Routledge
Taylor & Francis Group

Assessment in Education: Principles, Policy & Practice

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

More is not always better: the nonlinear relationship between formative assessment strategies and reading achievement

Yuyang Cai, Min Yang & Juanjuan Yao

To cite this article: Yuyang Cai, Min Yang & Juanjuan Yao (2022): More is not always better: the nonlinear relationship between formative assessment strategies and reading achievement, Assessment in Education: Principles, Policy & Practice, DOI: 10.1080/0969594X.2022.2158304

To link to this article: https://doi.org/10.1080/0969594X.2022.2158304

Published online: 17 Dec 2022.

ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE
https://doi.org/10.1080/0969594X.2022.2158304

Routledge
Taylor & Francis Group

More is not always better: the nonlinear relationship between formative assessment strategies and reading achievement

Yuyang Cai[a], Min Yang[b] and Juanjuan Yao[c]

[a]Centre for Language Education and Assessment Research (CLEAR), Shanghai University of International Business and Economics, Shanghai, China; [b]Department of Curriculum and Instruction, The Education University of Hong Kong, Hong Kong SAR, China; [c]School of Languages, Shanghai University of International Business and Economics, Shanghai, China

ABSTRACT

This study investigated the relationship between formative assessment and reading achievement in Hong Kong, a Confucian Heritage Culture (CHC) society. 4,837 Hong Kong students were surveyed in a nine-item questionnaire that was used as indicator variable of formative assessment strategies. The study used multi-group structural equation modelling (MG-SEM) to examine the effects of formative assessment strategies on reading achievement across low-, medium-, and high-achievers controlling for gender and social economic status (SES) effects. The result showed that after controlling for SES and gender effects, there was significant effect of formative assessment strategies with low- and medium-reading achievers but not with high-reading achievers. Implications are drawn to inform formative assessment research and practice relevant to students' reading achievement in CHC societies and other educational contexts.

ARTICLE HISTORY
Received 22 February 2021
Accepted 7 December 2022

KEYWORDS
Formative assessment; reading assessment; reading achievement; Confucian heritage culture (CHC)

1. Introduction

For decades, the power of formative assessment in supporting students' learning and academic achievement has been promoted by schools, governments, and international organisations (e.g. Organisation for Economic Co-operation and Development or OECD) in educational reforms worldwide (Wiliam, 2018; Yan, Li, etal., 2021). Based on an accumulative body of theoretical and empirical studies, researchers have argued for the integration of formative assessment into teachers' classroom practice. Ever since its inception in the 1960s when the concept of 'formative assessment' was originated, formative assessment has continued to be an active area of educational research (Klenowski & Wyatt-Smith, 2014; Wiliam, 2018).

Drawing on the existing literature, *formative assessment* is defined as the use of formal (e.g. examinations and tests) and informal (e.g. questioning and homework) assessment data to inform teachers' instructional adjustments and student learning improvement (Duckor & Holmberg, 2017; Wiliam, 2018), which concurs with current thinking in

CONTACT Min Yang myang@eduhk.hk Department of Curriculum and Instruction, Education University of Hong Kong, 10 Lo Ping Road, Taipo, New Territories, Hong Kong SRA, China

reading assessment research (Stahl et al., 2020). Black and Wiliam (2009) theorised formative assessment as comprising *five major strategies* (our interpretation):

(a) Sharing learning goals and success criteria or quality standards (Strategy 1);
(b) Eliciting information on student understanding through discussions and tasks in classrooms (Strategy 2);
(c) Providing feedback (Strategy 3);
(d) Involving students in one another's learning (Strategy 4);
(e) Nurturing students as independent learners (Strategy 5).

There are various manifestations of formative assessment. Examples may range from teachers' elicitation of information on students' progress through moment-to-moment interactions (e.g. whole-class or group-based discussions) that facilitate high-quality learning (Rubie-Davies, 2014), through students' direct involvement in assessment process (e.g. student participation in goal-setting, self- and peer-assessment) that develop their emerging expertise of evaluative judgement (Yang et al., 2021), to more formalised formative assessment activities (e.g. follow-up activities after a summative test) that help students use more effective learning strategies to address their weaknesses and self-correct answers (Carless, 2011). For formative assessment to benefit students' learning, associated strategies should be oriented towards increasing students' lifelong learning skills such as metacognition and self-regulation (Andrade & Heritage, 2018), which reflects a developmental perspective (Greyling et al., 2020) or a growth mindset (Box, 2019) regarding student achievement and learning as emphasised by reading researchers.

