Assessment in Education: Principles, Policy & Practice

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/caie20

Effects of self-assessment diaries on academic achievement, self-regulation, and motivation

Zi Yan , Ming Ming Chiu & Po Yuk Ko

To cite this article: Zi Yan , Ming Ming Chiu & Po Yuk Ko (2020) Effects of self-assessment diaries on academic achievement, self-regulation, and motivation, Assessment in Education: Principles, Policy & Practice, 27:5, 562-583, DOI: 10.1080/0969594X.2020.1827221

To link to this article: https://doi.org/10.1080/0969594X.2020.1827221

ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE
2020, VOL. 27, NO. 5, 562–583
https://doi.org/10.1080/0969594X.2020.1827221

Routledge
Taylor & Francis Group

Effects of self-assessment diaries on academic achievement, self-regulation, and motivation

Zi Yan [a], Ming Ming Chiu [b] and Po Yuk Ko [a]

[a]Department of Curriculum and Instruction, The Education University of Hong Kong, Hong Kong, China; [b]Department of Special Education and Counselling, The Education University of Hong Kong, Hong Kong, China

ABSTRACT

Although students who self-assess effectively often learn better, creating effective, low-cost interventions to help them do so is a critical challenge. This study examined the effects of a self-assessment diary intervention on 74 Form 1 (M_{age} = 12.2 years) students' academic achievement, self-regulation, and motivation. After each homework assignment, students in the experimental group (n = 37) completed a standardised self-assessment diary, while students in the control group (n = 37) did no additional work. *Difference-in-differences* analyses showed that self-assessment diaries significantly enhanced students' academic achievement, self-efficacy, and intrinsic value. Students with lower past achievement benefited more than other students from the intervention. The intervention had no significant impact on effort regulation and self-reflection. Furthermore, effort-regulation, self-reflection, self-efficacy, and intrinsic value all did not mediate the link between self-assessment diaries and academic achievement. The findings can inform researchers and educators aiming to help students self-assess effectively to improve their learning.

ARTICLE HISTORY
Received 3 October 2019
Accepted 15 September 2020

KEYWORDS
Self-assessment; diary; self-regulation; motivation; learning intervention

Introduction

Students who self-assess effectively can identify their academic strengths and weaknesses, and then allocate their educational resources and efforts (*self-regulation*) efficiently to form adaptive learning strategies, which likely lead to better learning performance than others (Boud, 1995; Yan & Brown, 2017). Hence, self-assessment is a crucial learning strategy for developing self-regulation and life-long learning (Panadero & Alonso-Tapia, 2013; Yan, 2016; Yan & Brown, 2017). Despite some theoretical and empirical support for a positive link between self-assessment and learning outcomes (Ibabe & Jauregizar, 2010; Ramdass & Zimmerman, 2008; Topping, 2003), its effects differ across studies, possibly because of variation in the intervention designs, implementations, teaching methods, learning environments, available supports for students, and so on (Brown & Harris, 2013; Panadero & Jonsson, 2013; Topping, 2003). Therefore, we need more empirical research

CONTACT Zi Yan zyan@eduhk.hk Department of Curriculum and Instruction, The Education University of Hong Kong, Hong Kong
Supplemental data for this article can be accessed here.

(especially with rigorous experimental designs) to identify effective interventions (especially low-cost, easy-to-implement ones). Furthermore, more work is necessary to (a) determine whether the effects of self-assessment differ across students by ability and (b) identify the possible factors mediating the effects of self-assessment on learning outcomes.

Ideally, self-assessment interventions can effectively develop students' self-assessment skills to aid their learning with little teacher effort, time or adaptation. In many education contexts, like Hong Kong, education reforms and evaluations demand much of teachers' effort, time and attention, so only such interventions are likely to be sustainable long-term. Hence, this study tests the effect of a self-assessment diary intervention on students' learning outcomes, specifically academic achievement, self-efficacy, intrinsic value, effort regulation, and self-reflection across students with different past achievement levels. It also tests for possible self-assessment mechanisms that might improve student learning. We argue that by presenting students with relevant questions within diaries, they can write diaries that effectively self-assess their learning processes to improve their learning outcomes. Repeated engagement with such diaries over an extended period of time might enhance student awareness or likelihood of doing self-assessments, which in turn, might improve their learning.

Self-assessment, self-regulation, motivation, and academic achievement

This study followed Yan and Brown (2017) process model and operationally defined self-assessment as 'a process during which students collect information about their own performance, evaluate and reflect on the quality of their learning process and outcomes according to selected criteria to identify their own strengths and weaknesses (p. 1248).' The model acknowledges the complexity of self-assessment and reveals the inner processes of self-assessment by depicting common and sequenced self-assessment actions, including determining the performance criteria, self-directed feedback seeking, and self-reflection. See Yan and Brown (2017) for details of this model. As demonstrated in recent review studies (e.g., Brown & Harris, 2013; Panadero et al., 2017), most studies about the impact of self-assessment focused on three groups of dependent variables: self-regulation (e.g., self-reflection), motivation (e.g., self-efficacy), and cognition (e.g., academic achievement).

Zimmerman (2000) defines self-regulation as 'self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals (p. 14).' He further stipulates that a self-regulated learner (1) monitors learning process; (2) evaluates qualities of learning outcomes; and (3) takes actions when there are differences between learning expectations and actual outcomes (Zimmerman, 2015). Hence, there is a close correlation between self-assessment and self-regulation. Studies (e.g., Panadero & Romero, 2014; Yan, 2020) found that self-assessment is an ongoing practice across the whole self-regulated learning process rather than a one-off action occurring only at the end of self-regulated learning. Zimmerman (2002) showed that self-assessment of one's own work enhances one's self-regulation competencies including self-observation, self-judgement, and self-reflection. As demonstrated in Yan and Brown (2017) process model, one key self-assessment action is self-reflecting on one's strengths and weaknesses. Similarly, students can self-assess to detect discrepancies between their goals and

their performances, which aids effort regulation (i.e., control of effort and attention while pursuing learning goals, Carver & Scheier, 2000). After identifying their strengths and weaknesses, students can regulate their efforts accordingly, such as allocating more time and effort to address their weaknesses, identifying application conditions for their strengths, and so on.

Although past studies have linked self-assessment to self-regulated learning, a review study concluded that the empirical evidence for the benefit of self-assessment interventions on self-regulated learning is inconclusive (Brown & Harris, 2013). A more recent meta-analysis study of 19 studies about effects of self-assessment intervention revealed a generally positive effect of self-assessment training on students' self-regulation strategies, but the effect sizes varied according to the types of self-regulation strategies and the measures of self-regulation (Panadero et al., 2017). The mean effect size for negative self-regulation and self-regulation measured with qualitative data[1] were 0.65 and 0.43 respectively. However, the mean effect size for learning self-regulation, mostly measured by self-report questionnaire, was only 0.24. Hence, the hypotheses proposed here are that self-assessment diary can enhance participants' effort regulation (H1) and self-reflection (H2).

