opportunities to relearn and/or revise (Andrade, 2010).

Including the self-assessment of *competence* in this definition is a little trickier. I hesitated to include it because of the risk of sneaking in global assessments of one's overall ability, self-esteem, and self-concept ("I'm good enough, I'm smart enough, and doggone it, people like me," Franken, 1992), which do not seem relevant to a discussion of feedback in the context of learning. Research on global self-assessment, or self-perception, is popular in the medical education literature, but even there, scholars have begun to question its usefulness in terms of influencing learning and professional growth (e.g., see Sargeant et al., 2008). Eva and Regehr (2008) seem to agree in the following passage, which states the case in a way that makes it worthy of a long quotation:

> Self-assessment is often (implicitly or otherwise) conceptualized as a personal, unguided reflection on performance for the purposes of generating an individually derived summary of one's own level of knowledge, skill, and understanding in a particular area. For example, this conceptualization would appear to be the only reasonable basis for studies that fit into what Colliver et al. (2005) has described as the "guess your grade" model of self-assessment research, the results of which form the core foundation for the recurring conclusion that self-assessment is generally poor. This unguided, internally generated construction of self-assessment stands in stark contrast to the model put forward

by Boud (1999), who argued that the phrase self-assessment should not imply an isolated or individualistic activity; it should commonly involve peers, teachers, and other sources of information. The conceptualization of self-assessment as enunciated in Boud's description would appear to involve a process by which one takes personal responsibility for looking outward, explicitly seeking feedback, and information from external sources, then using these externally generated sources of assessment data to direct performance improvements. In this construction, self-assessment is more of a pedagogical strategy than an ability to judge for oneself; it is a habit that one needs to acquire and enact rather than an ability that one needs to master (p. 15).

As in the K-16 context, self-assessment is coming to be seen as having value as much or more so in terms of pedagogy as in assessment (Silver et al., 2008; Brown and Harris, 2014). In the end, however, I decided that self-assessing one's competence to successfully learn a particular concept or complete a particular task (which sounds a lot like self-efficacy—more on that later) might be useful feedback because it can inform decisions about how to proceed, such as the amount of time to invest in learning how to play the flute, or whether or not to seek help learning the steps of the jitterbug. An important caveat, however, is that self-assessments of competence are only useful if students have opportunities to do something about their perceived low competence—that is, it serves the purpose of formative feedback for the learner.

HOW TO SELF-ASSESS?

Panadero et al. (2016a) summarized five very different taxonomies of self-assessment and called for the development of a comprehensive typology that considers, among other things, its purpose, the presence or absence of criteria, and the method. In response, I propose the taxonomy depicted in **Table 1**, which focuses on the *what* (competence, process, or product), the *why* (formative or summative), and the *how* (methods, including whether or not they include standards, e.g., criteria) of self-assessment. The collections of examples of methods in the table is inexhaustive.

I put the methods in **Table 1** where I think they belong, but many of them could be placed in more than one cell. Take *self-efficacy*, for instance, which is essentially a self-assessment of one's competence to successfully undertake a particular task (Bandura, 1997). Summative judgments of self-efficacy are certainly possible but they seem like a silly thing to do—what is the point, from a learning perspective? Formative self-efficacy judgments, on the other hand, can inform next steps in learning and skill building. There is reason to believe that monitoring and making adjustments to one's self-efficacy (e.g., by setting goals or attributing success to effort) can be productive (Zimmerman, 2000), so I placed self-efficacy in the formative row.

It is important to emphasize that self-efficacy is task-specific, more or less (Bandura, 1997). This taxonomy does not include general, holistic evaluations of one's abilities, for example, "I am good at math." Global assessment of competence does not provide the leverage, in terms of feedback, that is provided by task-specific assessments of competence, that is, self-efficacy. Eva and Regehr (2008) provided an illustrative example: "We suspect most people are prompted to open a dictionary as a result of encountering a word for which they are uncertain of the meaning rather than out of a broader assessment that their vocabulary could be improved" (p. 16). The exclusion of global evaluations of oneself resonates with research that clearly shows that feedback that focuses on aspects of a task (e.g., "I did not solve most of the algebra problems") is more effective than feedback that focuses on the self (e.g., "I am bad at math") (Kluger and DeNisi, 1996; Dweck, 2006; Hattie and Timperley, 2007). Hence, global self-evaluations of ability or competence do not appear in **Table 1**.

Another approach to student self-assessment that could be placed in more than one cell is *traffic lights*. The term *traffic lights* refers to asking students to use green, yellow, or red objects (or thumbs up, sideways, or down—anything will do) to indicate whether they think they have good, partial, or little understanding (Black et al., 2003). It would be appropriate for traffic lights to appear in multiple places in **Table 1**, depending on how they are used. Traffic lights seem to be most effective at supporting students' reflections on how well they understand a concept or have mastered a skill, which is line with their creators' original intent, so they are categorized as formative self-assessments of one's learning—which sounds like metacognition.

In fact, several of the methods included in **Table 1** come from research on metacognition, including *self-monitoring*, such as checking one's reading comprehension, and *self-testing*, e.g., checking one's performance on test items. These last two methods have been excluded from some taxonomies of self-assessment (e.g., Boud and Brew, 1995) because they do not engage students in explicitly considering relevant standards or criteria. However, new conceptions of self-assessment are grounded in theories of the self- and co-regulation of learning (Andrade and Brookhart, 2016), which includes self-monitoring of learning processes with and without explicit standards.

However, my research favors self-assessment with regard to standards (Andrade and Boulay, 2003; Andrade and Du, 2007; Andrade et al., 2008, 2009, 2010), as does related research

by Panadero and his colleagues (see below). I have involved students in self-assessment of stories, essays, or mathematical word problems according to *rubrics* or *checklists* with criteria. For example, two studies investigated the relationship between elementary or middle school students' scores on a written assignment and a process that involved them in reading a model paper, co-creating criteria, self-assessing first drafts with a rubric, and revising (Andrade et al., 2008, 2010). The self-assessment was highly scaffolded: students were asked to underline key phrases in the rubric with colored pencils (e.g., underline "clearly states an opinion" in blue), then underline or circle in their drafts the evidence of having met the standard articulated by the phrase (e.g., his or her opinion) with the same blue pencil. If students found they had not met the standard, they were asked to write themselves a reminder to make improvements when they wrote their final drafts. This process was followed for each criterion on the rubric. There were main effects on scores for every self-assessed criterion on the rubric, suggesting that guided self-assessment according to the co-created criteria helped students produce more effective writing.

Panadero and his colleagues have also done quasi-experimental and experimental research on standards-referenced self-assessment, using rubrics or lists of assessment criteria that are presented in the form of questions (Panadero et al., 2012, 2013, 2014; Panadero and Romero, 2014). Panadero calls the list of assessment criteria a *script* because his work is grounded in research on scaffolding (e.g., Kollar et al., 2006): I call it a *checklist* because that is the term used in classroom assessment contexts. Either way, the list provides standards for the task. Here is a script for a written summary that Panadero et al. (2014) used with college students in a psychology class:

- Does my summary transmit the main idea from the text? Is it at the beginning of my summary?
- Are the important ideas also in my summary?
- Have I selected the main ideas from the text to make them explicit in my summary?
- Have I thought about my purpose for the summary? What is my goal?

Taken together, the results of the studies cited above suggest that students who engaged in self-assessment using scripts or rubrics were more self-regulated, as measured by self-report questionnaires and/or think aloud protocols, than were students in the comparison or control groups. Effect sizes were very small to moderate ($\eta^2 = 0.06$–0.42), and statistically significant. Most interesting, perhaps, is one study (Panadero and Romero, 2014) that demonstrated an association between rubric-referenced self-assessment activities and all three phases of SRL; forethought, performance, and reflection.