In reading assessment literature, researchers have stressed the role of formative assessment in nurturing students' reading literacy development (Afflerbach et al., 2018; Shore et al., 2016). However, compared to the general formative assessment literature, there is only limited empirical research evidence relevant to the effects of formative assessment strategies on students' reading achievement (e.g. Hooley & Thorpe, 2017; Li, 2016). To address the research gap, we conducted an empirical investigation into the relationship between formative assessment and reading achievement among Hong Kong secondary school students, which drew data from the Programme for International Student Assessment (PISA, a large-scale assessment organised by the OECD).

Given the rising volume of studies on student achievement in PISA, it is somewhat surprising that very few studies examined the effects of formative assessment on reading achievement. Most PISA studies on factors influencing reading performance (e.g. gender, social-economic status or SES, school climate) have excluded formative assessment (e.g. Anaya & Zamarro, 2020; Asadullah et al., 2020; Ning et al., 2016). One exception was Li's (2016) study, which found that formative assessment was positively related with reading achievement both directly and indirectly through teacher-student relationship; however, the sample only consisted of students from the United States. Formative assessment researchers observed that teachers in Western and Eastern cultures differ in their approaches to formative assessment. While teachers in the West (e.g. the US, Australia) tend to use an extended approach which is student-driven (e.g. peer discussion, self- and peer-assessment), teachers in the East (e.g. Korea, China) are inclined to employ a restricted approach which is teacher-driven (e.g. whole class discussion, test follow-up learning activities) (Carless, 2011; Kennedy, 2016; Xiao & Yang, 2019); the

latter is regarded as a less effective form of formative assessment by researchers since it does not cultivate students' understanding of assessment criteria as a precondition for self-evaluation, reflection, and improvement (Biggs, 1998; Kennedy, 2016; Lam, 2016). Since teachers in Confucian heritage culture (CHC) societies adopt different formative assessment approaches than teachers in Western societies, a similar study as Li's (2016), which was conducted in the West would be valuable to do in the CHC society.

By focusing on students from Hong Kong, a CHC society, this study contributes to the debates around the possible reasons for CHC students' outstanding reading achievement (Frønes et al., 2020; Hopfenbeck et al., 2018). CHC societies (e.g. Hong Kong, Taiwan, Shanghai, Japan, Vietnam, and Singapore) are East- and South-East Asian societies that are traditionally influenced by Confucian culture. The notable outperformance of CHC students in large-scale assessments, such as PISA and PIRLS, has drawn international attention (Asadullah et al., 2020; Chen et al., 2020).

In part, the paradox between CHC students' outstanding performance and teacher-driven formative assessment they experience in classrooms can be attributed to the Confucian assessment tradition influenced by *keju*, the Chinese imperial examination originated in Han dynasty (206 B.C.E. to 220 C.E.). *Keju* offered candidates the opportunity to join the civil service system regardless of their backgrounds, providing them with strong incentives to work hard, which led to a longstanding examination-oriented learning culture in China and other CHC societies (Lam, 2016). On the other hand, *keju* reinforced the CHC belief that effortful learning leads to success, which is related to higher academic performance (Hau & Salili, 1990). It is also possible that the formative assessment practice of using summative tests to inform teaching (e.g. conducting test follow-up exercises) positively affected academic performance among CHC students.

Some researchers contended that the examination-oriented learning culture presented a significant barrier to Hong Kong's assessment reforms that promote formative assessment (Berry, 2008; Kennedy, 2016; Yan, Brown, 2021). Researchers also alluded to the prevalence of external examinations initiated in the name of formative assessment (e.g. the Territory-wide System Assessment or TSA as part of the centrally administered Basic Competence Assessment), which have intensified rather than elevated pressures on students to perform well in examinations (Kennedy, 2016; Lam, 2016).