Motivation is also influenced by self-assessment (McMillan & Hearn, 2008; Mortimer, 1998). In a meta-analysis of 166 studies, Sitzmann et al. (2010) concluded that self-assessment's links to motivational outcomes (mean effect size is 0.59) were stronger than its links to cognitive learning outcomes (mean effect size is 0.34); they argued that self-assessment 'is generally more useful as an indicator of how learners feel about a course than as an indicator of how much they learned (p. 180).' Self-assessment can shape one's self-efficacy (Kissling & O'Donnell, 2015; Panadero et al., 2017; Ramdass & Zimmerman, 2008). After recognising one's strengths and, more importantly, weaknesses for a task (effective self-assessment), a student can develop suitable strategies to capitalise on their strengths and dedicate study time to address their weaknesses (Yan & Brown, 2017). By doing so, these students can increase their autonomy and control over their efforts on the task, which enhances both their intrinsic motivation (Huang & Liaw, 2007; Mortimer, 1998) and self-efficacy (Vieira & Grantham, 2011).

The effects of assessment interventions on students' self-efficacy and intrinsic value are inconclusive. A meta-analysis reported a higher mean effect size of self-assessment interventions on self-efficacy (0.73) than on self-regulation (ranging from 0.24 to 0.65; Panadero et al., 2017), however, some of these reviewed studies found no significant effects of self-assessment training on self-efficacy across control and experimental groups (e.g., Andrade et al., 2009; Panadero et al., 2013). Meanwhile, studies measuring the effect of self-assessment training on intrinsic value are scarce. One study found that students who were extrinsically motivated showed poorer self-assessment skills, while those with a commitment to their self- and peer-assessment had greater intrinsic motivation (Mortimer, 1998); however, these results are based on interview data in an action research (rather than a controlled experiment). Another study showed that students with stronger mastery goal orientations engaged in more self-assessment, specifically self-directed feedback seeking (effect sizes of 0.77 for primary students and 0.83 for secondary students) and self-reflection (with an effect size of 0.92 for primary students and 0.88 for secondary students, Yan, 2018a). Theoretically, self-assessment is integral to mastery goals as it enables students to recognise their learning progress (McMillan & Hearn,

2008). The positive link between intrinsic value and mastery goal orientation for children, adolescents, and college students (Wigfield & Cambria, 2010) suggests a study of whether self-assessment training increases perceptions of intrinsic value. Hence, we test the hypotheses that self-assessment diary can enhance participants' self-efficacy (H3) or intrinsic value (H4).

Past studies show that students' self-assessment can aid their academic achievement across a range of grade levels and disciplines (McDonald & Boud, 2003; Topping, 2003); notably a review study reported a mean effect size of 0.54 but it varied substantially across studies (Brown & Harris, 2013), including some with no effect (e.g., Andrade & Boulay, 2003; Panadero et al., 2013). Furthermore, most of the studies only reported the overall effect of self-assessment on learning gains, but did not test for different effects across students by past achievement. Notably, a study of a teacher-directed intervention with a learning progress assessment increased the reading fluency of students with lower prior reading skills more than that of students those with higher prior reading skills (Förster et al., 2018). The authors speculated that after the teachers saw the assessment data, they might have targeted their follow-up instruction towards low achievers, thereby benefiting them more than higher achievers. Hence, this study tests whether this self-assessment diary intervention enhances participants' academic achievement overall (H5a) and whether this intervention enhances the academic achievement of students with lower prior academic achievements more than that of students with higher prior academic achievements (H5b).

However, scholars do not agree on a common self-assessment mechanism. Yan and Brown's (2017) model specified the self-assessment process as including concrete actions/strategies such as determining the criteria, seeking feedback, and self-reflection. But the mechanism by which these strategies improve learning is still unknown. Some argued that a student who self-
assesses engages in more cognitive and metacognitive learning activities such as clarifying the learning goals, identifying appropriate assessment criteria, monitoring the learning process, and self-reflecting (Andrade, 2010; Nicol & Macfarlane-Dick, 2006; Panadero et al., 2017; Ramdass & Zimmerman, 2008; Yan & Brown, 2017). In a similar fashion, self-assessment might increase students' self-efficacy and/or intrinsic value, which in turn, improve their learning outcomes. However, it is also possible that after students self-assess, they spend more time and effort to learn; for example, they might seek more feedback on their performance and, therefore, enhance their understanding of their own learning (Yan, 2020). Hence, this study tests whether effort regulation, self-reflection,self-efficacy, or intrinsic value mediate the link between self-assessment and academic achievement.

Diaries and reactivity effect

A diary study is a method for participants to reflect, write and perhaps act on their own thoughts, behaviours, and experiences regularly over a period of time. Using a diary as an educational intervention relies on behaviour changes in a desired direction (*reactivity effect*, Shapiro, 2014) as captured via recordings and/or reflections. This positive reactivity effect is closely related to self-monitoring, a systematic observation and recording of thoughts, feelings, and behaviours (Klug et al., 2018; Reynolds et al., 2015; Schmitz et al.,

2011). Mindful monitoring of one's own behaviour might support learning, though the behaviours being monitored might occur habitually or even outside of conscious awareness (e.g., Reynolds et al., 2015; Schmitz & Perels, 2011). For example, students who wrote in diaries with question prompts to promote self-regulation showed more self-regulated learning behaviours and superior mathematics test outcomes, compared to other students (Schmitz & Perels, 2011).

In addition to its potential academic benefit, using diary as an intervention also has important practical advantages. First, a diary does not occupy instruction time. Students can complete diaries outside of classroom (e.g., at home or during recess) so this activity does not disturb teachers' classroom teaching. Second, this intervention does not require much teacher effort. As many teachers have little time to engage in time-intensive interventions, they are more likely to faithfully implement an intervention that requires less time (such as student diaries) than one that requires more time. Third, this diary intervention is general rather than subject-specific, so it does not require much adaptation to different circumstances (e.g., different teachers, students or courses). Hence, we can use a well-designed diary repeatedly across different classes and different disciplines. Lastly, while used as an intervention method to promote target behaviours, diaries can serve as an informative data collection tool at the same time (Bolger et al., 2003; Iida et al., 2012). It is therefore regarded as a typical tool in the new wave of self-regulation measurement featured by the combination of intervention and measurement (Panadero et al., 2016). As diaries can help collect process data and serve as an intervention method simultaneously, researchers are increasingly recommending its use in educational studies (e.g., Schmidt et al., 2017; Schmitz et al., 2011).

The present study

The present study aimed to investigate the effects of self-assessment diaries on students' academic achievement, self-regulation, self-efficacy, and intrinsic value. We hypothesise that having students use self-assessment diaries enhances their effort regulation (H1), self-reflection (H2), self-efficacy (H3), intrinsic value (H4), or academic achievement (H5a). For the impact on academic achievement, it is hypothesised that students with lower prior academic achievements would benefit more from the diaries (H5b). Furthermore, we test whether effort regulation, self-reflection, self-efficacy, and/or intrinsic value mediate the link between self-assessment and academic achievement.

The setting

Hong Kong's education system is famous for its examination-orientation. In 2001, the Hong Kong government initiated a large-scale curriculum reform aiming to develop students' 'learning to learn' capacity (Curriculum Development Council, 2001). Accordingly, an assessment reform emphasising assessment-for-learning was promoted (Curriculum Development Council, 2002). However, despite the positive change at the system level (e.g., reducing the number of public examinations and introducing school-based assessment into high-stakes examinations) and teachers' awareness of the benefits of assessment for learning, the implementation of assessment for learning in classrooms is far from satisfactory (Berry, 2011; Yan & Cheng, 2015). In recent curriculum

documents, Curriculum Development Council (2014, 2017) took a further step by encouraging schools to integrate assessment-as-learning into the learning and teaching circle. However, it calls for a paradigm shift that re-conceptualises teacher and student roles in the assessment process (Yan & Brown, 2020).

