The Impact of Self-assessment on Academic Performance: A Meta-analysis Study

Pınar Karaman
Sinop University, Turkey

To cite this article:


The International Journal of Research in Education and Science (IJRES) is a peer-reviewed scholarly online journal. This article may be used for research, teaching, and private study purposes. Authors alone are responsible for the contents of their articles. The journal owns the copyright of the articles. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of the research material. All authors are requested to disclose any actual or potential conflict of interest including any financial, personal or other relationships with other people or organizations regarding the submitted work.

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.
The Impact of Self-assessment on Academic Performance: A Meta-analysis Study

Pınar Karaman

Abstract

This meta-analysis study synthesizing the results of experimental and quasi-experimental studies examined the effects of self-assessment interventions on student academic performance from primary education to higher education. A total of 16 studies with 46 effect sizes involving more than 7,650 participants were included in the analysis. Research synthesis showed that an overall small influence of self-assessment interventions on academic performance (g=.37, p< .05). Additionally, moderator analysis was used to examine moderating effects of some variables. The analysis indicated that traditional self-assessment interventions without external feedback have significantly a larger effect (g =.47, p<.05) than self-assessment with external feedback (g=.28, p<.05) on academic performance. However, effectiveness of other moderating variables (e.g. education level, assessment criteria type, self-assessment training) on academic performance were not statistically significant. The results suggest that further empirical studies are needed to reveal the moderating effects of self-assessment.

Introduction

Providing high quality feedback to students on their academic tasks is prominent in several decades (Sadler, 1989). Self-assessment as a central element of formative assessment and classroom assessment (Andrade & Valtcheva, 2009; Brown, Andrade, & Chen, 2015; Brown & Harris, 2013) provides feedback to promote student learning and academic performance (Andrade, 2010). Assessment for learning strategy such as self-assessment and peer assessment allows students’ active involvement in assessment (Black et al., 2003). Students collect information, identify, evaluate, and reflect about their own works based on explicit criteria and standards through self-assessment (Boud, 1986; Brown & Harris, 2013; Yan & Brown, 2016). Following specific criteria in their self assessment practices ensures that students maintain a degree of quality in their work and helps them direct their attentions to a particular task (Sadler, 1989). Self-assessment has been mostly utilized for formative purposes as a learning strategy but also used for summative purposes (Boud 1999; Panadero, Brown & Strijbos, 2016; Yan, 2016). On the other hand, accuracy of self-assessment for summative purposes has been argued because of some reliability issues (Brooks, 2002; Brown, Andrade and Chen, 2015). When comparing to teacher assessment and peer assessment, raters are generally more generous with self-assessment (Gürlen, Boztunç Öztürk, & Emiñoğlu, 2019; Karakaya, 2015). The research studies showed that some relationship exists between self-assessment and students’ academic achievement (Brown & Harris, 2013; McDonald & Boud, 2003).
studies also revealed that self-assessment is an effective assessment for learning strategy to develop self-regulated and lifelong learners (Panadero & Alonso-Tapia, 2013). Student involvement in assessment through self-assessment is an important component of self-reflection of self-regulation (Zimmerman, 2008). In self-regulated learning model, the sub-processes of monitoring and self-evaluation are related to self-assessment (Zimmerman, 2000). Students’ self-assessment practices in the classrooms increase their interest and motivation, support them to be more proficient in their own works, promote their self-regulated learning (Oscarson, 2013; van Loon and Roebers, 2017; Vasu et al., 2020), and improve their academic success (Desoete, Roeyers, & Buysse, 2001; Sharma et al., 2016; Winne, 2005; Zimmerman, 2008). Thus, students will be more proactive learners with accurate self-assessment (Boud, 2013). Even though self-assessment is an essential component of effective learning (Black & William, 1998), self-assessment is not applied commonly in many classrooms (Brown & Harris, 2013).