Two other Confucian values are argued to influence formative assessment practices in CHC classrooms. The first is an expectation of CHC students to show respect for teachers by being attentive during class, which makes them feel reluctant to engage in classroom interactions (Thanh Pham & Renshaw, 2015). This might explain why Hong Kong/other CHC teachers frequently employ teacher-led whole class questioning rather than student-led group discussions in classrooms (Carless, 2011). The second is the expectation of CHC teachers to be caring for students, which makes them approachable for students to ask questions and elicit help after class (Xiao & Yang, 2019).

The above analysis shows that formative assessment practices among CHC teachers are aligned with the assessment tradition and cultural values in CHC societies. This points to a need to examine the effects of formative assessment strategies on Hong Kong/CHC students' reading achievement by considering the social, cultural, and educational conditions that shape their learning experience.

Apart from a lack of research evidence on the effects of formative assessment on students from CHC cultures, another gap in existing studies pertains to the treatment of

students as if they were the same (e.g. Li, 2016; Yan, King, etal., 2021). Such studies failed to examine whether students of different achievement levels might respond to teachers' formative assessment strategies in similar or different ways. The need for studying whether the potential differential effects of formative assessment do exist for high-, low, and medium-achievers is supported by Shute's (2008) meta-analysis finding that high-achieving students preferred and benefitted delayed and elaborative feedback compared with low-achieving students who preferred and benefitted from immediate and corrective response feedback. Thus, different from the common approach of treating student participants as if they were from the same achievement group, this study employed multi-group structural equation modelling (MG-SEM) to examine the effects of formative assessment strategies on reading achievement across low-, medium-, and high-achievers with gender and SES as covariates, providing a more fine-grained analysis of how/whether formative assessment impacted reading achievement among Hong Kong students.

2. Effects of formative assessment on academic achievement and barriers to teachers' implementation in classrooms

2.1. Evidence of the effects of formative assessment on academic achievement

The sustained attention to formative assessment has been attributed to seminal literature reviews and meta-analyses (Black & Wiliam, 1998, 2010; Hattie & Timperley, 2007; Hattie, 2009; N. Kingston, B. Nash, 2011) that provided evidence on the effects of teachers' *appropriate* formative assessment and feedback practices (being recognised as a core element of formative assessment) on student learning gains. There was, however, inconsistencies in the results of these existing studies.

Two studies that provided stronger evidence were by Black and Wiliam (1998) and Hattie and Timperley (2007), who reported large effect sizes of 0.4–0.7 for formative assessment innovations (Black & Wiliam, 1998; based on 196 studies) and 0.79 for feedback interventions (Hattie & Timperley, 2007; based on 196 studies) respectively. On the other hand, N. Kingston, B. Nash, 2011 reported a moderate overall effect size of 0.20, with 0.32, 0.17, and 0.09 for English language arts (ELA), mathematics, and science respectively, indicating that formative assessment was more effective with ELA; they suggested that discrepancies in results might be due to the limited number of studies (just 13 out of over 300 studies) that provided sufficient information for meta-analysis.

Notwithstanding the promises of formative assessment to positively impact on students' achievement, which are both advocated by assessment researchers and evidenced in the above-discussed existing studies, the complexities involved in policy and practice of formative assessment (Gillis et al., 2016; Yan, Li, etal., 2021) and its nature as a cultural practice (Kennedy, 2016; Smagorinsky, 2009) make it an intricate business for teachers to implement it classrooms.

In the remainder of this section, we present a critical synthesis of research literature, analyse the primary barriers to teachers' effective implementation of formative assessment, and search for solutions to overcome such barriers and thus unleash the full potential of formative assessment to support achievement and learning. In so doing, we pay particular attention to reading research while taking account of the general

literature of formative assessment. The ways in which students are socialised into values of learning and assessment as embedded in their socio-cultural backgrounds also require special attention in rendering formative assessment a valuable educational tool (Frønes et al., 2020; Klenowski & Wyatt-Smith, 2014; Pryor & Crossouard, 2008), which is highly relevant to the current study as we shall demonstrate in the analysis of research themes below.

2.2. Barriers to teachers' effective implementation of formative assessment in CHC societies

A survey of the research literature reveals three major barriers that are likely to prohibit teachers' effective adoption of formative assessment, which are connected to the ways in which CHC teachers conduct classroom assessments for formative purposes.

