


## Student self-assessment practices: the role of gender, school level and goal orientation

Zi Yan


To cite this article: Zi Yan (2018) Student self-assessment practices: the role of gender, school level and goal orientation, *Assessment in Education: Principles, Policy & Practice*, 25:2, 183-199, DOI: [10.1080/0969594X.2016.1218324](https://doi.org/10.1080/0969594X.2016.1218324)

To link to this article: <https://doi.org/10.1080/0969594X.2016.1218324>

 View supplementary material [↗](#)

 Published online: 11 Aug 2016.

 Submit your article to this journal [↗](#)

 Article views: 232

 View related articles [↗](#)

 View Crossmark data [↗](#)

 Citing articles: 4 View citing articles [↗](#)



# Student self-assessment practices: the role of gender, school level and goal orientation

Zi Yan

Department of Curriculum and Instruction, The Education University of Hong Kong, Hong Kong, China

## ABSTRACT

The purpose of this study was to examine the effects of the key demographic variables of gender, school level and goal orientation on students' self-assessment practices, including self-directed feedback seeking (SDFS) and self-reflection (SR). A total of 8843 Hong Kong students were surveyed, ranging from Primary 4 to Secondary 6. The results showed that female students demonstrated higher levels of self-assessment practices than male students, including both SDFS and SR. School level was also significantly associated with secondary students' engagement in SR. Mastery goal orientation (MGO) was a significant and positive predictor of self-assessment practice including both SDFS and SR, while performance goal orientation (PGO) was a significantly negative predictor. Students with a higher level of MGO, or a lower level of PGO, were more likely to conduct self-assessment. The educational implications of the findings are discussed.

## ARTICLE HISTORY

Received 15 December 2015

Accepted 26 July 2016

## KEYWORDS

Self-assessment;  
measurement invariance;  
goal orientation;  
demographic variables

## Introduction

Self-assessment has attracted increased research interest paralleling the development of the concept of self-regulated learning. The current reforms being undertaken in major educational systems aim to develop students' self-regulated learning, which emphasises learning autonomy and control. As a fundamental process of self-regulated learning, self-assessment has the potential to improve student motivation and engagement in learning (Boud, 1995; Brown & Harris, 2013; McMillan & Hearn, 2008), autonomy in learning (Brown & Harris, 2013; Paris & Paris, 2001) and the development of metacognition (Puustinen & Pulkkinen, 2001; Zimmerman & Moylan, 2009).

Whether the potential benefits of self-assessment can be achieved in practice is still a debatable topic. The opinions of researchers and professionals on this issue have varied according to how they conceptualise it. Yan (*in press*) summarised three ways in which student self-assessment has been conceptualised in the literature: (1) self-assessment is regarded as a personal ability/skill for evaluating one's own knowledge, skills or performance; (2) self-assessment is used as a type of assessment serving summative purposes and (3) self-assessment, serving formative purposes, acts as a learning strategy or process with potential for improving students' learning. The literature indicates

**CONTACT** Zi Yan  [zyan@eduhk.hk](mailto:zyan@eduhk.hk)

 Supplemental data for this article can be accessed at <http://dx.doi.org/10.1080/0969594X.2016.1218324>.

that the educational merits of self-assessment might be greatly reduced if the first two conceptualisations are adopted. Eva and Regehr (2008) argued that it is difficult to develop self-assessment as a skill that can generate accurate assessment results. Many studies have also shown that realistic self-assessment appears to be very difficult to accomplish, due to human nature (e.g. Davis et al., 2006; Dunning, Heath, & Suls, 2004). Self-assessment has also been regarded as a less reliable indicator of student performance than other assessments (e.g. test scores, teacher ratings) (Kuncel, Crede, & Thomas, 2005; Ross, 2006). Such factors have weakened the status of self-assessment as a category of assessment, especially when serving a high-stakes summative purpose. However, as Boud (1999) argued, the formative use of self-assessment may have great educational merit for student learning by acting as a learning process or strategy, rather than as a substitute for other types of assessment.

In this regard, greater importance should be attached to the self-assessment practice serving formative purposes rather than the function of providing accurate summative evaluations. As Panadero, Brown, and Strijbos (2014, p. 11) pointed out: 'While it seems appropriate that an assessment be accurate, it might well be that there are benefits from engaging in self-assessment, even if it is unrealistic or inaccurate, as long as students overcome any unrealism or inaccuracy'. From a pedagogical perspective, what really helps students to learn is the engagement in self-assessment practices, rather than the scores or ratings generated by self-assessment. McMillan and Hearn (2008) noted that the effect of student self-assessment on academic performance is probably due to the fact that self-assessment enables students to control their own learning process and internalise the assessment criteria. In other words, a logical explanation of the positive relationship between self-assessment and learning gains is that engagement in self-assessment practices may encourage students to seek feedback and reflect on their own performances which, in turn, lead to improved learning.

From this perspective, self-assessment is defined in this study as a process during which students collect information about their own performance, reflect on and evaluate the quality of their learning process and outcomes according to explicitly stated criteria, and identify their own strengths and weaknesses (Andrade & Du, 2007; Andrade, Du, & Wang, 2008; Boud, 1995; Paris & Paris, 2001). In line with this definition, Yan (*in press*) further classified self-assessment practices into two major types of activities: self-directed feedback seeking (SDFS) and self-reflection (SR). In education literature, feedback is generally defined as specific information presented to a learner in order to facilitate reflection on his/her own performance and is usually used to inform the learner about his/her learning progress, strengths and weaknesses (Hattie & Timperley, 2007). SDFS, therefore, refers to the process whereby students take responsibility for seeking specific information in order to describe, evaluate and/or reflect their own learning (Boud, 1999; Kirby & Downs, 2007; Langendyk, 2006). When the need for self-assessment arises and feedback is not available or is insufficient, students need to gather feedback about the quality of their performance from various sources that enable them to reflect on the reasons behind their performance. It is clear that passively waiting for feedback is not an efficient way to learn, as effective feedback has to be provided in time (Maddox, Ashby, & Bohil, 2003; Park, Schmidt, Scheu, & DeShon, 2007). Thus, feedback seeking initiated by the students themselves is indispensable to self-assessment.