There are three types of self-assessment (Brown & Harris, 2013). One that is based on self-regulated learning is to allow students to compare their own performance with desired goals and to revise it accordingly (Andrade, 2010; Hattie & Timperley, 2007). Another self-assessment type is to require students to evaluate their own performance on a test by marking, grading, or ranking. Thirdly, self-assessment with rubrics, scripts, or checklists is also common for students to assess their own works. Self-assessment tools with rubrics and scripts including clear assessment criteria/standards help students self-grade their works (Panadero, Alanso-Tapia, & Huertas, 2012). Rubrics used as a self-assessment tool provide a list of criteria to determine the levels of quality for students’ specific tasks/performances (Andrade, 2000). A good rubric-referenced self-assessment tool provides feedback to students to guide them to make further revisions for their improvements (Andrade, 2008). A self-assessment tool with scripts gives specific questions to students to answer regarding the structured steps of tasks (Alanso-Tapia & Panadero, 2010; Panadero, Alanso-Tapia, & Huertas, 2012). Checklists also present a list of criteria to students to self-grade the process of their tasks step by step (Burke, 2010). Generally providing a self-assessment tool with rubrics, checklists, or scripts can guide students to understand the tasks deeply and monitor their own tasks for achievement (Andrade & Valtcheva, 2009; Panadero, Alanso-Tapia, & Huertas, 2012; Vasu et al., 2020; Veenman, 2011).

Student self-assessment without feedback is more common than self-assessment with external feedback coming from teachers or peers (Taras, 1999). Since self-assessment is crucial for self-regulated learning, feedback is necessary for the accuracy of this assessment (Andrade, 2018; Panadero, Fernandez-Ruiz, & Sanchez-Iglesias, 2020). However, the studies investigating the effectiveness of feedback on self-assessment are very limited (Panadero, Alanso-Tapia, & Huertas, 2012; Panadero, Fernandez-Ruiz, & Sanchez-Iglesias, 2020; Raaijmakers et al., 2019; Taras, 2003). There are also research studies suggesting that self-assessment training for students before self-assessment interventions contributed to increase effectiveness of self-assessment, self-regulated learning, and academic performance (Baars et al., 2014; Kostons, Van Gog, T., & Paas, 2010, 2012; McDonald & Boud, 2003). To sum up, different types of self-assessment interventions may have different impacts on student learning (Brown & Harris, 2013; Panadero, Jonsson, & Botella, 2017; Sitzmann et al., 2010). Therefore, examining the effect of different types of self-assessment interventions on learning outcomes is crucial.
Several studies have focused on meta-analytic review to explore the effectiveness of student self-assessment. Falchikov and Boud (1989) examined the validity and reliability of self-assessment to compare with teacher assessment. Brown and Harris (2013) reviewed 23 studies that covered K-12 students about the effectiveness of self-assessment. The median effect size between .40 and .45 suggested that self-assessment has a positive small impact on student learning. Sitzmann et al. (2010) reviewed and concluded that relationships between self-assessment and affective learning outcomes (motivation and satisfaction) was highest ($r = .59$; $r = .51$, respectively), but the relationship between self-assessment and cognitive learning outcome was moderate ($r = .34$). Li and Zhang (2020) specifically examined the relationship between self-assessment and language performance with meta-analysis method. They reported that overall correlation coefficient between self-assessment and language performance was .46. In Youde’s (2019) meta-analysis review, experimental studies regarding the impact of self-reflective assessment as instructional approach on academic achievement was investigated. The study indicated that self-reflective assessment as a cognitive and metacognitive strategy has an overall small effect size ($d = .46$) on academic achievement. Moreover, Panadero, Johnson, and Botella (2017) explored the effects of self-assessment on self-regulated learning and self-efficacy with four meta-analyses. The findings showed that effect sizes on different measures of self-regulated learning ranged from small to medium ($d = .23$, $d = .43$, and $d = .65$). And also effect size on self-efficacy as one of the motivational variables was obtained as .73. Besides, Andrade (2019) made a qualitative review of 76 empirical studies about self-assessment. The qualitative review suggested that self-assessment is useful for academic achievement and self-regulated learning. However, Andrade (2019) argued that effectiveness of self-assessment work is not clear. To better understand the influence of self-assessment interventions on academic performance, it seems that more research studies are necessary. Given qualitative and quantitative reviews and increasing emphasis on students’ involvement on assessment specifically self-assessment, current study aimed to use a meta-analysis method to statistically synthesize research findings regarding the effectiveness of self-assessment interventions from primary education to higher education. Following research questions were investigated in the present study:

a) What impact do self-assessment interventions have on student academic performance?

b) Do the moderating variables (education level, external feedback, self-assessment criteria type, self-assessment training) influence the effectiveness of the self-assessment processes?