The first barrier is related to the misinterpretation of formative assessment as a box-ticking activity by following discrete procedures (Box, 2019), such as setting objectives and criteria without considerations for supporting students' potential for attaining learning outcomes beyond the prescription of the curriculum. The mechanistic implementation of formative assessment procedures reflects a form of conformative assessment (Torrance, 2012), which seeks to increase students' test scores through coaching and practicing, resulting in students' convergent rather than divergent learning outcomes. Such conformative assessment practices frequently observed in CHC societies, such as drilling before term-end examinations (Carless, 2011).

The second barrier is educational authorities' over-reliance on standardised examinations to maintaining schools' accountability for promoting students' achievement enhancement, which can give rise to a performative assessment culture and reduce teachers' and students' engagement with formative assessment (Zhu, 2020). This is evident in Hong Kong, where TSA is regularly conducted with key stage 1–3 students to yield assessment data that are benchmarked against basic competence standards of Chinese language, English language, and mathematics (Kennedy, 2016). Similar critiques were also put forward by researchers in reading instruction and assessment (e.g. Afflerbach, 2016; Wixson, 2017).

The third barrier is associated with cultural norms in CHC societies. Confucian values such as prioritising academic success as personal responsibilities for family and effortful making can indeed improve students' academic performance (Gillis et al., 2016). However, when such values are over-emphasised to the extent that students feel pressurised to perform well, their learning attitudes would be negatively affected, as demonstrated by Chen et al'.s findings comparing Chinese-speaking students with English-speaking Western students in international reading assessments (Chen et al., 2020). Thus, despite CHC students' high-level performance in international assessments (Asadullah et al., 2020), it seems that some of the characteristics of formative assessment (e.g. using whole class and group discussions to elicit students' understanding and give them timely feedback) might be compromised (Black, 2015).

2.3. Interactive formative reading assessment activities that support students' reading literacy development

To maximise the power of formative assessment, educators and researchers called for a shift from narrowly focusing on assessing students' cognitive skills and strategies in tests and examinations towards adopting a rich array of *interactive formative assessment activities* (Afflerbach et al., 2018; Chen et al., 2020). These activities would help address students' needs in reading literacy development (Afflerbach, 2016), which entail that reading teachers:

(a) Identify students' zone of proximal development (i.e. what they know and can do on their own and with their teachers' and peers' support), which is important for scaffolding students' learning process (Afflerbach, 2016; Greyling et al., 2020) and differentiating instruction to address individual needs (Jones et al., 2016);
(b) Target both lower-order and higher-order cognitive skills to support students' needs for developing basic skills (e.g. word recognition) as well as higher-level reading comprehension strategies (e.g. summarising main ideas) (Chen et al., 2020; Jones et al., 2016);
(c) Nurture students' metacognition and self-assessment abilities to promote positive motivation and self-efficacy among students (e.g. using checklists to self-evaluate reading tasks; asking students to set and mark informal quizzes to encourage their goal setting and understanding quality standards) (Afflerbach, 2016; Xiao & Yang, 2019);
(d) Adapt formative assessment strategies to students' characteristics in a culturally appropriate manner (e.g. using summative tests for formative purposes, such as conducting test follow-up activities and engaging students in peer learning) (Carless, 2011; Smagorinsky, 2009; Xiao & Yang, 2019)
(e) Facilitate students' free exploration of an abundance of rich, authentic reading materials to promote their agency and interests as independent learners (Smagorinsky, 2009).

To sum up the above discussion, while summative tests and examinations provide limited feedback on teaching and learning because scores are a snapshot of how much students have learned (Afflerbach, 2016), the above-mentioned diverse formative assessment activities allow teachers to develop a deep understand of students' learning progress to cater for their learning needs. In such activities, students can obtain feedback from multiple sources (teacher, self, and peers, and learning materials) to develop and improve their reading literacy.

Despite the promises of formative assessment in supporting reading literacy development, we know relatively little about the efficacy of teachers enacted formative assessment strategies in improving students' reading achievement. Moreover, emergent research evidence has tended to indirectly rather than directly examine the relationship between formative assessment and reading achievement (e.g. Chen et al., 2020) with a few exceptions (e.g. Li, 2016).