There are surely many other methods of self-assessment to include in **Table 1**, as well as interesting conversations to be had about which method goes where and why. In the meantime, I offer the taxonomy in **Table 1** as a way to define and operationalize self-assessment in instructional contexts and as a framework for the following overview of current research on the subject.
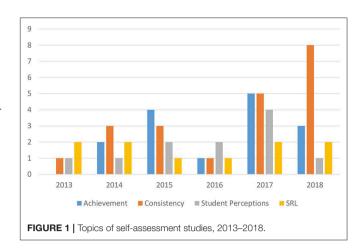
AN OVERVIEW OF CURRENT RESEARCH ON SELF-ASSESSMENT

Several recent reviews of self-assessment are available (Brown and Harris, 2013; Brown et al., 2015; Panadero et al., 2017), so I will not summarize the entire body of research here. Instead, I chose to take a birds-eye view of the field, with goal of reporting on what has been sufficiently researched and what remains to be done. I used the references lists from reviews, as well as other relevant sources, as a starting point. In order to update the list of sources, I directed two new searches[1], the first of the ERIC database, and the second of both ERIC and PsychINFO. Both searches included two search terms, "self-assessment" OR "self-evaluation." Advanced search options had four delimiters: (1) peer-reviewed, (2) January, 2013–October, 2016 and then October 2016–March 2019, (3) English, and (4) full-text. Because the focus was on K-20 educational contexts, sources were excluded if they were about early childhood education or professional development.

The first search yielded 347 hits; the second 1,163. Research that was unrelated to instructional feedback was excluded, such as studies limited to self-estimates of performance before or after taking a test, guesses about whether a test item was answered correctly, and estimates of how many tasks could be completed in a certain amount of time. Although some of the excluded studies might be thought of as useful investigations of self-monitoring, as a group they seemed too unrelated to theories of self-generated feedback to be appropriate for this review. Seventy-six studies were selected for inclusion in **Table S1** (Supplementary Material), which also contains a few studies published before 2013 that were not included in key reviews, as well as studies solicited directly from authors.

The **Table S1** in the Supplementary Material contains a complete list of studies included in this review, organized by the focus or topic of the study, as well as brief descriptions of each. The "type" column **Table S1** (Supplementary Material) indicates whether the study focused on formative or summative self-assessment. This distinction was often difficult to make due to a lack of information. For example, Memis and Seven (2015) frame their study in terms of formative assessment, and note that the purpose of the self-evaluation done by the sixth grade students is to "help students improve their [science] reports" (p. 39), but they do not indicate how the self-assessments were done, nor whether students were given time to revise their reports based on their judgments or supported in making revisions. A sentence or two of explanation about the process of self-assessment in the procedures sections of published studies would be most useful.

Figure 1 graphically represents the number of studies in the four most common topic categories found in the table— achievement, consistency, student perceptions, and SRL. The figure reveals that research on self-assessment is on the rise, with consistency the most popular topic. Of the 76 studies in the table in the appendix, 44 were inquiries into the consistency of students' self-assessments with other judgments (e.g., a

FIGURE 1 | Topics of self-assessment studies, 2013–2018.

test score or teacher's grade). Twenty-five studies investigated the relationship between self-assessment and achievement. Fifteen explored students' perceptions of self-assessment. Twelve studies focused on the association between self-assessment and self-regulated learning. One examined self-efficacy, and two qualitative studies documented the mental processes involved in self-assessment. The sum ($n = 99$) of the list of research topics is more than 76 because several studies had multiple foci. In the remainder of this review I examine each topic in turn.

Consistency

Table S1 (Supplementary Material) reveals that much of the recent research on self-assessment has investigated the accuracy or, more accurately, consistency, of students' self-assessments. The term *consistency* is more appropriate in the classroom context because the quality of students' self-assessments is often determined by comparing them with their teachers' assessments and then generating correlations. Given the evidence of the unreliability of teachers' grades (Falchikov, 2005), the assumption that teachers' assessments are accurate might not be well-founded (Leach, 2012; Brown et al., 2015). Ratings of student work done by researchers are also suspect, unless evidence of the validity and reliability of the inferences made about student work by researchers is available. Consequently, much of the research on classroom-based self-assessment should use the term *consistency*, which refers to the degree of alignment between students' and expert raters' evaluations, avoiding the purer, more rigorous term *accuracy* unless it is fitting.

In their review, Brown and Harris (2013) reported that correlations between student self-ratings and other measures tended to be weakly to strongly positive, ranging from $r \approx 0.20$ to 0.80, with few studies reporting correlations >0.60. But their review included results from studies of any self-appraisal of school work, including summative self-rating/grading, predictions about the correctness of answers on test items, and formative, criteria-based self-assessments, a combination of methods that makes the correlations they reported difficult to interpret. Qualitatively different forms of self-assessment, especially summative and formative types,

cannot be lumped together without obfuscating important aspects of self-assessment as feedback.

Given my concern about combining studies of summative and formative assessment, you might anticipate a call for research on consistency that distinguishes between the two. I will make no such call for three reasons. One is that we have enough research on the subject, including the 22 studies in **Table S1** (Supplementary Material) that were published after Brown and Harris's review (2013). Drawing only on studies included in **Table S1** (Supplementary Material), we can say with confidence that summative self-assessment tends to be inconsistent with external judgements (Baxter and Norman, 2011; De Grez et al., 2012; Admiraal et al., 2015), with males tending to overrate and females to underrate (Nowell and Alston, 2007; Marks et al., 2018). There are exceptions (Alaoutinen, 2012; Lopez-Pastor et al., 2012) as well as mixed results, with students being consistent regarding some aspects of their learning but not others (Blanch-Hartigan, 2011; Harding and Hbaci, 2015; Nguyen and Foster, 2018). We can also say that older, more academically competent learners tend to be more consistent (Hacker et al., 2000; Lew et al., 2010; Alaoutinen, 2012; Guillory and Blankson, 2017; Butler, 2018; Nagel and Lindsey, 2018). There is evidence that consistency can be improved through experience (Lopez and Kossack, 2007; Yilmaz, 2017; Nagel and Lindsey, 2018), the use of guidelines (Bol et al., 2012), feedback (Thawabieh, 2017), and standards (Baars et al., 2014), perhaps in the form of rubrics (Panadero and Romero, 2014). Modeling and feedback also help (Labuhn et al., 2010; Miller and Geraci, 2011; Hawkins et al., 2012; Kostons et al., 2012).

An outcome typical of research on the consistency of summative self-assessment can be found in row 59, which summarizes the study by Tejeiro et al. (2012) discussed earlier: Students' self-assessments were higher than marks given by professors, especially for students with poorer results, and no relationship was found between the professors' and the students' assessments in the group in which self-assessment counted toward the final mark. Students are not stupid: if they know that they can influence their final grade, and that their judgment is summative rather than intended to inform revision and improvement, they will be motivated to inflate their self-evaluation. I do not believe we need more research to demonstrate that phenomenon.

The second reason I am not calling for additional research on consistency is a lot of it seems somewhat irrelevant. This might be because the interest in accuracy is rooted in clinical research on calibration, which has very different aims. Calibration accuracy is the "magnitude of consent between learners' true and self-evaluated task performance. Accurately calibrated learners' task performance equals their self-evaluated task performance" (Wollenschläger et al., 2016). Calibration research often asks study participants to predict or postdict the correctness of their responses to test items. I caution about generalizing from clinical experiments to authentic classroom contexts because the dismal picture of our human potential to self-judge was painted by calibration researchers *before* study participants were effectively taught how to predict with accuracy, or provided with the tools they needed to be accurate, or motivated to do so.

Calibration researchers know that, of course, and have conducted intervention studies that attempt to improve accuracy, with some success (e.g., Bol et al., 2012). Studies of formative self-assessment also suggest that consistency increases when it is taught and supported in many of the ways any other skill must be taught and supported (Lopez and Kossack, 2007; Labuhn et al., 2010; Chang et al., 2012, 2013; Hawkins et al., 2012; Panadero and Romero, 2014; Lin-Siegler et al., 2015; Fitzpatrick and Schulz, 2016).