**Method**

**Inclusion Criteria**

The following inclusion criteria were applied to perform meta-analysis. The research studies:

1. designed to implement self-assessment interventions;
2. having true or quasi experimental design with at least one control group,
3. aimed to improve student academic performance,
4. covering articles, master’s and doctoral theses,
5. published between 1994 and 2021 (until May),
6. published in English, and
7. included sufficient statistical data to compute effect sizes.
Literature Research

To identify primary studies, the search criterion was used in the literature review process. The following keyword searches were performed: self-assessment or self-evaluation or self-feedback or self-rating and academic performance or academic achievement or learning or learning outcomes through databases such as Google Scholar, ERIC, and Springer Link.

There is a controversy about the inclusion of unpublished data in meta-analysis studies. However, some evidence showed that published studies tend to show greater treatment effect than unpublished studies (Conn et al., 2003; Driessen et al., 2015; Hopewell et al., 2007). The problem called as publication bias may cause over-estimating effect sizes in meta-analysis (Hopewell et al., 2007). Therefore, not only published studies but also unpublished studies (master’s and doctoral theses) were included in the search.

Prisma flow chart (adapted from Moher et al., 2009) was used to demonstrate how primary studies were included in the meta-analysis (Figure 1). Initial research resulted in 3,000 records. After excluding duplicated records, screening and analyzing were performed for eligibility of studies. It was found that 119 studies were eligible. However, some of the studies were excluded for several reasons (i.e. no experimental design, no control group, not designed for self-assessment interventions, not studied effect of self-assessment on academic performance).

![Prisma flow chart](image)

Figure 1. Study of Flow Chart

Two raters independently coded the eligible studies for four potential moderators. The coded moderators were education level, external feedback, self-assessment criteria type, and self-assessment training. Inter-rater
reliability for each coded moderator variable ranged from .78 to perfect agreement. Inter-rater reliability values for moderator variables had a perfect agreement for education level, .85 for external feedback, .78 for self-assessment criteria type, and .80 for self-assessment training. Education level variable was coded into primary education, secondary education, and tertiary education. The variable of external feedback on performance was coded as student self-assessment either including external feedback or not (coded as yes or no). Self-assessment criteria variable was coded as self-grading without using any specific tool and self-assessment with the use of specific tools (rubrics, control lists, and/or scripts). Whether the students have self-assessment training or not before self-assessment practices was coded as yes or no, accordingly. Ultimately, the raters had an agreement with 18 studies that met the inclusion criteria. Therefore, 18 primary studies were decided to be included in the study. The list of studies included in the meta-analysis was shown in Table 1.