To address the research gap, this study aimed to investigate the relation of formative assessment strategies to reading achievement drawing on Hong Kong 15-year-old

students' report of their teachers' use of formative assessment during reading lessons provided in OECD Programme for International Student Assessment (PISA) 2009 data (OECD, 2009). As studies continuously showed that reading achievement was positively affected by students' socio-economic status (SES) (Bernardo et al., 2021; Chiu & McBride-Chang, 2006; Xie et al., 2022; Yan, Cai, 2021) and by student gender in favour of girls (Bernardo et al., 2021; Chiu & McBride-Chang, 2006; Nalipay et al., 2020; Xie et al., 2022; Yan, Cai, 2021), we decided to include SES and gender as covariates to control for possible confounding effects from them.

The following research questions guided the study:

(1) To what extent is formative assessment strategies related to reading achievement after controlling for student gender and SES effects?
(2) Does this relation vary across students of different achievement levels after controlling for student gender and SES effects?

3. Methods

3.1. Data and measures

The current study used OECD PISA 2009 data provided by 4,837 students (47% girls and 53% boys) from Hong Kong. Nine students were excluded from the current study as they produced no response to variables selected for the current study. There were no missing values for other cases. The dependent variable was reading achievement. OECD scaled reading score to have a mean of 500 and a standard deviation of 100 (OECD, 2010). The mean of reading for Hong Kong was 534.16 with a standard deviation of 83.06. When grouping students, there are multiple approaches as usually appeared in empirical studies. One way would be to use the +1 and −1 standard units of the interested variable. However, this approach could result into imbalanced sample sizes in different groups which might subsequently incur biased estimates in favour of the larger group. Another approach is to use quantiles that can ensure equal sample size. In our study, students were grouped into low-, medium-, and high-achievers in reading of equal sample size (three quantiles) based on the cut-off scores of 505.34 and 574.65.

Formative assessment as the indicator variable were measured using nine items in PISA 2009 student questionnaire: ST38Q01 to ST38Q09 (OECD, 2009). In a previous study using PISA 2009 data by students in the United States, Li's (2016) explored the dimensionality and confirmed the one-factor structure of the formative assessment scale across three ethnic groups. In our study, we re-checked the factorial structure using exploratory factor analysis (EFA) with the Principal Axis Factoring plus the Promax Rotation Method, and the result suggested only one factor.

These items asked students to rate the frequency of their teachers' use of formative assessment strategies on a four-point scale (from 1 to 4, representing 'Never or hardly ever' to 'in all lessons'). An example item was: 'The teacher discusses students' work, after they have finished the <reading assignment>';. See Table 1 for all items.

As Li's (2016) argued, the nine items would suffice to indicate the basic meaning of formative assessment in using data from formal and informal assessment to inform learning and teaching; however, these items did not include self- and peer-assessment

Table 1. Descriptive statistics and reliability estimates.

		Low-Achievers		Medium-Achievers		High-Achievers	
Code	Content	Mean	SD	Mean	SD	Mean	SD
ST38Q01	The teacher explains beforehand what is expected of the students	2.17	0.89	2.11	0.81	2.07	0.82
ST38Q02	The teacher checks that students are concentrating while working on the <reading assignment>	2.53	0.89	2.57	0.81	2.65	0.83
ST38Q03	The teacher discusses students' work, after they have finished the <reading assignment>	2.40	0.88	2.50	0.81	2.64	0.83
ST38Q04	The teacher tells students in advance how their work is going to be judged	2.41	0.91	2.48	0.84	2.48	0.84
ST38Q05	The teacher asks whether every student has understood how to complete the <reading assignment>	2.50	0.91	2.55	0.84	2.59	0.83
ST38Q06	The teacher marks students' work	2.89	0.94	3.06	0.85	3.14	0.85
ST38Q07	The teacher gives students the chance to ask questions about the <reading assignment>	2.64	0.89	2.76	0.79	2.83	0.79
ST38Q08	The teacher poses questions that motivate students to participate actively	2.61	0.88	2.71	0.80	2.75	0.78
ST38Q09	The teacher tells students how well they did on the <reading assignment> immediately after	2.40	0.89	2.43	0.85	2.41	0.83
Overall		2.51	0.90	2.58	0.82	2.62	0.82
Cronbach's Alpha		.92		.86		.82	