SR is the activity in which students reflect on and evaluate the quality of their learning process and outcomes, based on relevant feedback, and identify their own strengths and weaknesses. SR is another integral and crucial component of the self-assessment process (McMillan & Hearn, 2008; Sargeant, Mann, van der Vleuten, & Metsemakers, 2008). When feedback regarding student performance is available, students can engage in mental endeavours, exploring and elaborating their understanding of problems encountered during learning. By reflective thinking, one links new experiences and information with existing knowledge and skills, and this facilitates new learning (Boud, 1995). Critical reflection is helpful to students in identifying their strengths and weaknesses, as well as enhancing students' understanding in terms of what they have learned, what they still need to work on, and how they can reach their goals.

Some activities that are relevant to self-assessment, such as goal-setting and selecting assessment criteria, are excluded from Yan's (in press) classification of self-assessment practice. Whilst not ignoring the important role of these activities in self-assessment, it is suggested that they are different from SDFS and SR. Although the goal of self-assessment is usually determined by students, there are cases where the goal might be specified by external parties (e.g. students may be required by the teacher to self-check the quality of their assignments). The goal, explicit or implicit, will guide students' self-assessment, but the goal-setting itself might not be an activity initiated and dominated by the students themselves, such as SDFS and SR. Similarly, assessment criteria for self-assessment might come from external parties but are not outcomes of self-initiated activities; consequently, this study focuses only on SDFS and SR as the activities in self-assessment.

### ***Demographic variables influencing self-assessment***

The literature shows that key demographic variables, including gender and school level (some studies used age instead of school level as the variable), appear to influence self-assessment practice. With regard to the influence of gender on feedback-seeking behaviour, previous studies have generated mixed findings; some (e.g. Milan, Dyche, & Fletcher, 2011) found no significant association between gender and feedback-seeking behaviour. However, other studies (e.g. Van der Rijt, van den Bossche, & Segers, 2013) concluded that males seek significantly more feedback than females, while still others (e.g. Miller & Karakowsky, 2005) favoured females in terms of feedback seeking in situations where the males' traditional gender role is threatened.

Findings reported in the literature are inconsistent regarding the relationship between school level/age and feedback seeking. Van der Rijt et al. (2013) found no significant correlations between age and the frequency of feedback seeking. However, Wu, Parker, and de Jong (2014) reported that age had a negative effect on feedback seeking; older respondents were less likely to seek feedback. Similarly, Gupta, Govindarajan, and Malhotra (1999) also noted that age was a significantly negative predictor of feedback-seeking behaviour through both monitoring and enquiry modes (the reader is referred to the Instrument section for more details on monitoring and enquiry modes).

Few studies have directly investigated the effect of gender or school level/age on SR. Nevertheless, some studies have gauged the relationship between demographic variables and the constructs relevant to reflection, such as self-regulation. In general, girls were reported to use significantly more self-regulated learning strategies than boys (e.g. de Bilde,

Vansteenkiste, & Lens, 2011; Kadiravan, 2012). Regarding school level, Pintrich and Zusho (2002) found that students acquired higher levels of cognitive and metacognitive strategies as they developed. Zimmerman and Martinez-Pons (1990) also observed that 11th-grade students exceeded 8th graders who, in turn, exceeded 5th graders in terms of self-regulated learning.

### **Goal orientation and self-assessment**

Goal orientation refers to an integrated pattern of beliefs that influence the way people approach, engage and respond to situations related to achievement (Dweck, 1986; Griffin, 2005; Schunk, Pintrich, & Meece, 2008). In the education literature, goal orientation has often been conceptualised as a two-dimensional construct, consisting of mastery and performance orientation. In mastery goal orientation (MGO), the aim of completing a task is to improve understanding, knowledge and skill, and to achieve a sense of mastery. In performance goal orientation (PGO), the aim is to gain favourable judgements of one's performance, and to achieve a sense of self-worth.

Given that a major intention of student self-assessment is to evaluate where they are in relation to their learning goals, goal orientation is probably an important variable influencing students' self-assessment. Goal orientation affects the way students perceive the cost and value of feedback seeking and reflection, leading to different self-assessment practices. Some studies have linked goal orientation and feedback seeking. For example, VandeWalle and Cummings (1997) reported that individuals with a high MGO were more likely to engage in feedback-seeking behaviour because they perceived greater benefits and lower costs to be associated with this. In contrast, individuals with a high PGO were less likely to engage in feedback seeking due to the higher level of perceived cost. More recent studies (e.g. Payne, Youngcourt, & Beaubien, 2007; Porath & Bateman, 2006) provided further evidence of a positive correlation between MGO and feedback seeking. Similarly, a positive relationship between MGO and self-assessment has been found in previous studies (e.g. McMillan & Hearn, 2008; Ng & Earl, 2008). McMillan and Hearn (2008) argued that self-assessment was integral to MGO because it equipped students with a better understanding of their learning processes and of how well they had improved their knowledge and skills. They further pointed out that individuals with MGO tended to monitor and evaluate their own learning through internal activities, while individuals with PGO were more likely to rely on external parties to monitor and evaluate their learning.

Compared to feedback seeking, the relationship between goal orientation and SR has not been so well studied. However, the relationship between goal orientation and self-regulation has been identified in previous studies. For example, some (e.g. Pintrich, 1999; Pintrich, Zusho, Schiefele, & Pekrun, 2001) found that MGO appeared to be more adaptive for self-regulated learning than PGO. Students with MGO were more engaged in various cognitive and metacognitive activities.