Method

Participants

A power analysis based on the mean effect size of similar past studies ($d = 0.54$; see Brown & Harris, 2013), with $\alpha = 0.05$ and statistical power $1\text{-}\beta = 0.80$ showed that our experimental design with a pre-test and a post-test requires a minimum of 30 students per group. Thus, the study recruited 74 participants (39 females, 35 males; $M_{age} = 12.2$, $SD_{age} = 0.42$) from two Form 1 (grade 7) classes in one secondary school in Hong Kong. The school is a direct subsidy school that is allowed to collect school fees in addition to government subsidies, so students are mainly from families with middle to high socio-economic status. Also, direct subsidy schools have greater flexibility in various areas including resources deployment, curriculum design and student admission. The school uses English as the medium of instruction. Simulations show that a confirmatory one-factor analysis with four variables and expected factor loadings of 0.8 require only a sample size of 60 (Wolf et al., 2013).

Design and procedure

We applied a quasi-experimental approach using intact classes to determine the effect of self-assessment in naturalistic settings. We randomly assign the two classes to one of the two conditions: self-assessment diary (experimental group; n = 37) or no diary (control group; n = 37). The same Integrated Humanities (IH) teacher taught both classes. We selected the IH course for two reasons: 1) IH assignments have many open-ended questions that give students opportunities to self-assess and reflect; and 2) the time and location of this course facilitated the participating school's logistic arrangements. Although the school did not give us access to data regarding students' academic performance on school-based assessments, the teacher said that the students' scores on their last school-based IH assessment indicated that these students in the two classes had similar academic abilities.

The teacher, all participating students and their parents/guardians all signed written consent forms. All participating students took the pre-intervention IH achievement test three days before the intervention, and completed the pre-intervention survey (regarding self-efficacy, intrinsic value, effort regulation, and self-reflection) two days before the intervention. During the five-week intervention, all participants received the same instruction from the same teacher and received eight homework assignments. All the homework assignments are consistent with the content under the unit 'Economic development in Hong Kong during 1900–1941.' The assignments contain mainly open-ended questions including short questions and scenario-based questions. After completing each homework assignment, students in the experimental group had to complete the standardised self-assessment diary. To close the self-feedback loop, teachers encouraged the

students to act on the self-assessment information (e.g., the weaknesses identified and the strategies proposed when doing the last assignment) on subsequent assignments. But teachers did not provide feedback either on students' self-assessment results or on their action upon the self-assessment information (to minimise the teacher's workload). In contrast, students in the control group only completed the assignments, not the diary. All participants took the same survey two days after the intervention and the same attainment test three days after the intervention. We invited four students with various academic performance (one high, two average, and one low) from the experimental group (recommended by the IH teacher) for a 30-minute focus group one week after the intervention. As this study requires the teacher's cooperation, a double-blind design was not feasible. Still, the research team told the teacher that the impact of the treatment is unknown in this exploratory study to reduce the likelihood that such expectations affect her behaviour (*Rosenthal effect*, Rosenthal & Jacobson, 1992).

Materials

Attainment test

To assess each participant's academic competence at each time, we administered two identical attainment tests before and after the intervention. The test was designed and scored by the teacher according to a pre-set rubric. During the intervention, the teacher taught 'Economic development in Hong Kong during 1900–1941'. The tests only have open-ended questions, which are divided into two sections. The first section has three short questions. The second section presents two sets of pictures/descriptions about economic development in Hong Kong, each of which is followed by four questions. The tasks aim to assess subject-based generic skills relevant to the topic, rather than curriculum-based knowledge. The maximum score on the test is 40. With this test design, the difference between pre- and post-tests (if any) is expected to reflects students' development of their covered subject matter knowledge and skills. To minimise any potential practice effect, we intentionally did not provide reference answers or discussions after the pre-test.

Survey

The survey consisted of questions for four scales. Three of them including self-efficacy, intrinsic value, and effort regulation, were from the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich et al., 1991), the most widely-used tool for measuring motivational and self-regulation strategies. Hilpert et al. (2013) re-examined the factor structure of the MSLQ and proposed a three-dimension structure with six variables, including expectancy (variables self-efficacy and control of learning), value (variables intrinsic value and task value), and self-regulation (variables effort-regulation and metacognitive regulation). Thus, this study selected one construct from each dimension as dependent variables to maximise coverage of dimensions while minimising the survey length. Self-assessment is more closely related to self-efficacy, intrinsic value, and effort-regulation than to their counterparts within the same dimension, so these three constructs were used in our study. These three subscales were adopted from the Chinese version (Lee et al., 2010) of the MSLQ including the *self-efficacy* scale (7 items, α =.91; e.g., I am sure I can do an excellent job on the class assignments and

homework.), the *intrinsic value* scale (9 items, α = .92; e.g., I prefer class work that is challenging so I can learn new things.), and the *effort regulation* scale (3 items, α = .79; e.g., Even when study materials are dull and uninteresting, I keep working until I finish.). We added an additional variable, self-reflection, as it is particularly important to self-assessment according to Yan and Brown (2017) self-assessment model. The *self-reflection* scale (7 items, α = .93; e.g., I seek out the reasons for mistakes I made after getting back marked work.) was from the Self-assessment Practice Scale (SaPS) developed by Yan (2018b).

Focus group

The focus groups attempted to collect students' experiences of completing the self-assessment diaries and identify possible explanations for the experimental study results. Hence, we only invited students from the experimental group. The protocol of the focus group contained questions on three domains including 1) students' perceptions and feelings about completing the diary (e.g., How did you feel when filling in the diary?); 2) the way the diary influences their learning, if any (e.g., Did filling the diary influence your learning in the IH subject? If yes, how?); and 3) feedback to improve the design and implementation of the diary (e.g., How can the diary be improved so that it is more useful to help your learning?).

The intervention: standardised self-assessment diaries

Due to the diary intervention's practical advantages over other self-assessment interventions (e.g., less requirements for instructional time and teacher workload), we developed a standardised diary based on the cyclical process model of self-assessment proposed by Yan and Brown (2017). The questions in the diary were deliberately designed to guide students to complete a self-assessment process, as depicted in the model. Consider the self-assessment on assignments for example. Students start the self-assessment process by first determining the assessment criteria for their assignments. Second, students choose the external source for feedback on the quality of their assignments via monitoring (e.g., from past assignments, text books, reference books) or via inquiry with relevant people (e.g., peers, parents). Third, students self-reflect, with the support of available feedback, aiming to identify learning strengths and weaknesses based on the quality of their assignments. Finally, they are encouraged to propose possible strategies for improvement. Relying on mindful self-monitoring (i.e., a systematic observation and recording of self-assessment behaviours), the target behaviour (i.e., self-assessment) might occur habitually or even outside of conscious awareness (e.g., Reynolds et al., 2015; Schmitz & Perels, 2011). Although it appears that the correlation between self-assessment and learning is stronger if students receive external feedback on their accuracy (Sitzmann et al., 2010), this study intentionally excluded teacher feedback to minimise teacher workload. We aimed to test whether students can benefit from the diary intervention without teacher feedback.

The standardised diary contains the following components (see Appendix A).