<table>
<thead>
<tr>
<th>Included study</th>
<th>Subject</th>
<th>Education level</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrade &amp; Boulay-A 2003</td>
<td>Literature</td>
<td>Secondary</td>
<td>119</td>
</tr>
<tr>
<td>Andrade &amp; Boulay-B 2003</td>
<td>Literature</td>
<td>Secondary</td>
<td>119</td>
</tr>
<tr>
<td>Andrade &amp; Boulay-C 2003</td>
<td>Literature</td>
<td>Secondary</td>
<td>98</td>
</tr>
<tr>
<td>Andrade &amp; Boulay-D 2003</td>
<td>Literature</td>
<td>Secondary</td>
<td>98</td>
</tr>
<tr>
<td>Goto Butler &amp; Lee-A 2010</td>
<td>Foreign Language</td>
<td>Primary</td>
<td>124</td>
</tr>
<tr>
<td>Goto Butler &amp; Lee-B 2010</td>
<td>Foreign Language</td>
<td>Primary</td>
<td>130</td>
</tr>
<tr>
<td>Goto Butler &amp; Lee-C 2010</td>
<td>Foreign Language</td>
<td>Primary</td>
<td>124</td>
</tr>
<tr>
<td>Goto Butler &amp; Lee-D 2010</td>
<td>Foreign Language</td>
<td>Primary</td>
<td>130</td>
</tr>
<tr>
<td>Goto Butler &amp; Lee-E 2010</td>
<td>Foreign Language</td>
<td>Primary</td>
<td>124</td>
</tr>
<tr>
<td>Goto Butler &amp; Lee-F 2010</td>
<td>Foreign Language</td>
<td>Primary</td>
<td>130</td>
</tr>
<tr>
<td>Clift 2015</td>
<td>Math</td>
<td>Primary</td>
<td>130</td>
</tr>
<tr>
<td>Fontana &amp; Fernandes-A 1994</td>
<td>Math</td>
<td>Primary</td>
<td>281</td>
</tr>
<tr>
<td>Fontana &amp; Fernandes-B 1994</td>
<td>Math</td>
<td>Primary</td>
<td>386</td>
</tr>
<tr>
<td>Guzman et al.-A 2007</td>
<td>Computer Science</td>
<td>Tertiary</td>
<td>61</td>
</tr>
<tr>
<td>Guzman et al.-B 2007</td>
<td>Computer Science</td>
<td>Tertiary</td>
<td>91</td>
</tr>
<tr>
<td>Hotard 2010</td>
<td>Math</td>
<td>Secondary</td>
<td>73</td>
</tr>
<tr>
<td>Mazloomi &amp; Khabiri-A 2018</td>
<td>Foreign Language</td>
<td>Tertiary</td>
<td>60</td>
</tr>
<tr>
<td>McDonald &amp; Boud-A 2003</td>
<td>Business Studies</td>
<td>Secondary</td>
<td>515</td>
</tr>
<tr>
<td>McDonald &amp; Boud-B 2003</td>
<td>Humanities</td>
<td>Secondary</td>
<td>515</td>
</tr>
<tr>
<td>McDonald &amp; Boud-C 2003</td>
<td>Science</td>
<td>Secondary</td>
<td>515</td>
</tr>
<tr>
<td>McDonald &amp; Boud-D 2003</td>
<td>Technical Studies</td>
<td>Secondary</td>
<td>515</td>
</tr>
<tr>
<td>Memiş &amp; Seven-A 2015</td>
<td>Science</td>
<td>Primary</td>
<td>67</td>
</tr>
<tr>
<td>Memiş &amp; Seven-B 2015</td>
<td>Science</td>
<td>Primary</td>
<td>67</td>
</tr>
<tr>
<td>Memiş &amp; Seven-C 2015</td>
<td>Science</td>
<td>Primary</td>
<td>67</td>
</tr>
<tr>
<td>Memiş &amp; Seven-D 2015</td>
<td>Science</td>
<td>Primary</td>
<td>67</td>
</tr>
<tr>
<td>Memiş &amp; Seven-E 2015</td>
<td>Science</td>
<td>Primary</td>
<td>67</td>
</tr>
<tr>
<td>Memiş &amp; Seven-F 2015</td>
<td>Science</td>
<td>Primary</td>
<td>67</td>
</tr>
<tr>
<td>Included study</td>
<td>Subject</td>
<td>Education level</td>
<td>Sample Size</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>------------------------</td>
<td>-----------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Papandero, Alonso-Tapia &amp; Reche-B 2013</td>
<td>New Technologies</td>
<td>Tertiary</td>
<td>49</td>
</tr>
<tr>
<td>Ross, Hogaboam-Gray &amp; Rolheiser 2002</td>
<td>Math</td>
<td>Primary</td>
<td>516</td>
</tr>
<tr>
<td>Vasileiadou &amp; Karadimitriou-A 2021</td>
<td>History</td>
<td>Primary</td>
<td>70</td>
</tr>
<tr>
<td>Vasileiadou &amp; Karadimitriou-B 2021</td>
<td>Language</td>
<td>Primary</td>
<td>70</td>
</tr>
<tr>
<td>Yu-A 2013</td>
<td>Math</td>
<td>Secondary</td>
<td>533</td>
</tr>
<tr>
<td>Yu-B 2013</td>
<td>Math</td>
<td>Secondary</td>
<td>533</td>
</tr>
<tr>
<td>Yu-C 2013</td>
<td>Math</td>
<td>Secondary</td>
<td>171</td>
</tr>
<tr>
<td>Yu-D 2013</td>
<td>Math</td>
<td>Secondary</td>
<td>203</td>
</tr>
<tr>
<td>Yuan, Savadatti &amp; Zheng-A 2021</td>
<td>Engineering Course</td>
<td>Tertiary</td>
<td>56</td>
</tr>
<tr>
<td>Yuan, Savadatti &amp; Zheng-B 2021</td>
<td>Engineering Course</td>
<td>Tertiary</td>
<td>56</td>
</tr>
<tr>
<td>Yuan, Savadatti &amp; Zheng-C 2021</td>
<td>Engineering Course</td>
<td>Tertiary</td>
<td>56</td>
</tr>
<tr>
<td>Yuan, Savadatti &amp; Zheng-D 2021</td>
<td>Engineering Course</td>
<td>Tertiary</td>
<td>56</td>
</tr>
<tr>
<td>Yuan, Savadatti &amp; Zheng-E 2021</td>
<td>Engineering Course</td>
<td>Tertiary</td>
<td>56</td>
</tr>
<tr>
<td>Yuan, Savadatti &amp; Zheng-F 2021</td>
<td>Engineering Course</td>
<td>Tertiary</td>
<td>56</td>
</tr>
<tr>
<td>Zamora, Suarez &amp; Ardura 2018</td>
<td>Natural Sciences</td>
<td>Secondary</td>
<td>130</td>
</tr>
<tr>
<td>deMarcos et al-A 2010</td>
<td>Technology Course</td>
<td>Secondary</td>
<td>100</td>
</tr>
<tr>
<td>deMarcos et al-B 2010</td>
<td>Physics Course</td>
<td>Secondary</td>
<td>98</td>
</tr>
<tr>
<td>deMarcos et al-C 2010</td>
<td>Nursery course</td>
<td>Tertiary</td>
<td>56</td>
</tr>
</tbody>
</table>