The means of reading across low to high achievers (with SD): 440.88 (54.21), 541.52(19.87), and 620.07 (35.07).
The means of SES across low to high achievers (with SD): −1.10 (0.99), −0.79(0.98), and −0.55(0.99).

that are also important formative assessment strategies (see also: Black, 2015; Duckor & Holmberg, 2017). When compared with the five major formative assessment strategies in Black and Wiliam's (2009) framework, the nine items captured the first three strategies, including: sharing clear goals and standards (Strategy 1: ST38Q01, ST38Q04, and ST38Q05); eliciting evidence on student understanding (Strategy 2: ST38Q02, ST38Q07, and ST38Q08); and giving feedback (Strategy 3: ST38Q03, ST38Q06, and ST38Q09). Two other formative assessment strategies (making students learning resources for one another; cultivating students' independence) were missing from the questionnaire.

3.2. Data analysis

This study used multi-group structural equation modelling (MG-SEM) to examine the effects of formative assessment strategies on reading achievement across low-, medium-, and high-achievers. Data analysis involved three steps: (1) conducting single-group confirmatory factor analysis (CFA) to assess the measurement validity of formative assessment (Model 1); (2) conducting multi-group CFA to assess the measurement invariance of formative assessment strategies across three groups of achievers (Model 2 to Model 4); and (3) conducting multi-group structural equation modelling to examine the effects of formative assessment on reading achievement across three levels of achievers by controlling for the covariate effects of students' gender and social economic status (SES).

Mplus 8.2 Muthén and Muthén (1998–2018) was used to conduct CFA and SEM models. Maximum Likelihood Robust (MLR) was estimator used for computation. For model evaluation, multiple indices were used: the root mean square error of approximation (RMSEA), standardised root mean square residual (SRMR), Tucker – Lewis index (TLI) and comparative fit index (CFI). It has been suggested that to support a model-data fit, a SRMR below 0.08, a RMSEA below 0.05, and a CFI or TLI above 0.95 are needed (Mueller & Hancock, 2019). A more lenient criteria for the CFI and TLI cut-off point for an acceptable model fit is above 0.90 (Browne & Cudeck, 1992; Raykov & Marcoulides, 2006).

To compare competitive models (i.e. measurement invariance models), change in CFI was referred and a decrease of .01 or less was regarded as evidence of invariance (Cheung & Rensvold, 2002). When interpreting estimates of path coefficients, we followed Hattie (2009) and take the values of .05, .15, and .24 and above as references for small, moderate and large positive effect. A small effect size of 0.05 or less is typically interpreted as trivial and should be ignored.

4. Results

4.1. Descriptive statistics and bivariate correlations

Table 1 shows the descriptive statistics and reliability estimates for formative assessment strategies across three groups of reading achievers. The overall mean of formative assessment strategies for low-achievers was 2.51 (S.D. = .90) out of 4, slightly lower

Table 2. Bivariate correlations among formative assessment strategies, reading and covariates.

	Low-Achievers			Medium-Achievers			High-Achievers		
	FA	SES	SEX	FA	SES	SEX	FA	SES	SEX
Reading	−.034	.096**	−.091**	−.022	.086**	−.049	−.035	.072**	−.037
Formative assessment (FA)		−.032	−.004		−.048	−.001		−.017	.051*
SES			.072**			.039			.046

**p<.01, *p<.05. Sex: 0=female, 1=male.

than the mean of 2.58 (S.D. = .82) for medium-achievers, which was again slightly lower than the mean of 2.61 (S.D. = .82) for the high-achievers.

The Cronbach's alphas of the formative assessment scale for the three groups were .92, .86 and .82, respectively. These values suggest all formative assessment items within each group performed consistently in measuring teachers' use of formative assessment strategies reported by students.