(1) Write down the assignment, i.e., the target of self-assessment;
(2) Specify criteria for their self-assessment;

(3) Self-rate overall performance on the assignments on a six-point scale ranging from poor to excellent;
(4) Reflect on their strengths and weaknesses in completing the assignment (when doing my IH homework, I know that I am best/worst at …);
(5) Propose possible improvements (To achieve the learning goals, I could …);
(6) Report their self-assessment practices when doing assignments on a modified Self-assessment Practice Scale (SaPS) (Yan, 2018b) covering two major self-assessment actions including self-directed feedback seeking (e.g., I used the textbook to check my work.) and self-reflection (e.g., I evaluated whether my study strategies help me learn.).

Students reported spending 5–10 minutes generally to complete a diary. As recency enhances the reactivity effect, students had to complete self-assessment diaries immediately after the assignment (Bandura, 1991; Schmitz et al., 2011). Self-monitoring nearer the monitored behaviour allows reflection on more recalled information and greater accuracy of such information. The teacher gave feedback to students on their assignments, as usual. To minimise teacher workload, teachers did not give feedback on the student diaries. By excluding teachers' feedback to the diaries, we tested whether the feedback was necessary for a successful self-assessment diary intervention. If not, our intervention can be effective with little teacher effort, time, or adaptation.

Data analysis

First, we analyse the test and survey data with a *difference-in-differences, system of equations*. Then, we use an inductive approach to analyse the data collected through focus groups.

Statistical analyses of controlled experiment data

Analytic issues and statistics strategies. To help answer the above research questions with these controlled experiment data, we must address nine analytic issues. These involve data, outcomes, and explanatory variables (see Table 1).

Data issues include missing data (3% in this study). Missing data can bias results, reduce estimation efficiency or complicate data analyses. *Markov Chain Monte Carlo multiple imputation* estimates the values of the missing data, which is more effective than deletion, mean substitution, or simple imputation according to computer simulations (Peugh & Enders, 2004).

Outcome issues include survey measurement errors and multiple outcomes. To minimise survey measurement error, we use multiple measures for each construct to create an index that is more precise than any single measure via confirmatory factor analysis (Joreskog & Sorbom, 2015). We analyse whether sets of questions that appear to test similar constructs (e.g., self-efficacy, intrinsic value) reflected one or more of them through *factor analyses* (Joreskog & Sorbom, 2015). Using Monte Carlo simulation studies, Hu and Bentler (1999) showed that using a combination of the standardised root mean residual (SRMR) and one of the following indices tends to minimise Type I and Type II errors under many conditions for both factor analyses and SEMs: Tucker-Lewis Index (TLI), incremental fit index (IFI), and root mean square error of

Table 1. Statistics strategies to address each analytic difficulty.

Analytic difficulty	Statistics strategy
Data Set	
· Missing data	· Markov Chain Monte Carlo multiple imputation (Peugh & Enders, 2004)
Outcome variables	
· Measurement errors on surveys	· Factor analysis (Joreskog & Sorbom, 2015)
· Multiple outcomes	· System of equations (Kennedy, 2008)
Explanatory variables	
· False positives	· Two-stage linear step-up procedure (Benjamini et al., 2006)
· Indirect effects	· Mediation test (Sobel, 1982)
· Intervention impact differences across students	· Intervention interaction terms (Kennedy, 2008)
· Effect size differences ($\beta_1 > \beta_2$?)	· Lagrange multiplier tests (Bertsekas, 2014)
· Robustness	· Single outcome models for each outcome (Kennedy, 2008) · Analyses of original, un-estimated data

approximation (RMSEA). The following threshold values separate good, moderate, and poor fits for each measure: SRMR (<.08; <.10; good fit if less than .08; moderate fit if between .08 and .10; poor fit if greater than .10), RMSEA (<.06; <.10), TLI (>.96; >.90) and IFI (>.96; >.90). As the factor analyses' resulting factor scores are more accurate than any single variable, we used them in the regression below. (Our sample size is too small for a structural equation model (Wolf et al., 2013).) As multiple outcomes can have correlated unexplained components (*residuals*) that underestimate standard errors in ordinary least square regressions, we model them properly with a system of equations (Kennedy, 2008).

Explanatory variable issues include many hypotheses' false positives, indirect effects, intervention impact differences, effect size differences, and robustness. As testing many hypotheses increases the possibility of a false positive, we reduce their likelihood via the *two-stage linear step-up procedure*, which outperformed 13 other methods in computer simulations (Benjamini et al., 2006). To test for indirect effects, we use a mediation test (Sobel, 1982). As the impact of the intervention might differ across students systematically, we test interaction terms (e.g., intervention * pre-academic achievement test). To test whether two or more standardised regression coefficients differ significantly, we use *Lagrange multiplier tests* (Bertsekas, 2014).

Lastly, we use two variations of the core model for robustness tests. First, a misspecified equation for any outcome in a system of equations can introduce errors in otherwise correctly specified equations, so we model each outcome variable separately. Second, we repeat the analyses for the original, un-estimated data.

Explanatory model. We tested for differences among the students in the control and experimental groups after the intervention with the following explanatory model.

$$\mathbf{Post-Measure}_{yi} = \beta_y + \mathbf{e}_{yi} + \mathit{Experimental} + \mathbf{Pre-Measure}_{yi} + \mathbf{Moderators}_{yi} + \mathbf{Mediators}_{yi} \quad (1)$$

In **Post-Measures**, a post-intervention measure y (*post-academic achievement, post-self-efficacy, post-intrinsic value, post-effort regulation* or *post-self-reflection*) of student i has a grand mean intercept β_y, with unexplained component e_{yi} (*residual*).

First, we enter variables indicating the experimental group *Experimental* and the pre-intervention value of the corresponding measure y (e.g., *pre-academic*

achievement for *post-academic achievement*; **Pre-Measure**). Then, we test whether the intervention systematically showed different effects across types of students (e.g., gender) with its variable (*girl*) and its interaction term (*girl*Experimental*). We tested for intervention differences with respect to gender and its pre-measure (e.g., *pre-academic achievement*Experimental*; **Moderators**). Lastly, we test for indirect effects by adding post-measures different from the post-measure *y* as mediation variables (e.g., for *post-academic achievement*, we test the other post-measures (*post-self-efficacy, post-intrinsic value, post-effort regulation* or *post-self-reflection*; **Mediators**).

We tested whether sets of explanatory variables were significant with a nested hypothesis test (χ^2 log likelihood, Kennedy, 2008). As only omitting *significant* variables might yield *omitted variable bias*, we safely remove non-significant variables to increase precision and reduce multicollinearity (Kennedy, 2008).

Analysis of data from focus groups

The focus groups were audio-recorded and transcribed. We used an inductive approach (Thomas, 2006) to analyse and classify the participant responses into three categories including (1) students' perceptions and feelings about completing the diary; (2) the way the diary influences their learning, if any; and (3) feedback to improve the design and implementation of the diary. We analysed the data as follows. First, two researchers (the first author and a well-trained research assistant) independently read the transcript thoroughly to become familiar with the data. Second, the two researchers extracted relevant content and coded them into meaning units. Third, we compared the units determined by the two researchers for consistency, and discussed disagreements to reach a consensus agreement. Finally, the first author coded the meaning units and clustered them under the above determined categories. These interview data illustrate the results of the experimental study.

Results

Results of controlled experiment data

The factor analysis of the four constructs' respective items (self-efficacy, intrinsic value, effort regulation, self-reflection) showed a good fit of a four-factor solution to the data (SRMR = .073; TLI = .960; IFI = .965; RMSEA = .085). As expected, factor loadings had a mean of 0.80 (minimum factor loading was 0.77).