Even clinical psychological studies that go beyond calibration to examine the associations between monitoring accuracy and subsequent study behaviors do not transfer well to classroom assessment research. After repeatedly encountering claims that, for example, low self-assessment accuracy leads to poor task-selection accuracy and "suboptimal learning outcomes" (Raaijmakers et al., 2019, p. 1), I dug into the cited studies and discovered two limitations. The first is that the tasks in which study participants engage are quite inauthentic. A typical task involves studying "word pairs (e.g., railroad—mother), followed by a delayed judgment of learning (JOL) in which the students predicted the chances of remembering the pair... After making a JOL, the entire pair was presented for restudy for 4 s [*sic*], and after all pairs had been restudied, a criterion test of paired-associate recall occurred" (Dunlosky and Rawson, 2012, p. 272). Although memory for word pairs might be important in some classroom contexts, it is not safe to assume that results from studies like that one can predict students' behaviors after criterion-referenced self-assessment of their comprehension of complex texts, lengthy compositions, or solutions to multi-step mathematical problems.

The second limitation of studies like the typical one described above is more serious: Participants in research like that are not permitted to regulate their own studying, which is experimentally manipulated by a computer program. This came as a surprise, since many of the claims were about students' poor study choices but they were rarely allowed to make actual choices. For example, Dunlosky and Rawson (2012) permitted participants to "use monitoring to effectively control learning" by programming the computer so that "a participant would need to have judged his or her recall of a definition entirely correct on three different trials, and once they judged it entirely correct on the third trial, that particular key term definition was dropped [by the computer program] from further practice" (p. 272). The authors note that this study design is an improvement on designs that did not require all participants to use the same regulation algorithm, but it does not reflect the kinds of decisions that learners make in class or while doing homework. In fact, a large body of research shows that students *can* make wise choices when they self-pace the study of to-be-learned materials and then allocate study time to each item (Bjork et al., 2013, p. 425):

In a typical experiment, the students first study all the items at an experimenter-paced rate (e.g., study 60 paired associates for 3 s each), which familiarizes the students with the items; after this familiarity phase, the students then either choose which items they want to restudy (e.g., all items are presented in an array, and the students select which ones to restudy) and/or pace their restudy of each item. Several dependent measures have been

widely used, such as how long each item is studied, whether an item is selected for restudy, and in what order items are selected for restudy. The literature on these aspects of self-regulated study is massive (for a comprehensive overview, see both Dunlosky and Ariel, 2011 and Son and Metcalfe, 2000), but the evidence is largely consistent with a few basic conclusions. First, if students have a chance to practice retrieval prior to restudying items, they almost exclusively choose to restudy unrecalled items and drop the previously recalled items from restudy (Metcalfe and Kornell, 2005). Second, when pacing their study of individual items that have been selected for restudy, students typically spend more time studying items that are more, rather than less, difficult to learn. Such a strategy is consistent with a discrepancy-reduction model of self-paced study (which states that people continue to study an item until they reach mastery), although some key revisions to this model are needed to account for all the data. For instance, students may not continue to study until they reach some static criterion of mastery, but instead, they may continue to study until they perceive that they are no longer making progress.

I propose that this research, which suggests that students' unscaffolded, unmeasured, informal self-assessments tend to lead to appropriate task selection, is better aligned with research on classroom-based self-assessment. Nonetheless, even this comparison is inadequate because the study participants were not taught to compare their performance to the criteria for mastery, as is often done in classroom-based self-assessment.

The third and final reason I do not believe we need additional research on consistency is that I think it is a distraction from the true purposes of self-assessment. Many if not most of the articles about the accuracy of self-assessment are grounded in the assumption that accuracy is necessary for self-assessment to be useful, particularly in terms of subsequent studying and revision behaviors. Although it seems obvious that accurate evaluations of their performance positively influence students' study strategy selection, which should produce improvements in achievement, I have not seen relevant research that tests those conjectures. Some claim that inaccurate estimates of learning lead to the selection of inappropriate learning tasks (Kostons et al., 2012) but they cite research that does not support their claim. For example, Kostons et al. cite studies that focus on the effectiveness of SRL interventions but do not address the accuracy of participants' estimates of learning, nor the relationship of those estimates to the selection of next steps. Other studies produce findings that support my skepticism. Take, for instance, two relevant studies of calibration. One suggested that performance and judgments of performance had little influence on subsequent test preparation behavior (Hacker et al., 2000), and the other showed that study participants followed their predictions of performance to the same degree, regardless of monitoring accuracy (van Loon et al., 2014).

Eva and Regehr (2008) believe that:
Research questions that take the form of "How well do various practitioners self-assess?" "How can we improve self-assessment?" or "How can we measure self-assessment skill?" should be considered defunct and removed from the research agenda [because] there have been hundreds of studies into these

questions and the answers are "Poorly," "You can't," and "Don't bother" (p. 18).

I almost agree. A study that could change my mind about the importance of accuracy of self-assessment would be an investigation that goes beyond attempting to improve accuracy just for the sake of accuracy by instead examining the relearning/revision behaviors of accurate and inaccurate self-assessors: Do students whose self-assessments match the valid and reliable judgments of expert raters (hence my use of the term *accuracy*) make better decisions about what they need to do to deepen their learning and improve their work? Here, I admit, is a call for research related to consistency: I would love to see a high-quality investigation of the relationship between accuracy in formative self-assessment, and students' subsequent study and revision behaviors, and their learning. For example, a study that closely examines the revisions to writing made by accurate and inaccurate self-assessors, and the resulting outcomes in terms of the quality of their writing, would be most welcome.

Table S1 (Supplementary Material) indicates that by 2018 researchers began publishing studies that more directly address the hypothesized link between self-assessment and subsequent learning behaviors, as well as important questions about the processes learners engage in while self-assessing (Yan and Brown, 2017). One, a study by Nugteren et al. (2018 row 19 in **Table S1** (Supplementary Material)), asked "How do inaccurate [summative] self-assessments influence task selections?" (p. 368) and employed a clever exploratory research design. The results suggested that most of the 15 students in their sample over-estimated their performance and made inaccurate learning-task selections. Nugteren et al. recommended helping students make more accurate self-assessments, but I think the more interesting finding is related to why students made task selections that were too difficult or too easy, given their prior performance: They based most task selections on interest in the content of particular items (not the overarching content to be learned), and infrequently considered task difficulty and support level. For instance, while working on the genetics tasks, students reported selecting tasks because they were fun or interesting, not because they addressed self-identified weaknesses in their understanding of genetics. Nugteren et al. proposed that students would benefit from instruction on task selection. I second that proposal: Rather than directing our efforts on accuracy in the service of improving subsequent task selection, let us simply teach students to use the information at hand to select next best steps, among other things.

Butler (2018, row 76 in **Table S1** (Supplementary Material)) has conducted at least two studies of learners' processes of responding to self-assessment items and how they arrived at their judgments. Comparing generic, decontextualized items to task-specific, contextualized items (which she calls *after-task items*), she drew two unsurprising conclusions: the task-specific items "generally showed higher correlations with task performance," and older students "appeared to be more conservative in their judgment compared with their younger counterparts" (p. 249). The contribution of the study is the detailed information it provides about *how* students generated their judgments. For

example, Butler's qualitative data analyses revealed that when asked to self-assess in terms of vague or non-specific items, the children often "contextualized the descriptions based on their own experiences, goals, and expectations," (p. 257) focused on the task at hand, and situated items in the specific task context. Perhaps as a result, the correlation between after-task self-assessment and task performance was generally higher than for generic self-assessment.

Butler (2018) notes that her study enriches our empirical understanding of the processes by which children respond to self-assessment. This is a very promising direction for the field. Similar studies of processing during formative self-assessment of a variety of task types in a classroom context would likely produce significant advances in our understanding of how and why self-assessment influences learning and performance.

Student Perceptions

Fifteen of the studies listed in **Table S1** (Supplementary Material) focused on students' perceptions of self-assessment. The studies of children suggest that they tend to have unsophisticated understandings of its purposes (Harris and Brown, 2013; Bourke, 2016) that might lead to shallow implementation of related processes. In contrast, results from the studies conducted in higher education settings suggested that college and university students understood the function of self-assessment (Ratminingsih et al., 2018) and generally found it to be useful for guiding evaluation and revision (Micán and Medina, 2017), understanding how to take responsibility for learning (Lopez and Kossack, 2007; Bourke, 2014; Ndoye, 2017), prompting them to think more critically and deeply (van Helvoort, 2012; Siow, 2015), applying newfound skills (Murakami et al., 2012), and fostering self-regulated learning by guiding them to set goals, plan, self-monitor and reflect (Wang, 2017).