**Final Sample**

After an inclusion/exclusion criterion was detailed above, a search for outliers was conducted. Effect size of each study for academic performance was examined in SPSS to identify outliers. Tukey fence method was used to detect outliers by using interquartile range (Tukey, 1977). A few studies having extreme effect sizes as outliers were found and removed (Cömert & Kutlu, 2018; Mazloomi & Khabiri, 2018 B effect size; Nbina & Viko, 2010; Vasileiadou & Karadimitriou, 2021 C effect size). Eventually, final sample of the study was 16 primary studies reporting 46 independent self-assessment experiments with a total of 7654 participants. It was found that most of the collected studies had multiple groups such as different schools, grades, and multiple assessments such as different subject areas, different self-assessment types. Therefore, some of the studies had more than one effect sizes.

**Analysis**

Pro-meta 3 was used to perform meta-analysis. The overall mean effect size by using Hedge's g (Hedges, 1981)
was calculated to estimate effect of self-assessment interventions on academic performance. This meta-analysis employed a random effects model. Random effects model assumes that true effect size may show differences due to differences of studies (Borenstein et al., 2009). Q statistic and $I^2$ were reported to measure heterogeneity in effect sizes. This statistics showed whether the variability in effect sizes was larger than sampling error alone (Lipsey & Wilson, 2001). Q statistic with a significant p value indicates the heterogeneity. Besides, moderator analysis was used to determine variability between studies by using Q statistics ($Q_{between}$).

**Publication Bias**

The possibility of publication bias was examined in the meta-analysis study. Initially, funnel plot method was used to assess if sample of studies distributed symmetrically around the mean effect size (Borenstein et al., 2009; Light & Pillemer, 1984). The funnel plot denoted that the major of the studies was distributed symmetrically (Figure 2). It may be an indication of the absence of severe publication bias.

![Funnel Plot](image)

Figure 2. Funnel Plot

Fail-safe N test method (Rosenthal, 1991) was used to estimate if publication bias existed in the sample of studies. According to Rosenthal’s criteria, the fail-safe number ($N=2560$) was considered as robust since the number was greater than 240 ($5k+10$). The non-significant results of Egger’s linear regression test ($p = .96$), and Begg and Mazumdar’s rank correlation test ($p = .77$) also suggested that there is no publication bias threat in the sample of studies.

**Results**

The distribution of effect sizes for student academic performance was presented in the forest plots in Figure 3. The diamond on the bottom indicated the overall effect size. The effect of self-assessment on academic achievement was analyzed. The results showed that 16 studies with 46 effect sizes yielded a small influence of self-assessment interventions on academic performance ($g = .37$, $SE = .04$, 95 % CI [ .29, .46], $p< .05$). Heterogeneity across effect sizes was significant [$Q (45) = 132.94$, $p< .05$, $I^2 = 66.15$]. $I^2$ suggested that 66.15
% of the observed heterogeneity considered as medium level (Higgins et al., 2003) was due to between study differences.