### **The present study**

In order to orient the reader in terms of the background for the current study, a brief introduction about the assessment context in Hong Kong is now presented. There are six years of schooling at the primary level which usually starts at the age of six. On completion of

primary schooling, all primary students are allocated a secondary school place through the Secondary School Places Allocation System. There is no public examination for primary students. As for secondary schooling, a significant restructuring of the curriculum was implemented in 2009 when the 5 + 2 secondary system (with examinations at Secondary 5 and Secondary 7) was replaced by a 3 + 3 year system. In the new system, all students take a single public examination (the Hong Kong Diploma in Secondary Education, or HKDSE) at the end of Secondary 6. Another school level at which students have relatively higher assessment pressure is Secondary 3. Although there is no public examination at Secondary 3, schools usually arrange internal assessments that will determine the elective subjects students are allowed to choose at Secondary 4 and sit in the HKDSE in Secondary 6.

In line with the above theoretical and empirical background, the purpose of the present study was to examine the effect of key variables, including gender, school level and goal orientation, on students' self-assessment practices. The specific research questions that guided this study are formulated below:

Research Question 1: Are there differences regarding self-assessment practice (SDFS and SR) across gender and school levels?

Research Question 2: Is there a relationship between goal orientation (MGO and PGO) and self-assessment practice (SDFS and SR) among primary and secondary school students?

Answers to these questions will provide teachers and researchers with a better understanding of the variables related to students' self-assessment practices.

## Method

### *Participants*

Using a stratified sampling method with school bands as the stratum, 20 Hong Kong schools (10 primary and 10 secondary) were invited to participate in the survey. The school band is an indicator of the general academic ability of students, with band 1 standing for the highest ability and band 3 for the lowest. For secondary schools, Secondary 1–6 (aged 12–18) were invited. For primary schools, only Primary 4–6 students (aged 9–11) were invited due to the concern that more junior primary students might not fully understand the meaning of the items. A total of 8843 students responded to the paper-pencil questionnaire. The details of the sample are presented in Table 1.

### *Instrument*

The Self-Assessment Practices Scale (SPS) developed by Yan (*in press*) and validated for use with Hong Kong samples was used to assess students' self-assessment practices. The scale consists of two subscales: the SDFS scale and the SR scale. The seven-item SDFS scale investigates students' self-directed activities aimed at seeking feedback from different sources with regard to their own performance. In line with Ashford and Cummings's (1983) seminal study, the SPS covers two sources of feedback seeking: enquiry and monitoring. Through enquiry, students seek feedback by directly asking relevant people for it, including teachers, family members and peers, for example, 'I will ask my group members to comment on my

**Table 1.** The details of the sample.

	N	%
Female	4319	48.8
Male	4427	50.1
Primary 4 (P4)	1019	11.5
Primary 5 (P5)	1048	11.9
Primary 6 (P6)	1107	12.5
Secondary 1 (S1)	878	9.9
Secondary 2 (S2)	1028	11.6
Secondary 3 (S3)	995	11.3
Secondary 4 (S4)	1038	11.7
Secondary 5 (S5)	1104	12.5
Secondary 6 (S6)	626	7.1
Overall	8843	100

work in group activities'. By monitoring, students identify and gather useful information regarding their performance from sources in the environment, such as performance records and past exam papers, for example, 'I will check whether I have fully understood course contents by reviewing past exam papers'. The three-item SR scale gauges students' reflective thinking/activities that may clarify reasons for mistakes or identify areas for improvement, for example, 'I will identify areas for improvement according to the teacher's comments on my assignment'. As the SPS was originally developed in a secondary school context, the appropriateness of the scale for primary schools was examined by a panel, consisting of three experts in the field of educational assessment and ten primary teachers. All items, with one exception, were rated as appropriate for use with Hong Kong primary students. One item in the SDFS scale ('I often check whether I have fully mastered the course contents through assessing myself using reference books or notes') was regarded as inappropriate because using reference books or notes was not a normal practice for Hong Kong primary students. As a result, this item was removed from the scale leaving six items for the current study.

Two four-item subscales from the Motivated Strategies for Learning Questionnaire were used to assess MGO and PGO (Pintrich, Smith, Garcia, & Mckeachie, 1991). Items in the MGO scale value the meaning of improving understanding, knowledge and skill, such as 'I prefer course material that can arouse my curiosity, even if it is difficult to learn'. Items in the PGO scale emphasise the favourable judgements of one's performance; for example, 'Getting a good grade is the most satisfying thing for me'. All items in the three scales (i.e. SPS, MGO and PGO) responded to by students were measured on a six-point Likert-type scale, ranging from Strongly Disagree (1), Disagree (2), Slightly Disagree (3), Slightly Agree (4), Agree (5) to Strongly Agree (6).

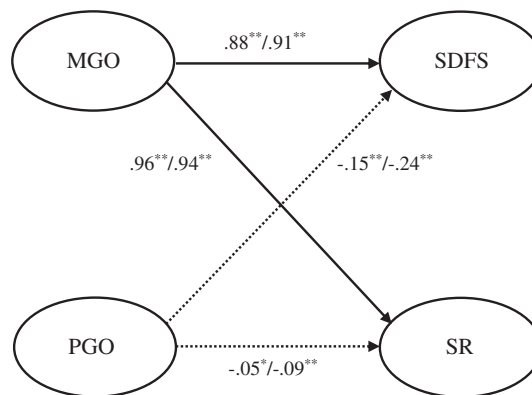
### **Data analysis**

In order to examine the differences regarding self-assessment practices across gender and school level (RQ1), the latent mean analysis, rather than the traditional approaches such as the t-test and analysis of variance, was conducted since the self-assessment practice is conceptualised as latent constructs that can only be indirectly measured through observable indicators (items). The test of latent mean differences can provide measures of the latent constructs by accounting for the random measurement error for the observed variables