Not surprisingly, positive perceptions of self-assessment were typically developed by students who actively engaged the formative type by, for example, developing their own criteria for an effective self-assessment response (Bourke, 2014), or using a rubric or checklist to guide their assessments and then revising their work (Huang and Gui, 2015; Wang, 2017). Earlier research suggested that children's attitudes toward self-assessment can become negative if it is summative (Ross et al., 1998). However, even summative self-assessment was reported by adult learners to be useful in helping them become more critical of their own and others' writing throughout the course and in subsequent courses (van Helvoort, 2012).

Achievement

Twenty-five of the studies in **Table S1** (Supplementary Material) investigated the relation between self-assessment and achievement, including two meta-analyses. Twenty of the 25 clearly employed the formative type. Without exception, those 20 studies, plus the two meta-analyses (Graham et al., 2015; Sanchez et al., 2017) demonstrated a positive association between self-assessment and learning. The meta-analysis conducted by Graham and his colleagues, which included 10 studies, yielded an average weighted effect size of 0.62 on writing quality. The Sanchez et al. meta-analysis revealed that, although 12 of the 44

effect sizes were negative, on average, "students who engaged in self-grading performed better ($g = 0.34$) on subsequent tests than did students who did not" (p. 1,049).

All but two of the non-meta-analytic studies of achievement in **Table S1** (Supplementary Material) were quasi-experimental or experimental, providing relatively rigorous evidence that their treatment groups outperformed their comparison or control groups in terms of everything from writing to dart-throwing, map-making, speaking English, and exams in a wide variety of disciplines. One experiment on summative self-assessment (Miller and Geraci, 2011), in contrast, resulted in no improvements in exam scores, while the other one did (Raaijmakers et al., 2017).

It would be easy to overgeneralize and claim that the question about the effect of self-assessment on learning has been answered, but there are unanswered questions about the key components of effective self-assessment, especially social-emotional components related to power and trust (Andrade and Brown, 2016). The trends are pretty clear, however: it appears that formative forms of self-assessment can promote knowledge and skill development. This is not surprising, given that it involves many of the processes known to support learning, including practice, feedback, revision, and especially the intellectually demanding work of making complex, criteria-referenced judgments (Panadero et al., 2014). Boud (1995a,b) predicted this trend when he noted that many self-assessment processes undermine learning by rushing to judgment, thereby failing to engage students with the standards or criteria for their work.

Self-Regulated Learning

The association between self-assessment and learning has also been explained in terms of self-regulation (Andrade, 2010; Panadero and Alonso-Tapia, 2013; Andrade and Brookhart, 2016, 2019; Panadero et al., 2016b). Self-regulated learning (SRL) occurs when learners set goals and then monitor and manage their thoughts, feelings, and actions to reach those goals. SRL is moderately to highly correlated with achievement (Zimmerman and Schunk, 2011). Research suggests that formative assessment is a potential influence on SRL (Nicol and Macfarlane-Dick, 2006). The 12 studies in **Table S1** (Supplementary Material) that focus on SRL demonstrate the recent increase in interest in the relationship between self-assessment and SRL.

Conceptual and practical overlaps between the two fields are abundant. In fact, Brown and Harris (2014) recommend that student self-assessment no longer be treated as an assessment, but as an essential competence for self-regulation. Butler and Winne (1995) introduced the role of self-generated feedback in self-regulation years ago:

> [For] all self-regulated activities, feedback is an inherent catalyst. As learners monitor their engagement with tasks, internal feedback is generated by the monitoring process. That feedback describes the nature of outcomes and the qualities of the cognitive processes that led to those states (p. 245).

The outcomes and processes referred to by Butler and Winne are many of the same *products* and *processes* I referred to earlier in the definition of self-assessment and in **Table 1**.

In general, research and practice related to self-assessment has tended to focus on judging the products of student learning, while scholarship on self-regulated learning encompasses both processes and products. The very practical focus of much of the research on self-assessment means it might be playing catch-up, in terms of theory development, with the SRL literature, which is grounded in experimental paradigms from cognitive psychology (de Bruin and van Gog, 2012), while self-assessment research is ahead in terms of implementation (E. Panadero, personal communication, October 21, 2016). One major exception is the work done on Self-regulated Strategy Development (Glaser and Brunstein, 2007; Harris et al., 2008), which has successfully integrated SRL research with classroom practices, including self-assessment, to teach writing to students with special needs.

> Nicol and Macfarlane-Dick (2006) have been explicit about the potential for self-assessment practices to support self-regulated learning:
> To develop systematically the learner's capacity for self-regulation, teachers need to create more structured opportunities for self-monitoring and the judging of progression to goals. Self-assessment tasks are an effective way of achieving this, as are activities that encourage reflection on learning progress (p. 207).

The studies of SRL in **Table S1** (Supplementary Material) provide encouraging findings regarding the potential role of self-assessment in promoting achievement, self-regulated learning in general, and metacognition and study strategies related to task selection in particular. The studies also represent a solution to the "methodological and theoretical challenges involved in bringing metacognitive research to the real world, using meaningful learning materials" (Koriat, 2012, p. 296).

FUTURE DIRECTIONS FOR RESEARCH

I agree with (Yan and Brown, 2017) statement that "from a pedagogical perspective, the benefits of self-assessment may come from active engagement in the learning process, rather than by being "veridical" or coinciding with reality, because students' reflection and metacognitive monitoring lead to improved learning" (p. 1,248). Future research should focus less on accuracy/consistency/veridicality, and more on the precise mechanisms of self-assessment (Butler, 2018).

An important aspect of research on self-assessment that is not explicitly represented in **Table S1** (Supplementary Material) is practice, or pedagogy: Under what conditions does self-assessment work best, and how are those conditions influenced by context? Fortunately, the studies listed in the table, as well as others (see especially Andrade and Valtcheva, 2009; Nielsen, 2014; Panadero et al., 2016a), point toward an answer. But we still have questions about how best to scaffold effective formative self-assessment. One area of inquiry is about the characteristics of the task being assessed, and the standards or criteria used by learners during self-assessment.

Influence of Types of Tasks and Standards or Criteria

Type of task or competency assessed seems to matter (e.g., Dolosic, 2018, Nguyen and Foster, 2018), as do the criteria (Yilmaz, 2017), but we do not yet have a comprehensive understanding of how or why. There is some evidence that it is important that the criteria used to self-assess are concrete, task-specific (Butler, 2018), and graduated. For example, Fastre et al. (2010) revealed an association between self-assessment according to task-specific criteria and task performance: In a quasi-experimental study of 39 novice vocational education students studying stoma care, they compared concrete, task-specific criteria ("performance-based criteria") such as "Introduces herself to the patient" and "Consults the care file for details concerning the stoma" to vaguer, "competence-based criteria" such as "Shows interest, listens actively, shows empathy to the patient" and "Is discrete with sensitive topics." The performance-based criteria group outperformed the competence-based group on tests of task performance, presumably because "performance-based criteria make it easier to distinguish levels of performance, enabling a step-by-step process of performance improvement" (p. 530).

This finding echoes the results of a study of self-regulated learning by Kitsantas and Zimmerman (2006), who argued that "fine-grained standards can have two key benefits: They can enable learners to be more sensitive to small changes in skill and make more appropriate adaptations in learning strategies" (p. 203). In their study, 70 college students were taught how to throw darts at a target. The purpose of the study was to examine the role of graphing of self-recorded outcomes and self-evaluative standards in learning a motor skill. Students who were provided with graduated self-evaluative standards surpassed "those who were provided with absolute standards or no standards (control) in both motor skill and in motivational beliefs (i.e., self-efficacy, attributions, and self-satisfaction)" (p. 201). Kitsantas and Zimmerman hypothesized that setting high absolute standards would limit a learner's sensitivity to small improvements in functioning. This hypothesis was supported by the finding that students who set absolute standards reported significantly less awareness of learning progress (and hit the bull's-eye less often) than students who set graduated standards. "The correlation between the self-evaluation and dart-throwing outcomes measures was extraordinarily high ($r = 0.94$)" (p. 210). Classroom-based research on specific, graduated self-assessment criteria would be informative.