Table 2 presents the group-based bivariate correlations among reading, formative assessment strategies and covariates. The correlations between reading and formative assessment appeared to be negative but non-significant across all groups of achievers: r = −0.034, −.022, and −.035 (all with $p > .05$) from low- to high- achievers, respectively. SES appeared to have largest effect on reading achievement across all groups of achievers: r = .096, .086 and .072 (all with $p < .01$), for low-, medium- and high-achievers, respectively. Gender appeared to have effect on reading only with low-achievers in favour of girls (r =-.091, $p < .01$), but not with medium- or higher-level achievers. However, the results out of bivariate correlations only provide a crude overview of the effect of formative assessment strategies and covariates on reading achievement. More accurate estimates need to be obtained from multivariate analysis based on latent variable analysis.

4.2. Model fit results

A first round CFA with the nine formative assessment items did not support a good fit: RMSEA (95% C.I.) = .088 (.084, .093), SRMR=.037, CFI=.940, TLI=.919. The modification indices provided by Mplus suggested correlations between two pairs of error terms: one between ST38Q07 and ST38Q08, and the other between ST38Q08 and ST38Q09. We went back to the items and found that both ST38Q07 and ST38Q08 dealt with the use of questioning and both ST38Q08 and ST38Q09 dealt with students' performance. Therefore, we allowed the two pairs of error terms to correlate. As a result, the model produced an acceptable fit with the data: RMSEA (95% C.I.) = .068 (.063, .073), SRMR=.032, CFI=.961, TLI=.944.

Table 3. Model fit results for invariance assessment and supplementary analysis.

	χ^2	df	χ2/df	p value	RMSEA (95% C.I.)	SRMR	CFI	TLI	ΔCFI
Model 1 Configural Invariance	614.089	74	8.30	<.001	.067 (.062, .072)	.030	.968	.954	–
Model 2 Metric invariance	814.539	91	8.95	<.001	.070 (.066, .075)	.039	.958	.950	.010
Model 3 Scalar invariance	993.322	107	9.28	<.001	.072 (.068, .076)	.043	.948	.948	.010

* RMSEA= root mean square error of approximation; SRMR=standardised root mean square residual; TLI=Tucker – Lewis index; CFI= comparative fit index.

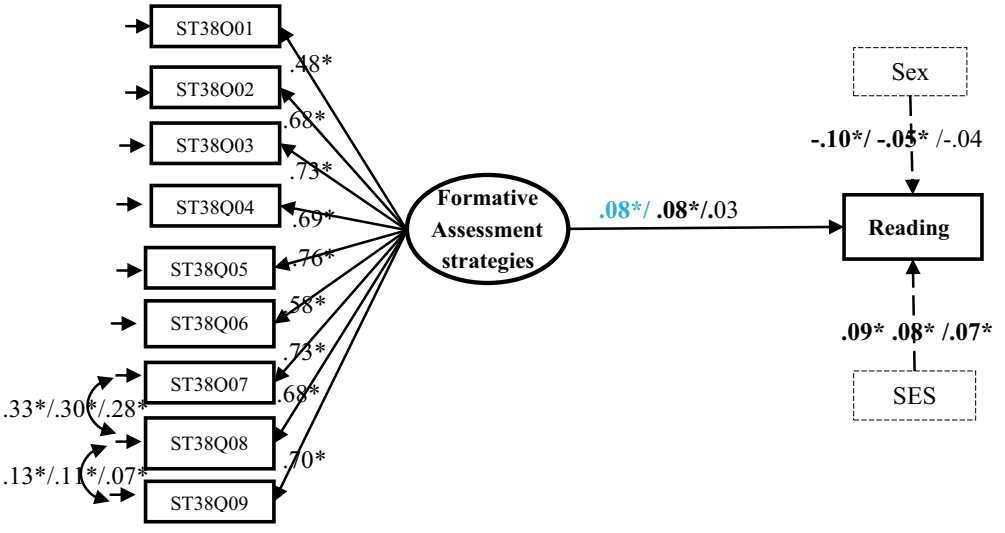

Figure 1. MG-SEM results with low, medium and high-reading achieves (estimates standardised). *P< 05 sex: 0= female, 1=male

Drawing on the modified model, we tested a series of MG-CFA models with constrains across groups to evaluate the measurement invariance of the formative assessment scale across different reading achievers. We first tested a configural MG-CFA (Model 1) with no parameters constrained, followed by a metric invariance model with factor loadings constrained (Model 2), and then a scalar invariance model with factor loadings and intercepts constrained (Model 3). The results are show in Table 3. As shown, the change in CFI was smaller than .01 at the scalar invariance level. Hence, we used Model 3 as the baseline model for parameter constraining for further MG-SEM.