(Aiken, Stein, & Bentler, 1994). It was conducted by constraining the latent means of the reference group to be zero while comparing the latent means of the reference group with other groups. Before comparing and interpreting the latent means, measurement invariance of the instruments, that is, the SDFS and SR scales, was examined across subgroups. Measurement invariance concerns whether the items and the underlying factors have the same meanings across subgroups (Borsboom, 2006). This was tested using a multi-group confirmatory factor analysis (MGCFA) approach with nested model comparisons (Vandenberg & Lance, 2000). A series of MGCFA models with increasing levels of cross-group equality constraints were tested. Firstly, the equivalence of the pattern of fixed and non-fixed parameters across groups (i.e. configural invariance) was tested. The configural invariance is to ensure the similar factor structure across groups. The metric invariance was tested by constraining factor loadings to be equivalent which implied that the same factor loadings should apply to different groups. The scalar invariance was tested by further constraining item intercepts as well as the factor loadings to be equivalent. In other words, students with the same value on the latent trait should have the same value on the observed indicator. The Chi-square difference test ( $\Delta\chi^2$ ) is commonly used for examining the measurement invariance. However,  $\Delta\chi^2$  is sensitive to sample size and a trivial difference may result in a significant value of  $\Delta\chi^2$  if the test is based on a large sample (Chen, 2007; Cheung & Rensvold, 2002). Alternatively, the difference of comparative fit index ( $\Delta CFI$ ) is recommended because it is less sensitive to both model complexity and sample size (Cheung & Rensvold, 2002; Mathyssek et al., 2013; Meade, Johnson, & Braddy, 2008). A value of  $\Delta CFI$  smaller than or equal to .01 (i.e.  $\Delta CFI \leq .01$ ) indicates measurement invariance (Cheung & Rensvold, 2002).

According to Vandenberg and Lance's (2000) suggestion, if configural, metric and scalar invariance are met between subgroups, the latent mean comparisons (across gender and school levels) could be further undertaken. The latent mean differences were assessed based on the critical ratio (CR) index. The mean difference was regarded as statistically significant if the CR was greater or equal to 1.96.

In order to investigate the relationship between goal orientation (MGO and PGO) and self-assessment practice (SDFS and SR) (RQ2), structural equation modelling (SEM) was conducted to test the hypothesised model. On the basis of theoretical framework and



**Figure 1.** Structural model with standardised coefficients.

Note: The number in front of the slash stands for primary data and behind the slash represents secondary data; \* $p < .05$ ; \*\* $p < .01$ .

empirical studies reviewed earlier, it was hypothesised that MGO would be a positive predictor of SDFS and SR, and PGO would be a negative predictor of SDFS and SR (see Figure 1).

Multiple fit indices were employed to examine the goodness of model-data fit including the comparative fit index (CFI), the normed fit index (NFI), the standardised root-mean-square error of approximation (RMSEA) and the standardised root-mean-square residual (SRMR). As a general rule, values of CFI and NFI greater than .90, and RMSEA and SRMR values less than .08 (McDonald & Ho, 2002) indicate an acceptable model fit. Given that a large sample ( $N = 8843$ ) was used in the current study, the chi-square test was not used as the index for the model-data fit due to its sensitivity to large samples. All analysis was carried out using AMOS 20.0 (Arbuckle, 2011).

## Results

### *Descriptive statistics*

The means, standard deviations and inter-correlations for the items in the four scales are presented in Table 2. Since multivariate statistical techniques have an assumption that the data follows a normal distribution, the skewness and kurtosis values were computed for each item to examine the normality of their distributions. The values for skewness and kurtosis for all items were less than the cut-off values of 3 and 7 recommended by Kline (2011) indicating that they met the univariate normality requirement of SEM sufficiently well.

### *Measurement invariance testing*

The measurement invariance of SPS across school sectors (primary vs. secondary) was first checked. The results showed that the configural invariance model had a good fit: CFI = .951, NFI = .949, RMSEA = .057, SRMR = .031. This indicates that the structural patterns were equivalent for primary and secondary students. The  $\Delta$ CFI between the configural invariance model and the metric invariance model was less than .01, indicating the metric invariance was supported. However, the  $\Delta$ CFI between the metric invariance and scalar invariance models was .03, indicating the scalar invariance was not satisfied. Therefore, the latent mean analysis and subsequent SEM was conducted for primary and secondary students separately. For each of the school sectors, measurement invariance across gender and school level was tested. The results showed that, for both primary and secondary students, the configural invariance models across gender and school levels had a good fit. The requirements of the metric and scalar invariance models were well satisfied because  $\Delta$ CFIs were all less than .01. Given that the scalar invariance across gender and school levels was supported, comparisons of latent means were subsequently conducted.

### *Comparisons of latent means*

In the current study, the comparisons of latent means were undertaken across gender (male vs. female) and school levels (P4–P6 for primary students; S1–S6 for secondary students). The model for comparing gender difference used male as the reference group by constraining the latent mean of the male group to be zero. Therefore the estimation of the latent mean of the female group stands for the mean differences between the two groups. The model for

**Table 2.** Means, standard deviations and inter-correlations for all items.

Item	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1. MGO1	4.37	1.24																	
2. MGO2	4.43	1.28	.602																
3. MGO3	4.27	1.28	.369	.398															
4. MGO4	4.25	1.27	.364	.317	.484														
5. PGO1	4.79	1.21	.298	.282	.398	.393													
6. PGO2	4.40	1.29	.233	.201	.309	.319	.598												
7. PGO3	4.62	1.26	.253	.252	.310	.255	.533	.549											
8. PGO4	4.51	1.25	.301	.275	.368	.362	.513	.499	.568										
9. SDFS1	3.55	1.33	.266	.226	.318	.308	.157	.201	.124	.204									
10. SDFS2	3.82	1.28	.338	.276	.361	.378	.231	.251	.193	.269	.522								
11. SDFS3	3.66	1.37	.297	.241	.302	.341	.212	.251	.165	.257	.458	.520							
12. SDFS4	3.74	1.27	.331	.286	.363	.363	.206	.227	.178	.253	.521	.571	.551						
13. SDFS5	3.64	1.32	.312	.231	.279	.338	.175	.204	.135	.238	.409	.411	.449	.452					
14. SDFS6	3.62	1.29	.312	.260	.329	.349	.194	.226	.158	.244	.455	.471	.472	.517	.590				
15. SR1	4.38	1.18	.342	.308	.337	.345	.291	.237	.232	.288	.292	.361	.279	.339	.285	.270			
16. SR2	4.50	1.19	.390	.358	.392	.388	.349	.284	.288	.344	.319	.431	.356	.440	.307	.322	.421		
17. SR3	4.24	1.18	.378	.352	.421	.412	.317	.280	.254	.339	.398	.467	.433	.477	.437	.433	.386	.563	