Cognitive and Affective Mechanisms of Self-Assessment

There are many additional questions about pedagogy, such as the hoped-for investigation mentioned above of the relationship between accuracy in formative self-assessment, students' subsequent study behaviors, and their learning. There is also a need for research on how to help teachers give students a central role in their learning by creating space for self-assessment (e.g., see Hawe and Parr, 2014), and the complex power dynamics

involved in doing so (Tan, 2004, 2009; Taras, 2008; Leach, 2012). However, there is an even more pressing need for investigations into the internal mechanisms experienced by students engaged in assessing their own learning. Angela Lui and I call this *the next black box* (Lui, 2017).

Black and Wiliam (1998) used the term *black box* to emphasize the fact that what happened in most classrooms was largely unknown: all we knew was that some inputs (e.g., teachers, resources, standards, and requirements) were fed into the box, and that certain outputs (e.g., more knowledgeable and competent students, acceptable levels of achievement) would follow. But what, they asked, is happening inside, and what new inputs will produce better outputs? Black and Wiliam's review spawned a great deal of research on formative assessment, some but not all of which suggests a positive relationship with academic achievement (Bennett, 2011; Kingston and Nash, 2011). To better understand why and how the use of formative assessment in general and self-assessment in particular is associated with improvements in academic achievement in some instances but not others, we need research that looks into the next black box: the cognitive and affective mechanisms of students who are engaged in assessment processes (Lui, 2017).

The role of internal mechanisms has been discussed in theory but not yet fully tested. Crooks (1988) argued that the impact of assessment is influenced by students' interpretation of the tasks and results, and Butler and Winne (1995) theorized that both cognitive and affective processes play a role in determining how feedback is internalized and used to self-regulate learning. Other theoretical frameworks about the internal processes of receiving and responding to feedback have been developed (e.g., Nicol and Macfarlane-Dick, 2006; Draper, 2009; Andrade, 2013; Lipnevich et al., 2016). Yet, Shute (2008) noted in her review of the literature on formative feedback that "despite the plethora of research on the topic, the specific mechanisms relating feedback to learning are still mostly murky, with very few (if any) general conclusions" (p. 156). This area is ripe for research.

CONCLUSION

Self-assessment is the act of monitoring one's processes and products in order to make adjustments that deepen learning and enhance performance. Although it can be summative, the evidence presented in this review strongly suggests that self-assessment is most beneficial, in terms of both achievement and self-regulated learning, when it is used formatively and supported by training.

What is not yet clear is why and how self-assessment works. Those of you who like to investigate phenomena that are maddeningly difficult to measure will rejoice to hear that the cognitive and affective mechanisms of self-assessment are the next black box. Studies of the ways in which learners think and feel, the interactions between their thoughts and feelings and their context, and the implications for pedagogy will make major contributions to our field.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2019.00087/full#supplementary-material

REFERENCES

Admiraal, W., Huisman, B., and Pilli, O. (2015). Assessment in massive open online courses. *Electron. J. e-Learning*, 13, 207–216.

Alaoutinen, S. (2012). Evaluating the effect of learning style and student background on self-assessment accuracy. *Comput. Sci. Educ.* 22, 175–198. doi: 10.1080/08993408.2012.692924

Al-Rawahi, N. M., and Al-Balushi, S. M. (2015). The effect of reflective science journal writing on students' self-regulated learning strategies. *Int. J. Environ. Sci. Educ.* 10, 367–379. doi: 10.12973/ijese.2015.250a

Andrade, H. (2010). "Students as the definitive source of formative assessment: academic self-assessment and the self-regulation of learning," in *Handbook of Formative Assessment,* eds H. Andrade and G. Cizek (New York, NY: Routledge), 90–105.

Andrade, H. (2013). "Classroom assessment in the context of learning theory and research," in *Sage Handbook of Research on Classroom Assessment,* ed J. H. McMillan (New York, NY: Sage), 17–34. doi: 10.4135/9781452218649.n2

Andrade, H. (2018). "Feedback in the context of self-assessment," in *Cambridge Handbook of Instructional Feedback,* eds A. Lipnevich and J. Smith (Cambridge: Cambridge University Press), 376–408.

Andrade, H., and Boulay, B. (2003). The role of rubric-referenced self-assessment in learning to write. *J. Educ. Res.* 97, 21–34. doi: 10.1080/00220670309596625

Andrade, H., and Brookhart, S. (2019). Classroom assessment as the co-regulation of learning. *Assessm. Educ. Principles Policy Pract.* doi: 10.1080/0969594X.2019.1571992

Andrade, H., and Brookhart, S. M. (2016). "The role of classroom assessment in supporting self-regulated learning," in *Assessment for Learning: Meeting the Challenge of Implementation*, eds D. Laveault and L. Allal (Heidelberg: Springer), 293–309. doi: 10.1007/978-3-319-39211-0_17

Andrade, H., and Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assess. Evalu. High. Educ.* 32, 159–181. doi: 10.1080/02602930600801928

Andrade, H., Du, Y., and Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assess. Educ.* 17, 199–214. doi: 10.1080/09695941003696172

Andrade, H., Du, Y., and 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. *Educ. Meas.* 27, 3–13. doi: 10.1111/j.1745-3992.2008.00118.x

Andrade, H., and Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Pract.* 48, 12–19. doi: 10.1080/00405840802577544

Andrade, H., Wang, X., Du, Y., and Akawi, R. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *J. Educ. Res.* 102, 287–302. doi: 10.3200/JOER.102.4.287-302

Andrade, H. L., and Brown, G. T. L. (2016). "Student self-assessment in the classroom," in *Handbook of Human and Social Conditions in Assessment,* eds G. T. L. Brown and L. R. Harris (New York, NY: Routledge), 319–334.

Baars, M., Vink, S., van Gog, T., de Bruin, A., and Paas, F. (2014). Effects of training self-assessment and using assessment standards on retrospective

and prospective monitoring of problem solving. *Learn. Instruc.* 33, 92–107. doi: 10.1016/j.learninstruc.2014.04.004

Balderas, I., and Cuamatzi, P. M. (2018). Self and peer correction to improve college students' writing skills. *Profile.* 20, 179–194. doi: 10.15446/profile.v20n2.67095

Bandura, A. (1997). *Self-efficacy: The Exercise of Control.* New York, NY: Freeman.

Barney, S., Khurum, M., Petersen, K., Unterkalmsteiner, M., and Jabangwe, R. (2012). Improving students with rubric-based self-assessment and oral feedback. *IEEE Transac. Educ.* 55, 319–325. doi: 10.1109/TE.2011.217 2981

Baxter, P., and Norman, G. (2011). Self-assessment or self deception? A lack of association between nursing students' self-assessment and performance. *J. Adv. Nurs.* 67, 2406–2413. doi: 10.1111/j.1365-2648.2011.05658.x

Bennett, R. E. (2011). Formative assessment: a critical review. *Assess. Educ.* 18, 5–25. doi: 10.1080/0969594X.2010.513678

Birjandi, P., and Hadidi Tamjid, N. (2012). The role of self-, peer and teacher assessment in promoting Iranian EFL learners' writing performance. *Assess. Evalu. High. Educ.* 37, 513–533. doi: 10.1080/02602938.2010. 549204

Bjork, R. A., Dunlosky, J., and Kornell, N. (2013). Self-regulated learning: beliefs, techniques, and illusions. *Annu. Rev. Psychol.* 64, 417–444. doi: 10.1146/annurev-psych-113011-143823

Black, P., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2003). *Assessment for Learning: Putting it into Practice.* Berkshire: Open University Press.

Black, P., and Wiliam, D. (1998). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan* 80, 139–144; 146–148.