Figure 3. Forest Plot for Self-Assessment Impacts on Academic Performance

Moderator Analysis

Since heterogeneity of test was significant, moderator analysis was performed to investigate variations in the study. The results of the moderator analyses for student academic performance were presented in Table 2. Mixed effect analysis showed a significant difference in effect size for external feedback moderator \([Q(1) = 5.86, p < .05]\). This finding denoted that self-assessment interventions without external feedback \((g = .47, p < .05)\) have significantly larger effect than self-assessment interventions with external feedback \((g = .28, p < .05)\) on academic performance. However, other variables (education levels, self-assessment criteria, and self-assessment instrument type) did not have significant moderating effects. Mixed effects analysis indicated no significant difference in effect sizes for education level \([Q (2) = 1.05, p = .59]\). The effect of self-assessment at primary education level \((g = .43, p < .05)\) and secondary education level \((g = .37, p < .05)\) were statistically significant small effect but at tertiary education level was not significant on academic performance \((g = .28, p > .05)\). The influence of self-assessment criteria type on academic performance was examined with moderator analysis. The results showed that effect sizes were not statistically different between self-assessment interventions using any specific assessment tool (rubric, scripts etc.) \((g = .42, p < .05)\) and not using any specific tool \((g = .36, p < .05)\) on academic performance \([Q (1) = .31, p = .57]\). The third moderator variable as self-assessment training was
also tested with mixed effect analysis. The results showed no significant difference in effect sizes between students exposed to training in self-assessment ($g = .36, p < .05$) and not exposed to training ($g = .38, p < .05$) on learning outcomes [$Q (1) = .05, p = .81$].

### Table 2. Results of the Moderator Analysis for Academic Performance

<table>
<thead>
<tr>
<th>Moderator variable</th>
<th>k</th>
<th>g</th>
<th>95% CI</th>
<th>p</th>
<th>SE</th>
<th>N</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Education Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>12</td>
<td>.43</td>
<td>[.30, .56]</td>
<td>.00</td>
<td>.07</td>
<td>2215</td>
<td>1.05</td>
</tr>
<tr>
<td>Secondary</td>
<td>22</td>
<td>.37</td>
<td>[.26, .47]</td>
<td>.00</td>
<td>.05</td>
<td>4737</td>
<td></td>
</tr>
<tr>
<td>Tertiary</td>
<td>12</td>
<td>.28</td>
<td>[-.05, .60]</td>
<td>.09</td>
<td>.16</td>
<td>702</td>
<td></td>
</tr>
<tr>
<td><strong>External feedback</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>27</td>
<td>.28</td>
<td>[.15, .41]</td>
<td>.00</td>
<td>.06</td>
<td>3532</td>
<td>5.86*</td>
</tr>
<tr>
<td>No</td>
<td>19</td>
<td>.47</td>
<td>[.38, .57]</td>
<td>.00</td>
<td>.05</td>
<td>3726</td>
<td></td>
</tr>
<tr>
<td><strong>Self-assessment criteria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not using any assessment tool</td>
<td>28</td>
<td>.36</td>
<td>[.27, .45]</td>
<td>.00</td>
<td>.05</td>
<td>5801</td>
<td>.31</td>
</tr>
<tr>
<td>Using specific assessment tool (rubrics, scripts etc.)</td>
<td>18</td>
<td>.42</td>
<td>[.23, .61]</td>
<td>.00</td>
<td>.10</td>
<td>1853</td>
<td></td>
</tr>
<tr>
<td><strong>Self-assessment training</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>13</td>
<td>.36</td>
<td>[.20, .51]</td>
<td>.00</td>
<td>.08</td>
<td>3867</td>
<td>.05</td>
</tr>
<tr>
<td>No</td>
<td>33</td>
<td>.38</td>
<td>[.28, .48]</td>
<td>.00</td>
<td>.05</td>
<td>3787</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05; k = number of effects; SE = standard error