**Table 3.** Tests of latent mean differences.

	Mean	CR	<i>p</i>	ES	Mean	CR	<i>p</i>	ES
<b>Primary</b>								
<i>Across gender (the latent mean of the male group equals to zero)</i>								
Female	.283	7.284	.000	.3	.191	7.540	.000	.3
<i>Across school levels (the latent mean of P4 equals to zero)</i>								
P5	.089	1.821	.069	.1	.015	.456	.648	0
P6	.079	1.677	.094	.1	.041	1.305	.192	0
<b>Secondary</b>								
<i>Across gender (the latent mean of the male group equals to zero)</i>								
Female	.086	3.539	.000	.1	.145	7.991	.000	.3
<i>Across school levels (the latent mean of S1 equals to zero)</i>								
S2	-.029	-.644	.519	0	-.066	-2.034	.042	.1
S3	-.064	-1.431	.152	.1	-.109	-3.343	.000	.2
S4	.003	.079	.937	0	-.077	-2.417	.016	.1
S5	-.140	-3.303	.000	.2	-.090	-2.883	.004	.2
S6	-.060	-1.255	.209	.1	-.138	-3.786	.000	.2

comparing school level difference used P4 (for primary students) and S1 (for secondary students) as the reference groups by constraining the latent mean of P4/S1 to be zero. The estimations of the latent means of other school levels represent the mean differences relative to the reference group (i.e. P4 or S1). The latent mean differences across gender and school levels on the SDFS and SR are presented in Table 3.

As shown in Table 3, female primary students had a significantly higher mean than males on SDFS ( $M = .283$ ,  $CR = 7.284$ ,  $p = .000$ ) and SR ( $M = .191$ ,  $CR = 7.540$ ,  $p = .000$ ). The effect size (ES), calculated according to Hancock's (2001) suggestion, for the gender differences on both scales were .3. There was no significant effect of school level, although P5 and P6 students had a slightly higher mean on SDFS and SR than P4 students.

Female secondary students also had a significantly higher mean than males on SDFS ( $M = .086$ ,  $CR = 3.539$ ,  $p = .000$ ,  $ES = .1$ ) and SR ( $M = .145$ ,  $CR = 7.991$ ,  $p = .000$ ,  $ES = .3$ ). Regarding the impact of school level, S2, S3, S5 and S6 students had slightly lower means than S1 students on SDFS, but only the difference for S5 students reached statistical significance ( $M = -.140$ ,  $CR = -3.303$ ,  $p = .000$ ,  $ES = .2$ ). S4 students had a slightly higher mean on SDFS than S1 students but the difference was not significant. S2 to S6 students all had a significantly lower mean on SR than S1 students, with a small ES ranging from .1 to .2.

### **Relationship between goal orientation and self-assessment practice**

In order to investigate the relationship between goal orientation and self-assessment practice, a structural model was specified with two goal orientations (MGO and PGO) as predictors, and two types of self-assessment practices (SDFS and SR) as outcomes. The correlations between the latent variables are presented in Table 4. For both primary and secondary students, MGO and PGO were strongly positively correlated. SDFS and SR held positive and strong correlations with MGO and PGO.

The SEM was applied to both primary and secondary data. The results indicated that the proposed model had acceptable fit between the data and the model. For primary students,  $CFI = .915$ ,  $NFI = .909$ ,  $RMSEA = .063$ ,  $SRMR = .046$ ; for secondary students,  $CFI = .913$ ,

**Table 4.** Correlation coefficients between the latent variables.

Variable	1	2	3	4
1. MGO	–	<b>.535</b>	<b>.799</b>	<b>.934</b>
2. PGO	.696	–	<b>.321</b>	<b>.461</b>
3. SDFS	.742	.395	–	<b>.752</b>
4. SR	.882	.569	.665	–

Note. The numbers shown in bold type represent primary data, other numbers represent secondary data.

NFI = .911, RMSEA = .076, SRMR = .047. The standardised path coefficients are presented in Figure 1. For both primary and secondary students, MGO is a positive and significant predictor of SDFS (primary:  $\beta = .88, p < .01$ ; secondary:  $\beta = .91, p < .01$ ) and SR (primary:  $\beta = .96, p < .01$ ; secondary:  $\beta = .94, p < .01$ ). Students with a higher level of MGO were more likely to seek feedback regarding the quality of their performance as well as to arrive at the reasons underlying their performance by reflection. However, PGO appears to be a significantly negative predictor of SDFS (primary:  $\beta = -.15, p < .01$ ; secondary:  $\beta = -.24, p < .01$ ) and SR (primary:  $\beta = -.05, p < .05$ ; secondary:  $\beta = -.09, p < .01$ ), indicating that students with a higher level of PGO were less likely to engage in SDFS and SR.

*Note:* Further analytical output from the tests of measurement invariance and the structural equation modelling are available as supplementary material online

## Discussion

This study aimed to investigate the effect of gender, school level and goal orientation on students' self-assessment practices. The two specific research questions were addressed through the results of latent mean comparisons and SEM.