Blanch-Hartigan, D. (2011). Medical students' self-assessment of performance: results from three meta-analyses. *Patient Educ. Counsel.* 84, 3–9. doi: 10.1016/j.pec.2010.06.037

Bol, L., Hacker, D. J., Walck, C. C., and Nunnery, J. A. (2012). The effects of individual or group guidelines on the calibration accuracy and achievement of high school biology students. *Contemp. Educ. Psychol.* 37, 280–287. doi: 10.1016/j.cedpsych.2012.02.004

Boud, D. (1995a). *Implementing Student Self-Assessment, 2nd Edn.* Australian Capital Territory: Higher Education Research and Development Society of Australasia.

Boud, D. (1995b). *Enhancing Learning Through Self-Assessment.* London: Kogan Page.

Boud, D. (1999). Avoiding the traps: Seeking good practice in the use of self-assessment and reflection in professional courses. *Soc. Work Educ.* 18, 121–132. doi: 10.1080/02615479911220131

Boud, D., and Brew, A. (1995). Developing a typology for learner self-assessment practices. *Res. Dev. High. Educ.* 18, 130–135.

Bourke, R. (2014). Self-assessment in professional programmes within tertiary institutions. *Teach. High. Educ.* 19, 908–918. doi: 10.1080/13562517.2014.934353

Bourke, R. (2016). Liberating the learner through self-assessment. *Cambridge J. Educ.* 46, 97–111. doi: 10.1080/0305764X.2015.1015963

Brown, G., Andrade, H., and Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. *Assess. Educ.* 22, 444–457. doi: 10.1080/0969594X.2014.996523

Brown, G. T., and Harris, L. R. (2013). "Student self-assessment," in *Sage Handbook of Research on Classroom Assessment,* ed J. H. McMillan (Los Angeles, CA: Sage), 367–393. doi: 10.4135/9781452218649.n21

Brown, G. T. L., and Harris, L. R. (2014). The future of self-assessment in classroom practice: reframing self-assessment as a core competency. *Frontline Learn. Res.* 3, 22–30. doi: 10.14786/flr.v2i1.24

Butler, D. L., and Winne, P. H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Rev. Educ. Res.* 65, 245–281. doi: 10.3102/00346543065003245

Butler, Y. G. (2018). "Young learners' processes and rationales for responding to self-assessment items: cases for generic can-do and five-point Likert-type formats," in *Useful Assessment and Evaluation in Language Education,* eds J. Davis et al. (Washington, DC: Georgetown University Press), 21–39. doi: 10.2307/j.ctvvngrq.5

Chang, C.-C., Liang, C., and Chen, Y.-H. (2013). Is learner self-assessment reliable and valid in a Web-based portfolio environment for high school students? *Comput. Educ.* 60, 325–334. doi: 10.1016/j.compedu.2012.05.012

Chang, C.-C., Tseng, K.-H., and Lou, S.-J. (2012). A comparative analysis of the consistency and difference among teacher-assessment, student self-assessment and peer-assessment in a Web-based portfolio assessment environment for high school students. *Comput. Educ.* 58, 303–320. doi: 10.1016/j.compedu.2011.08.005

Colliver, J., Verhulst, S, and Barrows, H. (2005). Self-assessment in medical practice: a further concern about the conventional research paradigm. *Teach. Learn. Med.* 17, 200–201. doi: 10.1207/s15328015tlm1703_1

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Rev. Educ. Res.* 58, 438–481. doi: 10.3102/00346543058004438

de Bruin, A. B. H., and van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learn. Instruct.* 22, 245–252. doi: 10.1016/j.learninstruc.2012.01.003

De Grez, L., Valcke, M., and Roozen, I. (2012). How effective are self- and peer assessment of oral presentation skills compared with teachers' assessments? *Active Learn. High. Educ.* 13, 129–142. doi: 10.1177/1469787412441284

Dolosic, H. (2018). An examination of self-assessment and interconnected facets of second language reading. *Read. Foreign Langu.* 30, 189–208.

Draper, S. W. (2009). What are learners actually regulating when given feedback? *Br. J. Educ. Technol.* 40, 306–315. doi: 10.1111/j.1467-8535.2008.00930.x

Dunlosky, J., and Ariel, R. (2011). "Self-regulated learning and the allocation of study time," in *Psychology of Learning and Motivation,* Vol. 54 ed B. Ross (Cambridge, MA: Academic Press), 103–140. doi: 10.1016/B978-0-12-385527-5.00004-8

Dunlosky, J., and Rawson, K. A. (2012). Overconfidence produces underachievement: inaccurate self evaluations undermine students' learning and retention. *Learn. Instr.* 22, 271–280. doi: 10.1016/j.learninstruc.2011.08.003

Dweck, C. (2006). *Mindset: The New Psychology of Success.* New York, NY: Random House.

Epstein, R. M., Siegel, D. J., and Silberman, J. (2008). Self-monitoring in clinical practice: a challenge for medical educators. *J. Contin. Educ. Health Prof.* 28, 5–13. doi: 10.1002/chp.149

Eva, K. W., and Regehr, G. (2008). "I'll never play professional football" and other fallacies of self-assessment. *J. Contin. Educ. Health Prof.* 28, 14–19. doi: 10.1002/chp.150

Falchikov, N. (2005). *Improving Assessment Through Student Involvement: Practical Solutions for Aiding Learning in Higher and Further Education.* London: Routledge Falmer.

Fastre, G. M. J., van der Klink, M. R., Sluijsmans, D., and van Merrienboer, J. J. G. (2012). Drawing students' attention to relevant assessment criteria: effects on self-assessment skills and performance. *J. Voc. Educ. Train.* 64, 185–198. doi: 10.1080/13636820.2011.630537

Fastre, G. M. J., van der Klink, M. R., and van Merrienboer, J. J. G. (2010). The effects of performance-based assessment criteria on student performance and self-assessment skills. *Adv. Health Sci. Educ.* 15, 517–532. doi: 10.1007/s10459-009-9215-x

Fitzpatrick, B., and Schulz, H. (2016). "Teaching young students to self-assess critically," *Paper presented at the Annual Meeting of the American Educational Research Association* (Washington, DC).

Franken, A. S. (1992). *I'm Good Enough, I'm Smart Enough, and Doggone it, People Like Me! Daily affirmations by Stuart Smalley.* New York, NY: Dell.

Glaser, C., and Brunstein, J. C. (2007). Improving fourth-grade students' composition skills: effects of strategy instruction and self-regulation procedures. *J. Educ. Psychol.* 99, 297–310. doi: 10.1037/0022-0663.99.2.297

Gonida, E. N., and Leondari, A. (2011). Patterns of motivation among adolescents with biased and accurate self-efficacy beliefs. *Int. J. Educ. Res.* 50, 209–220. doi: 10.1016/j.ijer.2011.08.002

Graham, S., Hebert, M., and Harris, K. R. (2015). Formative assessment and writing. *Elem. Sch. J.* 115, 523–547. doi: 10.1086/681947

Guillory, J. J., and Blankson, A. N. (2017). Using recently acquired knowledge to self-assess understanding in the classroom. *Sch. Teach. Learn. Psychol.* 3, 77–89. doi: 10.1037/stl0000079

Hacker, D. J., Bol, L., Horgan, D. D., and Rakow, E. A. (2000). Test prediction and performance in a classroom context. *J. Educ. Psychol.* 92, 160–170. doi: 10.1037/0022-0663.92.1.160

Harding, J. L., and Hbaci, I. (2015). Evaluating pre-service teachers math teaching experience from different perspectives. *Univ. J. Educ. Res.* 3, 382–389. doi: 10.13189/ujer.2015.030605

Harris, K. R., Graham, S., Mason, L. H., and Friedlander, B. (2008). *Powerful Writing Strategies for All Students.* Baltimore, MD: Brookes.