The difference-in-differences, system of equations results showed significant intervention differences on several measures, some significant moderation, and no significant mediation (see Table 2; for descriptive statistics, see Appendix Table A.1 and A.2). After the intervention, the experimental group outperformed the control group in academic achievement (13.581), controlling for their prior academic achievement, which supports H5a. As expected, students showing higher academic achievement on the pre-test scored higher on its post-test (0.806). Furthermore, students in the experimental group with higher academic achievement on the pre-test than other students benefited less from the intervention (−0.816), supporting H5b. Together, these explanatory variables accounted for nearly 29% of the differences in students' post-academic achievement (R^2 = .287).

After the intervention, the experimental group outperformed the control group in both self-efficacy (2.086) and intrinsic value (2.104), controlling for each pre-measure,

Table 2. Summaries of a difference-in-differences system of equations modelling students' academic achievement, self-efficacy, intrinsic value, effort regulation, and self-reflection in the controlled experiment, with unstandardised coefficients (and standard errors in parentheses).

	Post-intervention Dependent Variables									
Explanatory variable	Academic achievement		Self-efficacy		Intrinsic value		Effort regulation		Self-reflection	
Constant	6.896	**	2.238	***	2.588	***	3.439	***	3.114	***
	(2.350)		(0.440)		(0.412)		(0.554)		(0.402)	
Experimental	13.581	***	2.086	*	2.104	*	0.729		1.574	†
	(3.725)		(0.925)		(0.842)		(1.035)		(0.887)	
Pre-Academic achievement test	0.806	***								
	(0.170)									
Experimental * Pre-Academic achievement test	−0.816	**								
	(0.278)									
Pre-Self-efficacy			0.414	***						
			(0.120)							
Experimental * Pre-Self-efficacy			−0.417							
			(0.228)							
Pre-Intrinsic					0.302	**				
					(0.109)					
Experimental * Pre-Intrinsic					−0.389					
					(0.203)					
Pre-Effort regulation							0.181			
							(0.131)			
Experimental * Pre-Effort regulation							−0.131			
							(0.236)			
Pre-Self-reflection									0.241	*
									(0.098)	
Experimental * Pre-Self-reflection									−0.301	†
									(0.205)	
R^2	0.287		0.231		0.232		0.041		0.128	

†p <.10, *p <.05, **p <.01, ***p <.001

supporting H3 and H4. No moderators were significant. Together, these explanatory variables accounted for over 23% of the variances in students' self-efficacy and intrinsic value (R^2 = .231 and .232).

The experimental group did not significantly outperform the control group after the intervention on effort regulation or self-reflection (no support for H1 and H2), though the self-reflection difference (1.574) was nearly significant (at $p < .1$). Also, the interaction term *Experimental*Pre-self-reflection* was also nearly significant (–.301, $p < .1$).

Box plots of pre- and post-measures for each dependent variable in control and experimental groups, as presented in Figure 1, showed that outliers or extreme scores did not cause the intervention effects. All other variables that were not shown in the results were not significant (e.g., gender).

Results of focus group data

Students' responses in the focus groups were categorised into three themes including (1) perceptions and feelings about completing the diary; (2) the way the diary influences their learning; and (3) feedback to improve the design and implementation of the diary. As the main purpose of the focus groups was to explore possible explanations for the experimental study results, we briefly summarised main points of the results here, leaving

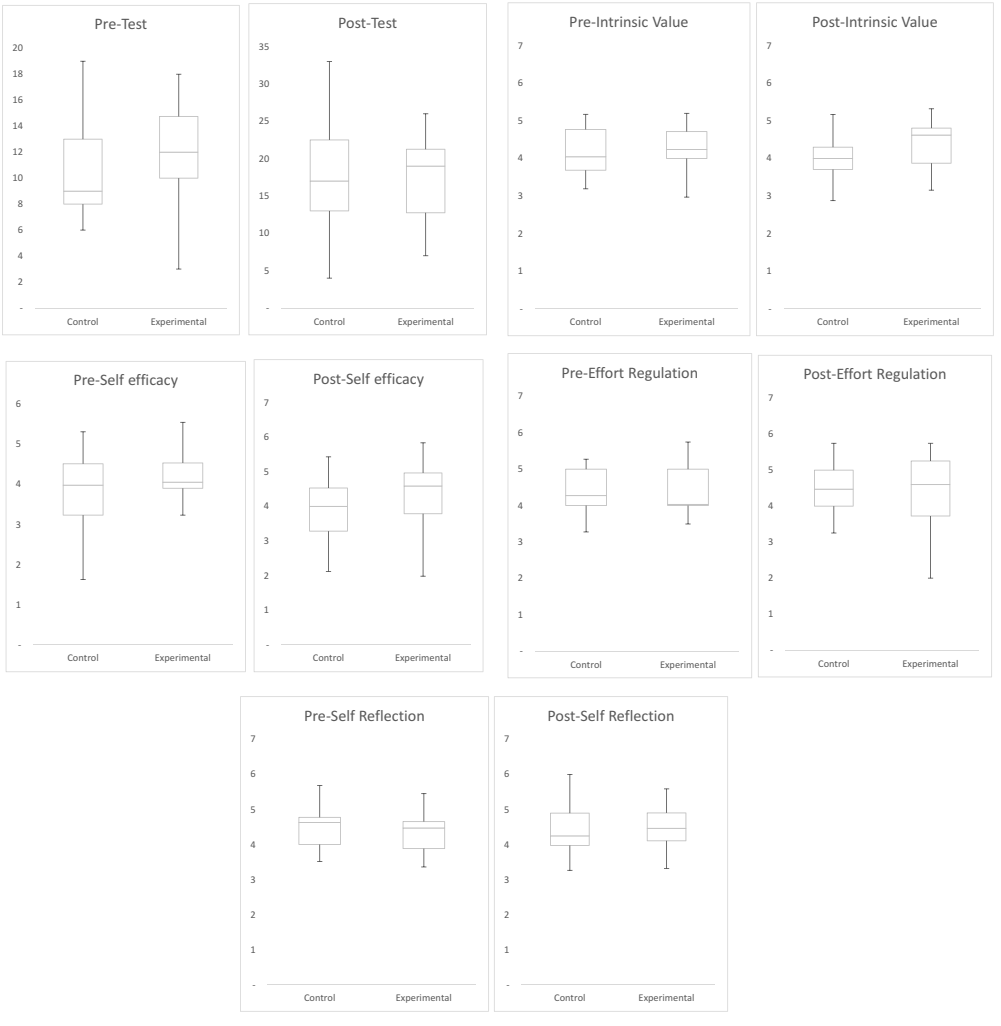

Figure 1. Box plots of pre- and post-measures for each dependent variable in control and experimental groups.

the direct excerpts and interpretation in the discussion about the results of the experimental data.

As for perceptions and feelings about completing the diary (Theme 1), all four students found this self-assessment diary differed substantially from their usual routines. Their previous self-assessment worksheets mainly asked them to self-grade their performance and identify areas of difficulties. All four students felt that completing the diary was interesting at the beginning, but one of them reported that his interest decreased after completing the diary 3–4 times. With regard to the way the diary influences their learning (Theme 2), all four students found the diary useful for their learning in some way. Three students viewed answering the questions in the diary as a reminder to check their own work after finishing the assignments. Two students reported that the diary taught them some self-assessment strategies, such as seeking family members' advice and

reviewing the parts of work that might be wrong. All four students agreed that the diary could be useful in learning subjects other than IH, but their opinions regarding which subject varied among their weaker subjects (N = 2), interesting subjects (N = 1), or difficult subjects (N = 1). Students also gave some feedback to improve the design and implementation of the diary (Theme 3). One student said the questions in the diary could vary from time to time to maintain students' interest. One suggested completing the diary online. Another said the diary could be shorter.