Research question 1 concerned differences regarding self-assessment practices (SDFS and SR) across gender and school levels. Before comparing latent means across subgroups, measurement invariance was first examined. It was found that scalar invariance was not supported across primary and secondary school. Therefore, the sample was split and subsequent analyses were conducted on primary and secondary data separately. The results showed that, for both primary and secondary students, the configural, metric and scalar invariance across gender and school levels were all supported.

Latent mean comparisons indicated that there were significant gender differences on self-assessment practices, including both SDFS and SR. Female students demonstrated higher levels of self-assessment practices than male students. This finding provides further evidence that females are more willing to seek feedback. Female students' higher performances on SR also echoed previous studies favouring females in terms of using self-regulated learning strategies (e.g. de Bilde et al., 2011; Kadhiravan, 2012), cognitive and metacognitive learning strategies (Bidjerano, 2005) and self-consciousness (Csank & Conway, 2004). One possible reason might be that female students reflected more on their learning experiences and were more willing to apply self-assessment practices to check and monitor learning processes and outcomes. Another possibility could be that female students were more willing to report the use of self-regulated learning strategies, including self-assessment (Bidjerano, 2005). Considering the central role self-assessment plays in self-regulated learning, the results of this study suggest that teachers should find ways to encourage and facilitate student self-assessment, especially for male students, while designing instruction

plans. The possible strategies might include providing clear assessment criteria, models of self-assessment and opportunities for self-assessment, as well as highlighting successful past experiences in self-assessment.

With regard to the effect of school level on students' self-assessment practices, the findings showed no significant difference among P4 to P6 students. For secondary students in general, senior level students had lower means on both SDFS and SR than junior level students, indicating that older students were less likely to engage in self-assessment practices. The difference among school levels was significant on SR. This is an interesting finding that seems incompatible with the results of previous studies and theories grounded in developmental psychology. Previous studies conducted in Western cultures have demonstrated that, as students grow older, they become increasingly capable of applying both basic and advanced cognitive and metacognitive strategies (Pintrich & Zusho, 2002) as well as self-regulated learning (Zimmerman & Martinez-Pons, 1990). However, studies conducted in Hong Kong students pertaining to the relevant constructs demonstrated a different picture. For example, with a sample of 404 sub-degree students in Hong Kong, Chan and Chan (2010) reported no significant difference across age in terms of the learning approach (deep or surface) adopted by students. Law, Chan, and Sachs (2008) surveyed 417 Grade 5 and 420 Grade 6 Hong Kong students on their beliefs about learning and found no grade differences with regard to constructivist beliefs. However, the same research team (Chan & Sachs, 2001) reported contradictory findings in another study with English-speaking children. Therefore, they attributed the lack of difference between Grade 5 and Grade 6 Hong Kong students to contextual factors, and speculated that public assessments faced by Grade 6 students in Hong Kong might have influenced their beliefs about learning. This view is echoed in the current study. As shown in Table 3, student performance on the SR scale did not decrease with their school levels monotonically. Students' measures remained relatively constant from P4 to P6. Among all secondary school students, S1 students had generally higher measures than senior students (S2–S6), while S3 and S6 students had the lowest measures. It is highly possible that the differences in levels demonstrated here did not reflect ability issues but, rather, were related to students' willingness to engage in self-assessment. In an examination-dominant education culture, the high-stakes internal assessment faced by S3 students and the public assessment (i.e. HKDSE) faced by S6 students were likely to place excessive pressure on students and such pressure, in turn, may have promoted 'efficient' learning strategies perceived by students, such as rote memorisation, but inhibiting a deeper, yet more time-consuming, learning approach, such as engagement in SR. This result implies that it will be more productive to develop student self-assessment in a learning environment with low pressure. How to create more space for students' deep thinking and maintain their initiatives in learning, such as self-assessment, when facing pressure caused by high-stakes assessment is an issue that should be considered and addressed by teachers, school administrators and policy-makers.

Regarding SDFS, the difference among levels compared to S1 was not significant except for S5. However, the general pattern of senior level students having lower means than junior level students still holds for SDFS. This finding is consistent with a number of previous studies (e.g. Gupta et al., 1999; Wu et al., 2014) reporting that older respondents were less inclined to seek feedback through both monitoring and enquiry modes compared with their younger peers. There may be at least two possible explanations for this. First, as pointed out

by Gupta et al. (1999), older respondents may be likely to undervalue feedback due to their resistance to change entailed by the feedback, with regard to their fixed cognitive patterns and behaviour. The second reason might be related to cultural factors because feedback seeking, as a communication-based process, should be understood and interpreted within a particular cultural context (Brutus & Greguras, 2008). Erez (2010) found that members of cultures characterised by individualistic self-concepts (e.g. the United States) were more active in seeking feedback than individuals from collectivistic and high-powered distance cultures (e.g. Hong Kong). Fear of losing face is therefore an important impediment to feedback-seeking behaviour in a Chinese cultural context (MacDonald, Sulsky, Spence, & Brown, 2013). The effect of culture may also be more significant for older people (Tony, 2013) because 'culture and life experience reinforce each other' (Swidler, 2013, p. 94). In this case, older Hong Kong students were likely to perceive more face-loss in seeking feedback compared to their younger peers and were thus more reluctant to seek it.

Concerning the relationship between goal orientation and students' self-assessment practices (RQ2), the results showed that MGO was a positive and significant predictor of student self-assessment practices, including both SDFS and SR, while PGO was a significantly negative predictor of self-assessment practices.

It is not surprising that MGO had a positive effect on self-assessment. Many studies have identified a positive correlation between MGO and feedback seeking or self-regulated learning (e.g. McMillan & Hearn, 2008; Ng & Earl, 2008; Pintrich et al., 2001; VandeWalle & Cummings, 1997). It is likely that individuals who have higher levels of MGO attach greater value and lower cost to feedback seeking and were therefore more inclined to engage in feedback seeking (VandeWalle & Cummings, 1997). Furthermore, individuals with higher levels of MGO may have been more likely to internalise the process of self-monitoring and self-evaluation because they were focused on the task at hand and paid attention to how to improve their understanding and skills (McMillan & Hearn, 2008). Therefore, students with a MGO may be more likely to continually self-assess their progress by seeking feedback and reflection, and maintaining motivation to do so during the learning process.