Harris, L. R., and Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: case studies into teachers' implementation. *Teach. Teach. Educ.* 36, 101–111. doi: 10.1016/j.tate.2013.07.008

Hattie, J., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77, 81–112. doi: 10.3102/003465430298487

Hawe, E., and Parr, J. (2014). Assessment for learning in the writing classroom: an incomplete realization. *Curr. J.* 25, 210–237. doi: 10.1080/09585176.2013.862172

Hawkins, S. C., Osborne, A., Schofield, S. J., Pournaras, D. J., and Chester, J. F. (2012). Improving the accuracy of self-assessment of practical clinical skills using video feedback: the importance of including benchmarks. *Med. Teach.* 34, 279–284. doi: 10.3109/0142159X.2012.658897

Huang, Y., and Gui, M. (2015). Articulating teachers' expectations afore: Impact of rubrics on Chinese EFL learners' self-assessment and speaking ability. *J. Educ. Train. Stud.* 3, 126–132. doi: 10.11114/jets.v3i3.753

Kaderavek, J. N., Gillam, R. B., Ukrainetz, T. A., Justice, L. M., and Eisenberg, S. N. (2004). School-age children's self-assessment of oral narrative production. *Commun. Disord. Q.* 26, 37–48. doi: 10.1177/15257401040260010401

Karnilowicz, W. (2012). A comparison of self-assessment and tutor assessment of undergraduate psychology students. *Soc. Behav. Person.* 40, 591–604. doi: 10.2224/sbp.2012.40.4.591

Kevereski, L. (2017). (Self) evaluation of knowledge in students' population in higher education in Macedonia. *Res. Pedag.* 7, 69–75. doi: 10.17810/2015.49

Kingston, N. M., and Nash, B. (2011). Formative assessment: a meta-analysis and a call for research. *Educ. Meas.* 30, 28–37. doi: 10.1111/j.1745-3992.2011.00220.x

Kitsantas, A., and Zimmerman, B. J. (2006). Enhancing self-regulation of practice: the influence of graphing and self-evaluative standards. *Metacogn. Learn.* 1, 201–212. doi: 10.1007/s11409-006-9000-7

Kluger, A. N., and DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol. Bull.* 119, 254–284. doi: 10.1037/0033-2909.119.2.254

Kollar, I., Fischer, F., and Hesse, F. (2006). Collaboration scripts: a conceptual analysis. *Educ. Psychol. Rev.* 18, 159–185. doi: 10.1007/s10648-006-9007-2

Kolovelonis, A., Goudas, M., and Dermitzaki, I. (2012). Students' performance calibration in a basketball dribbling task in elementary physical education. *Int. Electron. J. Elem. Educ.* 4, 507–517.

Koriat, A. (2012). The relationships between monitoring, regulation and performance. *Learn. Instru.* 22, 296–298. doi: 10.1016/j.learninstruc.2012.01.002

Kostons, D., van Gog, T., and Paas, F. (2012). Training self-assessment and task-selection skills: a cognitive approach to improving self-regulated learning. *Learn. Instruc.* 22, 121–132. doi: 10.1016/j.learninstruc.2011.08.004

Labuhn, A. S., Zimmerman, B. J., and Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: the influence of feedback and self-evaluative standards *Metacogn. Learn.* 5, 173–194. doi: 10.1007/s11409-010-9056-2

Leach, L. (2012). Optional self-assessment: some tensions and dilemmas. *Assess. Evalu. High. Educ.* 37, 137–147. doi: 10.1080/02602938.2010.515013

Lew, M. D. N., Alwis, W. A. M., and Schmidt, H. G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assess. Evalu. High. Educ.* 35, 135–156. doi: 10.1080/02602930802687737

Lin-Siegler, X., Shaenfield, D., and Elder, A. D. (2015). Contrasting case instruction can improve self-assessment of writing. *Educ. Technol. Res. Dev.* 63, 517–537. doi: 10.1007/s11423-015-9390-9

Lipnevich, A. A., Berg, D. A. G., and Smith, J. K. (2016). "Toward a model of student response to feedback," in *The Handbook of Human and Social Conditions in Assessment,* eds G. T. L. Brown and L. R. Harris (New York, NY: Routledge), 169–185.

Lopez, R., and Kossack, S. (2007). Effects of recurring use of self-assessment in university courses. *Int. J. Learn.* 14, 203–216. doi: 10.18848/1447-9494/CGP/v14i04/45277

Lopez-Pastor, V. M., Fernandez-Balboa, J.-M., Santos Pastor, M. L., and Aranda, A. F. (2012). Students' self-grading, professor's grading and negotiated final grading at three university programmes: analysis of reliability and grade

difference ranges and tendencies. *Assess. Evalu. High. Educ.* 37, 453–464. doi: 10.1080/02602938.2010.545868

Lui, A. (2017). *Validity of the responses to feedback survey: operationalizing and measuring students' cognitive and affective responses to teachers' feedback* (Doctoral dissertation). University at Albany—SUNY: *Albany NY.*

Marks, M. B., Haug, J. C., and Hu, H. (2018). Investigating undergraduate business internships: do supervisor and self-evaluations differ? *J. Educ. Bus.* 93, 33–45. doi: 10.1080/08832323.2017.1414025

Memis, E. K., and Seven, S. (2015). Effects of an SWH approach and self-evaluation on sixth grade students' learning and retention of an electricity unit. *Int. J. Prog. Educ.* 11, 32–49.

Metcalfe, J., and Kornell, N. (2005). A region of proximal learning model of study time allocation. *J. Mem. Langu.* 52, 463–477. doi: 10.1016/j.jml.2004.12.001

Meusen-Beekman, K. D., Joosten-ten Brinke, D., and Boshuizen, H. P. A. (2016). Effects of formative assessments to develop self-regulation among sixth grade students: results from a randomized controlled intervention. *Stud. Educ. Evalu.* 51, 126–136. doi: 10.1016/j.stueduc.2016.10.008

Micán, D. A., and Medina, C. L. (2017). Boosting vocabulary learning through self-assessment in an English language teaching context. *Assess. Evalu. High. Educ.* 42, 398–414. doi: 10.1080/02602938.2015.1118433

Miller, T. M., and Geraci, L. (2011). Training metacognition in the classroom: the influence of incentives and feedback on exam predictions. *Metacogn. Learn.* 6, 303–314. doi: 10.1007/s11409-011-9083-7

Murakami, C., Valvona, C., and Broudy, D. (2012). Turning apathy into activeness in oral communication classes: regular self- and peer-assessment in a TBLT programme. *System* 40, 407–420. doi: 10.1016/j.system.2012.07.003

Nagel, M., and Lindsey, B. (2018). The use of classroom clickers to support improved self-assessment in introductory chemistry. *J. College Sci. Teach.* 47, 72–79.

Ndoye, A. (2017). Peer/self-assessment and student learning. *Int. J. Teach. Learn. High. Educ.* 29, 255–269.

Nguyen, T., and Foster, K. A. (2018). Research note—multiple time point course evaluation and student learning outcomes in an MSW course. *J. Soc. Work Educ.* 54, 715–723. doi: 10.1080/10437797.2018.1474151

Nicol, D., and Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Stud. High. Educ.* 31, 199–218. doi: 10.1080/03075070600572090

Nielsen, K. (2014), Self-assessment methods in writing instruction: a conceptual framework, successful practices and essential strategies. *J. Res. Read.* 37, 1–16. doi: 10.1111/j.1467-9817.2012.01533.x

Nowell, C., and Alston, R. M. (2007). I thought I got an A! Overconfidence across the economics curriculum. *J. Econ. Educ.* 38, 131–142. doi: 10.3200/JECE.38.2.131-142

Nugteren, M. L., Jarodzka, H., Kester, L., and Van Merriënboer, J. J. G. (2018). Self-regulation of secondary school students: self-assessments are inaccurate and insufficiently used for learning-task selection. *Instruc. Sci.* 46, 357–381. doi: 10.1007/s11251-018-9448-2

Panadero, E., and 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. *Electron. J. Res. Educ. Psychol.* 11, 551–576. doi: 10.14204/ejrep.30.12200

Panadero, E., Alonso-Tapia, J., and Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learn. Individ. Differ.* 22, 806–813. doi: 10.1016/j.lindif.2012.04.007