Discussion

The present study investigated the effects of self-assessment diaries on student learning. We hypothesised that engagement with self-assessment diaries would enhance participants' effort regulation (H1), self-reflection (H2), self-efficacy (H3), intrinsic value (H4), and academic achievement (H5a), and students with lower prior academic achievements would benefit more from the diaries (H5b). The results did not support H1 or H2, but generally supported hypotheses H3, H4, H5a, and H5b.

Effect on effort-regulation, self-reflection, self-efficacy, and intrinsic value

The results showed that self-assessment diaries did not significantly enhance students' self-regulation, notably their effort regulation or self-reflection. In contrast, self-assessment diaries could significantly enhance students' self-efficacy, consistent with past studies (Kissling & O'Donnell, 2015; Panadero et al., 2017; Ramdass & Zimmerman, 2008). Self-assessment might have increased mastery experiences, which in turn, enhanced self-efficacy. Self-assessment diaries also significantly increased students' perceived intrinsic value, consistent with Mortimer's (1998) action research, in which enhanced competence and self-awareness helped them differentiate between the difficulty levels of various learning tasks. By self-assessing, these students might have enhanced their insight into their own progression in learning (McMillan & Hearn, 2008; Mortimer, 1998). Such self-awareness might reinforce their engagement and enjoyment in learning, namely the intrinsic value they attach to the learning tasks.

Notably, self-efficacy and intrinsic value are closely related to motivational and affective processes, while self-reflection and effort regulation are closer to cognitive processes. Hence, these results are consistent with Sitzmann et al.'s (2010) meta-analysis, which found that self-assessment was more strongly linked to motivational outcomes than to cognitive learning outcomes. We speculate that the lack of self-assessment instruction and support might reduce the effectiveness of the intervention on self-reflection and effort regulation. The intentional exclusion of supporting measures to self-assessment, such as explicit rubrics and external feedback minimised the teacher's workload, but might, at the same time, reduce the instructional benefits of self-assessment. Without such supports, students might have more difficulty using effective self-reflection or effort regulation

strategies. Nevertheless, this speculation remains untested because the internal mechanisms of self-assessment are still unknown. Some studies (e.g., Yan & Brown, 2017) unpacked the behavioural process within self-assessment, but the cognitive and affective mechanisms for student self-assessment remain a black box (Andrade, 2019).

The short duration of the intervention (five weeks) might also partially explain the non-significant impacts on self-reflection and effort regulation. In a meta-analysis of intervention studies promoting self-regulated learning in primary and secondary schools, Dignath and Büttner (2008) found that longer interventions yielded larger effect sizes. Although there is no ideal duration for effective interventions, five weeks might not suffice for students to learn, practice, and master deep self-regulation strategies, such as self-reflection and effort regulation.

Another potential explanation for the lack of significant impact of self-assessment diaries on self-regulation is related to the inaccuracy in measuring self-regulation through self-reported questionnaire. Panadero et al. (2017) criticised the validity and accuracy in measurement of self-regulation using self-reported questionnaires because of (a) students' inability in recording their self-regulation actions when they are concentrating on a task; (b) the lack of standards for students to evaluate their self-regulation; and (c) a general tendency of inaccuracy in self-assessing self-regulation, especially in retrospective situations. In contrast, some qualitative methods, such as think-aloud protocols, capture self-regulation immediately, directly, and therefore likely to provide a valid account of students' actual use of self-regulation. Panadero et al. (2012) study showed that the effect of self-assessment on self-regulation was greater when measured via think-aloud protocols than via self-reported questionnaires. Future studies might consider superior approaches to collecting self-regulation data.

Effect on academic achievement and the interaction with students' prior academic achievements

The results indicated that completing self-assessment diaries significantly enhanced students' academic achievement. This result is consistent with most past studies of self-assessment interventions (e.g., Boud et al., 2013; Brown & Harris, 2013). As a student said during the focus group, the diary reminds her to self-assess (seek feedback and self-reflect) when doing assignments, 'the diary reminded me to check whether my answers in the assignment are correct … now I would ask my friends or classmates if I don't understand'. Another student said, 'Before this (completing the diaries), it is done when I finished the assignment. But now I am reminded to check what I have done well, and what should be further improved.'

Students with lower pre-test scores tended to have higher increases in their post-test scores, suggesting that students with lower prior achievement benefited more from the diary intervention than students with higher prior achievement. As pre-test scores and self-reflection were not significantly correlated (r = .05), neither low-achieving students nor high-achieving students were more likely to self-reflect, thereby eliminating two possible explanations ([a] low-achieving students did not self-reflect more to learn more, and [b] high-achieving students did not already self-reflect more, which otherwise would have rendered the diary intervention redundant). This result is consistent with the

finding reported in a study of an assessment-based instruction intervention showing that it did not benefit all students (Förster et al., 2018). Even within the same intervention (e.g., self-assessment diary), accommodating ability differences among students might require differentiated designs.

Unlike Förster et al.'s (2018) study, teachers did not provide feedback in this study. Hence, this study shows that student diaries, even without teacher feedback, aided student learning. (Note that we cannot rule out the possibility that the novelty of self-assessment diaries triggered greater engagement, resulting in (a) higher achievement overall, or (b) greater engagement by low achievers and hence still greater achievement). For example, an interviewee with low academic achievement said, 'I seldom did self-reflection after completing assignments, but, after doing these diaries, I frequently check my assignments and try to figure out which area should be improved'.

Another possible explanation is student motivation. According to Bandura (1991), the reactivity effect of self-monitoring is more salient for persons with higher motivation; those with greater desire to change their behaviour often attend to their behaviour more. In the focus groups, the low and average achievers were generally positive about completing the diaries, while the high-achieving student was bored, saying 'I felt bored after completing the diary 3–4 times as the diary items remained the same.' While the diary intervention benefited many students' learning, maintaining high-achieving students' motivation is a challenge for future diary intervention designs.

All students in our focus groups mentioned that these diaries can be useful in learning other academic subjects, but their opinions regarding which subject differed. Two students said that it might be particularly useful in their weaker subjects. For example, one student said 'I am not good at Chinese and I think this diary might help me learn Chinese … by making me think more and reflect more.' Another student said that it would be useful for interesting subjects (e.g., English language and Mathematics), and the last one said that it would be useful for difficult subjects, 'It might be useful in Chinese history since there is a huge amount of learning contents that provides more space for self-reflection.'

This study explored the mechanism(s) by which self-assessment influences academic performance. None of the explanatory variables significantly mediated the link between self-assessment and academic achievement. Notably, self-regulative learning activities (effort regulation and self-reflection) did not mediate this link in these data, contrary to the theoretical claims of past researchers (e.g., Andrade, 2010; Panadero et al., 2017; Ramdass & Zimmerman, 2008; Yan & Brown, 2017). This discrepancy might result from the short duration of the intervention and inaccuracy of measuring self-regulation with a survey, as discussed above. We also cannot exclude the possibility that other self-regulated activities not covered in this study mediated the link between self-assessment and academic achievement. Also, the positive impact of self-assessment on academic achievement did not operate via greater self-efficacy or intrinsic value. Possibly, when students self-assessed, they spent more time and effort on unmeasured productive learning activities, such as feedback seeking, which in turn improved their learning outcomes (Yan, 2020).