The results revealed that PGO had a negative effect on self-assessment, although the magnitude of the effect was not so substantial as MGO. This finding echoed previous studies that viewed PGO as being less adaptive than MGO. For example, Pintrich et al. (2001) argued that adopting performance goals or relative ability goals could have a negative effect on self-regulated learning. Svanum and Aigner (2011) reported a correlation between performance goals and less effective learning strategies with poorer performance. VandeWalle and Cummings (1997) found a negative relationship between PGO and feedback seeking. An important implication of this finding is that teachers should help students to adopt mastery goals in their learning because students with a MGO are more likely to engage in self-assessment including both SDFS and SR. Students' achievement motivation and goal orientations should be monitored before appropriate instructional strategies, such self-assessment training, is designed and implemented. It is suggested that teachers should use well-designed and rewarding systems and strategies so as to increase students' appreciation of competence and mastery, rather than relying only upon the grades to indicate ability. Productive and enjoyable learning environments could be set up by providing interesting and challenging learning tasks; self-assessment could be promoted in such learning environments.

Due to the limitations of cross-sectional designs for data collection, the current study was unable to make a strong claim on the causal relationship between goal orientations and self-assessment practices. As McMillan and Hearn (2008) pointed out, student self-assessment, if appropriately implemented, can promote a MGO, internal motivation, and a more meaningful learning experience. In their findings, self-assessment acted as a cause, while MGO was an outcome. The application of SEM is able to estimate the relationships among latent factors, but it cannot determine causality in non-experimental designs (Beatty & Brew, 2005; Kline, 2011). Therefore, a meaningful direction for future study could be the investigation of the causal relationship between goal orientation and self-assessment, using an experimental or longitudinal design.

The overall aim of the current educational reforms being undertaken by major education systems is to help students discover how to learn, and ultimately to equip them with the tools they need for lifelong learning. However, this is a challenging task, especially in an examination-oriented education system such as Hong Kong, where most students emphasise high examination scores rather than learning itself. As a necessary skill for self-regulated learning and lifelong learning, self-assessment skills, including feedback seeking and reflective thinking, should and can be explicitly taught and facilitated in the formal instruction environment. The relationship between student characteristics (gender, school level and goal orientation) and self-assessment practices, as demonstrated by this study, will shed light on the reasons that might explain the differences in self-assessment practices among students. Such understanding will be helpful in identifying the facilitators and constraints of student self-assessment in the Hong Kong context. This, in turn, may promote teachable self-assessment practices, contributing to a realistic approach that will enable students to become self-regulated and lifelong learners.

## Disclosure statement

No potential conflict of interest was reported by the author.

## Notes on contributor

*Zi Yan* is an assistant professor and associate head in the Department of Curriculum and Instruction at The Education University of Hong Kong. His main publication and research interests are in Rasch measurement and assessment in the school and higher education context with an emphasis on student self-assessment and self-regulated learning.

## References

- Aiken, L. S., Stein, J. A., & Bentler, P. M. (1994). Structural equation analyses of clinical subpopulation differences and comparative treatment outcomes: Characterizing the daily lives of drug addicts. *Journal of Consulting and Clinical Psychology, 62*, 488–499.
- Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education, 32*, 159–181.
- Andrade, H., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice, 27*, 3–13.
- Arbuckle, J. L. (2011). *IBM SPSS Amos 20 user's guide*. Armonk, NY: IBM.

- Ashford, S. J., & Cummings, L. L. (1983). Feedback as an individual resource: Personal strategies of creating information. *Organizational Behavior and Human Performance*, 32, 370–398.
- Beatty, B., & Brew, C. (2005). Measuring student sense of connectedness with school: The development of an instrument for use in secondary schools. *Leading and Managing*, 11, 103–118.
- Bidjerano, T. (2005). *Gender differences in self-regulated learning*. Paper presented at the Annual Meeting of the Northeastern Educational Research Association, Kerhonkson, NY.
- Borsboom, D. (2006). When does measurement invariance matter? *Medical Care*, 44, 176–181.
- Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.
- Boud, D. (1999). Avoiding the traps: Seeking good practice in the use of self assessment and reflection in professional courses. *Social Work Education*, 18, 121–132.
- Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 367–393). Thousand Oaks, CA: Sage.
- Brutus, S., & Greguras, G. J. (2008). Self-construals, motivation, and feedback-seeking behaviors. *International Journal of Selection and Assessment*, 16, 282–291.
- Chan, Y. M., & Chan, C. M. S. (2010). Approach to learning of sub-degree students in Hong Kong. *Research in Education*, 84, 65–78.
- Chan, C. K. K., & Sachs, J. (2001). Beliefs about learning in children's understanding of science texts. *Contemporary Educational Psychology*, 26, 192–210.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14, 464–504.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 235–255.
- Csank, P. A., & Conway, M. (2004). Engaging in self-reflection changes self-concept clarity: On differences between women and men, and low- and high-clarity individuals. *Sex Roles*, 50, 469–480.
- Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: A systematic review. *Journal of the American Medical Association*, 296, 1094–1102.
- de Bilde, J., Vansteenkiste, M., & Lens, W. (2011). Understanding the association between future time perspective and self-regulated learning through the lens of self-determination theory. *Learning and Instruction*, 21, 332–344.
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5, 69–106.
- Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist*, 41, 1040–1048.
- Erez, M. (2010). Culture and job design. *Journal of Organizational Behaviour*, 31, 389–400.
- Eva, K. W., & Regehr, G. (2008). “I’ll never play professional football” and other fallacies of self-assessment. *Journal of Continuing Education in the Health Professions*, 28, 14–19.
- Griffin, R. W. (2005). *Management* (7th ed.). Boston, MA: Houghton Mifflin.
- Gupta, A. K., Govindarajan, V., & Malhotra, A. (1999). Feedback-seeking behavior within multinational corporations. *Strategic Management Journal*, 20, 205–222.
- Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and mimic approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika*, 66, 373–388.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.
- Kadhiravan, S. (2012). Self-regulated learning of adolescents in relation to their achievement motivation. *Journal of Psychosocial Research*, 7, 211–218.
- Kirby, N. F., & Downs, C. T. (2007). Self-assessment and the disadvantaged student: Potential for encouraging self-regulated learning? *Assessment & Evaluation in Higher Education*, 32, 475–494.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.
- Kuncel, N. R., Crede, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*, 75, 63–82.
- Langendyk, V. (2006). Not knowing that they do not know: Self-assessment accuracy of third-year medical students. *Medical Education*, 40, 173–179.