### Discussion

The purpose of this meta-analysis study was to statistically synthesize empirical studies on the effect of self-assessment interventions on academic performance within educational contexts from primary education to higher education. A total of 46 effect sizes from 16 studies was calculated. The results showed that self-assessment interventions had a small positive impact on student academic performance within educational contexts ($g = .37$). This finding was similar to results from previous meta-analysis studies about the effectiveness of self-assessment (Brown, & Harris, 2013; Panadero, Johnson, and Botella, 2017; Youde, 2019). On the other hand, some of the meta-analytic studies in the literature examined the correlation between self-assessment and learning outcomes, and found a moderate relationship (Sitzmann et al., 2010; Li & Zhang, 2020). The overall results showed that these effect ranges varied from small to moderate. The differences may occur due to using different moderators in the studies (Dignath, Büttner, & Langfeldt, 2008).
The present meta-analysis study helped to understand the impact of self-assessment interventions (self-assessment with and without external feedback) on academic performance. It was found that there was a significant difference in effect sizes between self-assessment interventions with feedback and without feedback on academic performance/learning outcomes. Even though self-assessment intervention without feedback are usually traditionally used (Taras, 1999), self-assessment procedure with feedback is also crucial for self-regulated learning (Andrade, 2018). Self-assessment with tutor or peer feedback would be useful for accuracy of self-assessment and improving learning process (Andrade, 2018; Taras, 1999, 2001). One of the meta-analysis studies (Sitzmann et al., 2010) examined the relationship between self-assessment and learning and concluded that self-assessment with feedback had a higher (r=.28) relationship than without feedback (r=.14) in the courses. However, the present meta-analysis study showed that traditional self-assessment interventions without feedback have a larger effect size on academic performance than self-assessment interventions with feedback. Since research studies suggested that self-assessment with feedback has a crucial role on student learning, more self-assessment research studies with specific types of self-assessment feedback are needed.

The impact of other moderating variables such as educational level, assessment criteria type, and self-assessment training were also examined. None of these variables had a significant effect on academic performance. When examining the impact of education levels separately, the result suggested that there is a statistically positive effect of self-assessment on academic performance at different education levels such as primary school level and secondary level. Besides, the results showed that self-assessment criteria type and self-assessment training were not significant moderators of the effect on academic performance.

The research studies suggested that self-assessment with specific assessment tool (rubrics, scripts etc.) has a positive influence on academic performance and learning (Andrade & Valtcheva, 2009; Panadero & Jonsson, 2013). Students who were exposed to training in self-assessment can also improve their academic performance. However, this study indicated that moderating effects of these educational characteristics were not statistically significant on academic performance. Besides, it is crucial to point out that uncovering the effect of these moderating variables on student achievement was difficult due to limited number of studies included in the meta-analysis. The further studies may help to rule out generalization of the results.

**Conclusion**

The present meta-analysis study was designed to address some questions that previous research did not answer such as moderating effects of external feedback, self-assessment training and self-assessment tools. The meta-analytic review suggested that there is a positive impact of self-assessment on student academic performance within educational contexts. Moderating analysis showed that traditional self-assessment interventions without external feedback have a larger impact than self-assessment interventions with external feedback on academic performance. However, effect of some of the moderating variables (education levels, self-assessment tools, self-assessment training) on academic performance were not statistically significant.

Several studies suggested that self-assessment with feedback, training, rubrics, scripts etc. may have a potential
influence to improve student learning and academic performance. However, the meta-analysis results found a little evidence about the effectiveness of some moderator variables on self-assessment processes. Therefore, more empirical studies are needed to investigate the effectiveness of self-assessment practices allowing students’ active involvement on academic performance.

**Limitations**

One of the limitations of this meta-analysis study was the limited number of the empirical studies having experimental or quasi experimental design to examine the impact of self-assessment on academic achievement. To obtain more evidence in meta-analysis studies, more experimental studies are needed to examine the effectiveness of self-assessment. The present meta-analysis study investigated the moderating effect of some variables. In further studies, more potential moderating variables such as accuracy of self-assessment and subject areas should be investigated. Another limitation of the study was the difficulty in identifying the coding scheme of included studies that have not reported the student self-assessment interventions in detail. A detailed report in self-assessment interventions for empirical studies would be useful for meta-analysis studies to collect more evidence in that area.

**References**

Studies included in the meta-analysis marked as an *.


Putting it into practice. Maidenhead: Open University Press.


Karaman

& practice, 10(2), 209-220. http://dx.doi.org/10.1080/0969594032000121289


**Author Information**

Pınar Karaman  
https://orcid.org/0000-0002-2218-2701  
Sinop University  
Sinop  
Turkey  
Contact e-mail: pkaraman@sinop.edu.tr