Implications for practice

In addition to the theoretical implications discussed in the above sections, this study has practical implications for classroom pedagogy. Identifying effective ways to help students develop productive self-assessment practice is challenging (Deneen, 2014). As most students do not learn to self-assess on their own, someone needs to teach and guide them to do so (Boud, 1995; Yan, 2020; Yan et al., 2020).

Self-assessment diaries provide a standard tool, which students can easily use, with sufficient flexibilities to cater for different teaching and contextual scenarios. For instance, teachers can easily revise the items/questions in the diaries to cater for various requirements imposed by different subjects and learning targets (e.g., fractions vs. plurals). In contrast to ad hoc self-assessment activities led by teachers in classroom, self-assessment diaries enable students to repeatedly engage in interactions with assessment criteria and feedback on their performance, which can enhance their long-term evaluative judgement (Tai et al., 2018).

Moreover, the self-assessment diary is simple and entails little teacher effort or classroom instruction time, unlike past interventions requiring extensive teacher training or teacher feedback (e.g., McDonald & Boud, 2003). As shown in these results, students who wrote their diaries without any teacher feedback still showed greater learning outcomes. Hence, students can benefit from diary interventions without substantially increasing teacher workloads, an especially important concern in contexts like Hong Kong with heavily-burdened teachers. If future studies replicate the results reported in this study, then self-assessment diaries' academic benefits would suggest their incorporation into classroom practice as a sustainable solution for aligning assessment and learning. Future studies can test whether having teachers or others facilitate students' self-assessment with external feedback, rubrics, or scripts, enhances their learning outcomes more than diaries alone.

Limitations and future studies

This study's limitations include a small sample size, a short intervention period of five weeks, only one academic subject, and a simple control group. Future studies can use larger samples. Notably, the nearly significant self-reflection results merit further study with larger samples. Moreover, the intervention period in this study was only five weeks, so future studies can test the effects of shorter versus longer interventions. As several students suggested that the diaries might be helpful in supporting learning in other academic subjects, future studies can test whether a well-designed diary template improves learning outcomes across classes and disciplines. Lastly, future studies can assign additional activities to control groups so that they parallel the experimental group more closely.

Although self-assessment diaries were linked to superior learning outcomes, the effect was greater for low-achieving students. Hence, educators can consider how to improve the intervention design for (bored) high-achieving students, such as maintaining their interest over time. For example, future studies might use fewer reminders or include self-assessment tasks that are more challenging (e.g., identify not only mistakes but also possible reasons for them) to determine if either of them increases the learning outcomes for higher achievers.

Another point worth noting is that self-assessment interventions have two major considerations. One is to enhance students' engagement in self-assessment, the other is to improve the quality of self-assessment. The current study emphasised the former to design a simple intervention without increasing teaching workload. Still, we deliberately designed the prompts questions in the diary to guide students' self-assessment process and improve their quality. Students' responses in the focus group appeared to support the general applicability of the self-assessment tasks. Hence, this study verified that engagement in self-assessment, even in a very simple way, is useful. Future studies can determine 'how best to scaffold effective formative self-assessment' (Andrade, 2019, p. 9).

Conclusion

In this study, self-assessment diaries (without teacher feedback to the completed diaries) showed the potential to enhance students' learning performance. Specifically, the benefits of self-assessment diaries include greater academic achievement, self-efficacy, and intrinsic value. However, the effect on effort regulation or self-reflection were not significant. As this self-assessment diary does not require much teaching workload or instruction time, it is likely more sustainable than interventions that are more demanding, especially in contexts where manpower and resource are limited. These findings can inform researchers and educators seeking to foster students' self-assessments to enhance their learning.

Note

1. Effect sizes can be calculated from qualitative data as long as that data is processed quantitatively. For instance, data collected with observations (qualitative data) could be used to calculate frequencies and then run quantitative analysis (e.g., *t* tests).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by a Research Cluster Fund 2017/2018 from the Education University of Hong Kong [Project No.: RG 51/2017-2018R].

ORCID

Zi Yan http://orcid.org/0000-0001-9305-884X
Ming Ming Chiu http://orcid.org/0000-0002-5721-1971
Po Yuk Ko http://orcid.org/0000-0003-0926-9754

References

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

Andrade, H. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, *4*, 87. http://doi.org/10.3389/feduc.2019.00087

Andrade, H., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *The Journal of Educational Research*, *97*(1), 21–30. http://doi.org/10.1080/00220670309596625

Andrade, H., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research*, *102*(4), 287–302. http://doi.org/10.3200/JOER.102.4.287-302

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, *50*(2), 248–287. http://doi.org/10.1016/0749-5978(91)90022-L

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, *93*(3), 491–507. http://doi.org/10.1093/biomet/93.3.491

Berry, R. (2011). Educational assessment in Mainland China, Hong Kong and Taiwan. In R. Berry & B. Adamson (Eds.), *Assessment reform in education: Policy and practice* (pp. 49–62). Springer.

Bertsekas, D. P. (2014). *Constrained optimization and Lagrange multiplier methods*. Academic.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*(1), 579–616. http://doi.org/10.1146/annurev.psych.54.101601.145030

Boud, D. (1995). *Enhancing learning through self-assessment*. Kogan Page.

Boud, D., Lawson, R., & Thompson, D. G. (2013). Does student engagement in self-assessment calibrate their judgement over time? *Assessment & Evaluation in Higher Education*, *38*(8), 941–956. http://doi.org/10.1080/02602938.2013.769198

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

Carver, C. S., & Scheier, M. F. (2000). Scaling back goals and recalibration of the affect system are processes in normal adaptive self-regulation: Understanding "response shift" phenomena. *Social Science & Medicine*, *50*(12), 1715–1722. http://dx.doi.10.1016/S0277-9536(99)00412-8

Curriculum Development Council. (2001). *Learning to learn: The way forward in curriculum development consultation document*. Author.

Curriculum Development Council. (2002). *Basic education curriculum guide: Building on strengths (Primary 1 – Secondary 3)*. Author.

Curriculum Development Council. (2014). *Basic education curriculum guide: To sustain, deepen and focus on learning to learn (Primary 1 – 6)*. Author.

Curriculum Development Council. (2017). *Secondary education curriculum guide*. Author.

Deneen, C. (2014). How good am I? Self-evaluation in an examination culture. In J. Curry & P. Hanstedt (Eds.), *Reading Hong Kong, reading ourselves* (pp. 230–249). City University of Hong Kong Press.

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition Learning*, *3*(3), 231–264. http://doi.org/10.1007/s11409-008-9029-x

Förster, N., Kawohl, E., & Souvignier, E. (2018). Short- and long-term effects of assessment-based differentiated reading instruction in general education on reading fluency and reading comprehension. *Learning and Instruction*, *56*, 98–109. http://doi.org/10.1016/j.learninstruc.2018.04.009

Hilpert, J. C., Stempien, J., van der Hoeven Kraft, K. J., & Husman, J. (2013). Evidence for the latent factor structure of the MSLQ: A new conceptualization of an established questionnaire. *SAGE Open*, *3*(4), 1–10. http://doi.org/10.1177/2158244013510305

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55. doi:10.1080/10705519909540118

Huang, H.-M., & Liaw, -S.-S. (2007). Exploring learners' self-efficacy, autonomy, and motivation toward e-learning. *Perceptual and Motor Skills*, *105*(2), 581–586. http://doi.org/10.2466/pms.105.2.581-586

Ibabe, I., & Jauregizar, J. (2010). Online self-assessment with feedback and metacognitive knowledge. *Higher Education*, *59*(2), 243–258. http://doi.org/10.1007/s10734-009-9245-6

Iida, M., Shrout, P. E., Laurenceau, J.-P., & Bolger, N. (2012). Using diary methods in psychological research. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 1. Foundations, planning, measures, and psychometrics* (pp. 277–305). American Psychological Association.