- Law, Y., Chan, C. K. K., & Sachs, J. (2008). Beliefs about learning, self-regulated strategies and text comprehension among Chinese children. *British Journal of Educational Psychology*, 78, 51–73.
- MacDonald, H. A., Sulsky, L. M., Spence, J. R., & Brown, D. J. (2013). Cultural differences in the motivation to seek performance feedback: A comparative policy-capturing study. *Human Performance*, 26, 211–235.
- Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 650–662.
- Mathyssek, C. M., Olino, T. M., Hartman, C. A., Ormel, J., Verhulst, F. C., & Van Oort, F. V. (2013). Does the Revised Child Anxiety and Depression Scale (RCADS) measure anxiety symptoms consistently across adolescence? The TRAILS study. *International Journal of Methods in Psychiatric Research*, 22, 27–35.
- McDonald, R. P., & Ho, M. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7, 64–82.
- McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, 87, 40–49.
- Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, 93, 568–592.
- Milan, F. B., Dyche, L., & Fletcher, J. (2011). “How am I doing?” Teaching medical students to elicit feedback during their clerkships. *Medical Teacher*, 33, 904–910.
- Miller, D. L., & Karakowsky, L. (2005). Gender influences as an impediment to knowledge sharing: When men and women fail to seek peer feedback. *The Journal of Psychology*, 139, 101–118.
- Ng, J. R., & Earl, J. K. (2008). Accuracy in self-assessment: The role of ability, feedback, self-efficacy and goal orientation. *Australian Journal of Career Development*, 17, 39–50.
- Panadero, E., Brown, G. T. L., & Strijbos, J.-W. (2014). *The future of student self-assessment: Known unknowns and probable directions*. Paper presented at the biennial conference of the Assessment & Evaluation SIG, Madrid, Spain.
- Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist*, 36, 89–101.
- Park, G., Schmidt, A. M., Scheu, C., & DeShon, R. P. (2007). A process model of goal orientation and feedback seeking. *Human Performance*, 20, 119–145.
- Payne, S. C., Youngcourt, S. S., & Beaubien, J. M. (2007). A meta-analytic examination of the goal orientation nomological net. *Journal of Applied Psychology*, 92, 128–150.
- Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, 31, 459–470.
- Pintrich, P. R., Smith, D. A., Garcia, T., & Mckeachie, W. J. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning, University of Michigan.
- Pintrich, P. R., & Zusho, A. (2002). The development of academic self-regulation: The role of cognitive and motivational factors. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 249–284). San Diego, CA: Academic.
- Pintrich, P. R., Zusho, A., Schiefele, U., & Pekrun, R. (2001). Goal orientation and self-regulated learning in the college classroom: A cross-cultural comparison. In F. Salili, C.-Y. Chiu, & Y.-Y. Hong (Eds.), *Student motivation: The culture and context of learning* (pp. 149–169). New York, NY: Plenum.
- Porath, C. L., & Bateman, T. S. (2006). Self-regulation: From goal orientation to job performance. *Journal of Applied Psychology*, 91, 185–192.
- Puustinen, M., & Pulkkinen, L. (2001). Models of self-regulated learning: A review. *Scandinavian Journal of Educational Research*, 45, 269–286.
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment Research & Evaluation*, 11, 1–13.
- Sargeant, J., Mann, K., van der Vleuten, C., & Metsemakers, J. (2008). Directed self-assessment: Practice and feedback within a social context. *Journal of Continuing Education in the Health Profession*, 28, 47–54.

- Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2008). *Motivation in education: Theory, research and applications* (3rd ed.). Upper Saddle River, NJ: Merrill-Prentice Hall.
- Svanum, S., & Aigner, C. (2011). The influences of course effort, mastery and performance goals, grade expectancies, and earned course grades on student ratings of course satisfaction. *British Journal of Educational Psychology*, 81, 667–679.
- Swidler, A. (2013). *Talk of love: How culture matters*. Chicago, IL: University of Chicago Press.
- Tony, D. (2013). *Culture and neighborhoods*. Retrieved from [http://embed.policyreview.tv/media/documents/SE241\\_TONY\\_DUGGAN.pdf](http://embed.policyreview.tv/media/documents/SE241_TONY_DUGGAN.pdf)
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4–70.
- Van der Rijt, J., van den Bossche, P., & Segers, M. S. (2013). Understanding informal feedback seeking in the workplace: The impact of the position in the organizational hierarchy. *European Journal of Training and Development*, 37, 72–85.
- VandeWalle, D., & Cummings, L. L. (1997). A test of the influence of goal orientation on the feedback seeking process. *Journal of Applied Psychology*, 82, 390–400.
- Wu, C. H., Parker, S. K., & de Jong, J. P. (2014). Feedback seeking from peers: A positive strategy for insecurely attached team-workers. *Human Relations*, 67, 441–464.
- Yan, Z. (in press). The self-assessment practices of Hong Kong secondary students: Findings with a new instrument. *Journal of Applied Measurement*.
- Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, 82, 51–59.
- Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299–315). New York, NY: Routledge.