Panadero, E., Alonso-Tapia, J., and Huertas, J. A. (2014). Rubrics vs. self-assessment scripts: effects on first year university students' self-regulation and performance. *J. Study Educ. Dev.* 3, 149–183. doi: 10.1080/02103702.2014.881655

Panadero, E., Alonso-Tapia, J., and Reche, E. (2013). Rubrics vs. self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers. *Stud. Educ. Evalu.* 39, 125–132. doi: 10.1016/j.stueduc.2013.04.001

Panadero, E., Brown, G. L., and Strijbos, J.-W. (2016a). The future of student self-assessment: a review of known unknowns and potential directions. *Educ. Psychol. Rev.* 28, 803–830. doi: 10.1007/s10648-015-9350-2

Panadero, E., Jonsson, A., and Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: four meta-analyses. *Educ. Res. Rev.* 22, 74–98. doi: 10.1016/j.edurev.2017.08.004

Panadero, E., Jonsson, A., and Strijbos, J. W. (2016b). "Scaffolding self-regulated learning through self-assessment and peer assessment: guidelines for classroom implementation," in *Assessment for Learning: Meeting the Challenge of Implementation,* eds D. Laveault and L. Allal (New York, NY: Springer), 311–326. doi: 10.1007/978-3-319-39211-0_18

Panadero, E., and Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assess. Educ.* 21, 133–148. doi: 10.1080/0969594X.2013.877872

Papanthymou, A., and Darra, M. (2018). Student self-assessment in higher education: The international experience and the Greek example. *World J. Educ.* 8, 130–146. doi: 10.5430/wje.v8n6p130

Punhagui, G. C., and de Souza, N. A. (2013). Self-regulation in the learning process: actions through self-assessment activities with Brazilian students. *Int. Educ. Stud.* 6, 47–62. doi: 10.5539/ies.v6n10p47

Raaijmakers, S. F., Baars, M., Paas, F., van Merriënboer, J. J. G., and van Gog, T. (2019). *Metacognition and Learning*, 1–22. doi: 10.1007/s11409-019-09189-5

Raaijmakers, S. F., Baars, M., Schapp, L., Paas, F., van Merrienboer, J., and van Gog, T. (2017). Training self-regulated learning with video modeling examples: do task-selection skills transfer? *Instr. Sci.* 46, 273–290. doi: 10.1007/s11251-017-9434-0

Ratminingsih, N. M., Marhaeni, A. A. I. N., and Vigayanti, L. P. D. (2018). Self-assessment: the effect on students' independence and writing competence. *Int. J. Instruc.* 11, 277–290. doi: 10.12973/iji.2018.11320a

Ross, J. A., Rolheiser, C., and Hogaboam-Gray, A. (1998). "Impact of self-evaluation training on mathematics achievement in a cooperative learning environment," *Paper presented at the annual meeting of the American Educational Research Association* (San Diego, CA).

Ross, J. A., and Starling, M. (2008). Self-assessment in a technology-supported environment: the case of grade 9 geography. *Assess. Educ.* 15, 183–199. doi: 10.1080/09695940802164218

Samaie, M., Nejad, A. M., and Qaracholloo, M. (2018). An inquiry into the efficiency of whatsapp for self- and peer-assessments of oral language proficiency. *Br. J. Educ. Technol.* 49, 111–126. doi: 10.1111/bjet.12519

Sanchez, C. E., Atkinson, K. M., Koenka, A. C., Moshontz, H., and Cooper, H. (2017). Self-grading and peer-grading for formative and summative assessments in 3rd through 12th grade classrooms: a meta-analysis. *J. Educ. Psychol.* 109, 1049–1066. doi: 10.1037/edu0000190

Sargeant, J. (2008). Toward a common understanding of self-assessment. *J. Contin. Educ. Health Prof.* 28, 1–4. doi: 10.1002/chp.148

Sargeant, J., Mann, K., van der Vleuten, C., and Metsemakers, J. (2008). "Directed" self-assessment: practice and feedback within a social context. *J. Contin. Educ. Health Prof.* 28, 47–54. doi: 10.1002/chp.155

Shute, V. (2008). Focus on formative feedback. *Rev. Educ. Res.* 78, 153–189. doi: 10.3102/0034654307313795

Silver, I., Campbell, C., Marlow, B., and Sargeant, J. (2008). Self-assessment and continuing professional development: the Canadian perspective. *J. Contin. Educ. Health Prof.* 28, 25–31. doi: 10.1002/chp.152

Siow, L.-F. (2015). Students' perceptions on self- and peer-assessment in enhancing learning experience. *Malaysian Online J. Educ. Sci.* 3, 21–35.

Son, L. K., and Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *J. Exp. Psychol.* 26, 204–221. doi: 10.1037/0278-7393.26.1.204

Tan, K. (2004). Does student self-assessment empower or discipline students? *Assess. Evalu. Higher Educ.* 29, 651–662. doi: 10.1080/0260293042000227209

Tan, K. (2009). Meanings and practices of power in academics' conceptions of student self-assessment. *Teach. High. Educ.* 14, 361–373. doi: 10.1080/13562510903050111

Taras, M. (2008). Issues of power and equity in two models of self-assessment. *Teach. High. Educ.* 13, 81–92. doi: 10.1080/13562510701794076

Tejeiro, R. A., Gomez-Vallecillo, J. L., Romero, A. F., Pelegrina, M., Wallace, A., and Emberley, E. (2012). Summative self-assessment in higher education: implications of its counting towards the final mark. *Electron. J. Res. Educ. Psychol.* 10, 789–812.

Thawabieh, A. M. (2017). A comparison between students' self-assessment and teachers' assessment. *J. Curri. Teach.* 6, 14–20. doi: 10.5430/jct.v6n1p14

Tulgar, A. T. (2017). Selfie@ssessment as an alternative form of self-assessment at undergraduate level in higher education. *J. Langu. Linguis. Stud.* 13, 321–335.

van Helvoort, A. A. J. (2012). How adult students in information studies use a scoring rubric for the development of their information literacy skills. *J. Acad. Librarian.* 38, 165–171. doi: 10.1016/j.acalib.2012.03.016

van Loon, M. H., de Bruin, A. B. H., van Gog, T., van Merriënboer, J. J. G., and Dunlosky, J. (2014). Can students evaluate their understanding of cause-and-effect relations? The effects of diagram completion on monitoring accuracy. *Acta Psychol.* 151, 143–154. doi: 10.1016/j.actpsy.2014.06.007

van Reybroeck, M., Penneman, J., Vidick, C., and Galand, B. (2017). Progressive treatment and self-assessment: Effects on students' automatisation of grammatical spelling and self-efficacy beliefs. *Read. Writing* 30, 1965–1985. doi: 10.1007/s11145-017-9761-1

Wang, W. (2017). Using rubrics in student self-assessment: student perceptions in the English as a foreign language writing context. *Assess. Evalu. High. Educ.* 42, 1280–1292. doi: 10.1080/02602938.2016.1261993

Wollenschläger, M., Hattie, J., Machts, N., Möller, J., and Harms, U. (2016). What makes rubrics effective in teacher-feedback? Transparency of learning goals is not enough. *Contemp. Educ. Psychol.* 44–45, 1–11. doi: 10.1016/j.cedpsych.2015.11.003

Yan, Z., and Brown, G. T. L. (2017). A cyclical self-assessment process: towards a model of how students engage in self-assessment. *Assess. Evalu. High. Educ.* 42, 1247–1262. doi: 10.1080/02602938.2016.1260091

Yilmaz, F. N. (2017). Reliability of scores obtained from self-, peer-, and teacher-assessments on teaching materials prepared by teacher candidates. *Educ. Sci.* 17, 395–409. doi: 10.12738/estp.2017.2.0098

Zimmerman, B. J. (2000). Self-efficacy: an essential motive to learn. *Contemp. Educ. Psychol.* 25, 82–91. doi: 10.1006/ceps.1999.1016

Zimmerman, B. J., and Schunk, D. H. (2011). "Self-regulated learning and performance: an introduction and overview," in *Handbook of Self-Regulation of Learning and Performance,* eds B. J. Zimmerman and D. H. Schunk (New York, NY: Routledge), 1–14.