Joreskog, K., & Sorbom, D. (2015). *LISREL 9.2*. Scientific Software International.

Kennedy, P. (2008). *Guide to econometrics*. Wiley-Blackwell.

Kissling, E. M., & O'Donnell, M. E. (2015). Increasing language awareness and self-efficacy of FL students using self-assessment and the ACTFL proficiency guidelines. *Language Awareness*, *24* (4), 283–302. http://doi.org/10.1080/09658416.2015.1099659

Klug, J., Schultes, M.-T., & Spiel, C. (2018). Assessment at school – Teachers' diary-supported implementation of a training program. *Teaching and Teacher Education*, *76*, 298–308. http://doi.org/10.1016/j.tate.2017.10.014

Lee, J. C. K., Yin, H., & Zhang, Z. (2010). Adaptation and analysis of motivated strategies for learning questionnaire in the Chinese setting. *International Journal of Testing*, *10*(2), 149–165. http://doi.org/10.1080/15305050903534670

McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice*, *10*(2), 209–220. http://doi.org/10.1080/0969594032000121289

McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, *87*(1), 40–49. http://www.jstor.org/stable/42923742

Mortimer, J. (1998). Motivating student learning through facilitating independence: Self and peer assessment of reflective practice - an action research project. In S. Brown, S. Armstrong, & G. Thompson (Eds.), *Motivating students*(pp. 173–187). Kogan Page.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218. http://doi.org/10.1080/03075070600572090

Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology*, *11*(2), 551–576. http://dx.doi.10.14204/ejrep.30.12200

Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, *22*(6), 806–813. http://doi.org/10.1016/j.lindif.2012.04.007

Panadero, E., Alonso-Tapia, J., & Reche, E. (2013). Rubrics vs. self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers. *Studies in Educational Evaluation*, *39*(3), 125–132. http://doi.org/10.1016/j.stueduc.2013.04.001

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, *9*, 129–144. http://doi.org/10.1016/j.edurev.2013.01.002

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74–98. http://doi.org/10.1016/j.edurev.2017.08.004

Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: When measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research*, *60*(6), 723–735. http://doi.org/10.1080/00313831.2015.1066436

Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 133–148. https://doi.org/10.1080/0969594X.2013.877872

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research. *Review of Educational Research*, *74*(4), 525–556. http://doi.org/10.3102/00346543074004525

Pintrich, P. R., Smith, D. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. University of Michigan.

Ramdass, D., & Zimmerman, B. J. (2008). Effects of self-correction strategy training on middle school students' self-efficacy, self-evaluation, and mathematics division learning. *Journal of Advanced Academics*, *20*(1), 18–41. http://doi.org/10.4219/jaa-2008-869

Reynolds, B. M., Robles, T. F., & Repetti, R. L. (2015). Measurement reactivity and fatigue effects in daily diary research with families. *Developmental Psychology*, *52*(3), 442–456. http://doi.org/10.1037/dev0000081

Rosenthal, R., & Jacobson, L. (1992). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. (Newly expanded ed.). Crown House Pub.

Schmidt, J., Klusman, U., Lüdtke, O., Möller, J., & Kunter, M. (2017). What makes good and bad days for beginning teachers? A diary study on daily uplifts and hassles. *Contemporary Educational Psychology*, *48*, 85–97. http://doi.org/10.1016/j.cedpsych.2016.09.004

Schmitz, B., Klug, J., & Schmidt, M. (2011). Assessing self-regulated learning using diary measures with university students. In B. J. Zimmerman & D. H. Schunk (Eds.), *Educational psychology handbook series. Handbook of self-regulation of learning and performance* (pp. 251–266). Routledge/Taylor & Francis Group.

Schmitz, B., & Perels, F. (2011). Self-monitoring of self-regulation during math homework behaviour using standardized diaries. *Metacognition Learning*, *6*(3), 255–273. http://doi.org/10.1007/s11409-011-9076-6

Shapiro, E. (2014). *Behavioral assessment in school psychology*. Psychology Press.

Sitzmann, T., Ely, K., Brown, K. G., & Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education*, *9*(2), 169–191. http://doi.org/10.5465/amle.9.2.zqr169

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological methodology* (pp. 290–312). Jossey.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, *76*(3), 467–481. http://doi.org/10.1007/s10734-017-0220-3

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, *27*(2), 237–246. http://doi.org/10.1177/1098214005283748

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

Vieira, E. T., Jr., & Grantham, S. (2011). University students setting goals in the context of autonomy, self-efficacy and important goal-related task engagement. *Educational Psychology*, *31*(2), 141–156. http://doi.org/10.1080/01443410.2010.536508

Wigfield, A., & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review, 30*(1), 1–35. https://doi.org/10.1016/j.dr.12.001

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, *73*(6), 913–934. http://doi.org/10.1177/0013164413495237

Yan, Z. (2016). The self-assessment practices of Hong Kong secondary students: Findings with a new instrument. *Journal of Applied Measurement*, *17*(3), 335–353.

Yan, Z. (2018a). Student self-assessment practices: The role of gender, year level, and goal orientation. *Assessment in Education: Principles, Policy & Practice*, *25*(2), 183–199. http://doi.org/10.1080/0969594X.2016.1218324

Yan, Z. (2018b). The Self-assessment Practice Scale (SaPS) for students: Development and psychometric studies. *The Asia-Pacific Education Researcher*, *27*(2), 123–135. http://doi.org/10.1007/s40299-018-0371-8

Yan, Z. (2020). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment and Evaluation in Higher Education*, *45*(2), 224–238. http://doi.org/10.1080/02602938.2019.1629390

Yan, Z., & Brown, G. T. L. (2017). A cyclical self-assessment process: Towards a model of how students engage in self-assessment. *Assessment & Evaluation in Higher Education*, *42*(8), 1247–1262. http://doi.org/10.1080/02602938.2016.1260091

Yan, Z., & Brown, G. T. L. (2020, April 17–21).*Assessment for learning in the Hong Kong assessment reform: A case of policy borrowing. Paper accepted for 2020 AERA Annual Meeting*, San Francisco, CA, United States. (Conference canceled).

Yan, Z., Brown, G. T. L., Lee, J. C. K., & Qiu, X. L. (2020). Student self-assessment: Why do they do it? *Educational Psychology*, *40*(4), 509–532. http://doi.org/10.1080/01443410.2019.1672038

Yan, Z., & Cheng, E. C. K. (2015). Primary teachers' attitudes, intentions and practices regarding formative assessment. *Teaching and Teacher Education*, *45*, 128–136. http://doi.org/10.1016/j.tate.2014.10.002

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

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, *41* (2), 64–70. https://www.jstor.org/stable/1477457 http://doi.org/10.1207/s15430421tip4102_2

Zimmerman, B. J. (2015). Self-regulated learning: Theories, measures, and outcomes. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 541–546). Elsevier.