To test the MG-SEM, we regressed reading achievement on the latent variable of formative assessment strategies. To control for covariate effects, we also regressed reading achievement on the covariates (i.e. gender and SES). The final model produced an acceptable fit to the data: RMSEA (with 95% C.I.) = 0.059 (.056, .063), SRMR =.040, CFI =.925, TLI=.921.

4.3. Estimates of path coefficients

Figure 1 shows the results of MG-SEM across the three reading achievement groups. The effect of formative assessment strategies on reading achievement across the low-, medium- and high-reading achievers were β = .08 (p = .005), .08 (p = .005), and .03 (p = .226), respectively, suggesting significant effect of formative assessment strategies with low- and medium-reading achievers. To compare the difference in the formative assessment effects between low- and medium- achievers, we conducted Wald Test and the results suggested non-significant difference: X^2/degree of freedom = 2.228/1, p = .136.

The effect of gender effect on reading was slightly larger with low-achievers (β = −.10, p < .01) than with medium-achievers (β = −.05, p < .01) in favour of girls, whereas this effect was not significant with high-achievers (β = −.04, p = .110).

The effects of SES on reading were significant but small with all achievers: β = .09 (p < .01) with low-achievers, β = .08 (p < .01) with medium-achievers, and β = .07 (p < .01) with high-achievers.

5. Discussion

Our study found that the relation between formative assessment strategies and reading achievement varied across the low-, medium- and high-proficiency readers. The most salient effect that formative assessment strategies had was on low- and medium-reading achievers and that formative assessment strategies did not work with high-reading achievers. These effects was smaller than the median effect of d = .20 found in a meta-analysis of the relationship between formative assessment and achievement (N. Kingston, B. Nash, 2011) or the effect size d = .33in literacy (Lee et al., 2020). There seemed to be a nonlinear relationship between formative assessment strategies and reading achievement. We offer two explanations of the result.

First, the result might be attributed to the selective inclusion of formative assessment strategies in the PISA student survey (Li, 2016). As we discussed earlier, the survey only included three categories of formative assessment strategies (sharing clear goals and standards, obtaining evidence on student understanding, and providing feedback) (Black & Wiliam, 2009). Such strategies are teacher-driven, thus reflect a restricted approach to formative assessment (Carless, 2011; Xiao & Yang, 2019), which might be more compatible with the learning needs of low- and medium-reading achievers who tend to be more reliant on teachers' guidance. On the other hand, two other categories of formative assessment strategies (making students learning resources for one another; cultivating students' independence) that are excluded from the PISA student survey reflect an extended approach to formative assessment (ibid), which might serve high achievers' learning needs and preferences better, which are more oriented to developing students' higher-order thinking skills (e.g. critical thinking) as well as metacognition and self-assessment abilities (e.g. self-evaluation and reflection) (Afflerbach, 2016; Lee et al., 2020). Due to the possible mismatch between the formative assessment strategies examined in the PISA student survey, future studies may investigate whether more inclusive formative assessment strategies will have similar or differing effects on reading achievements among students of high-, medium, and low-level language proficiency. Given the inconsistency of this result compared with what Li found among US students (that formative assessment positively affected all students' reading achievement in PISA), future research is needed to further examine the relationship between formative assessment and reading achievement for students of different achievement levels across Western and Eastern cultures.

Second, specific to high-reading achievers, it was also possible that their reading proficiency passed certain threshold, so that their proficiency was high enough for them to sail by themselves to comprehend the reading materials. This varying effect of formative assessment strategies across the low-, medium- and high- achievers displays a pattern of 'more is less', partly echoing the Island Ridge Curve (IRC) originally explaining the effect of strategy use on EFL reading performance with medium- and high-achievers (Cai, 2020; Cai, Yang, 2022; Cai, 2020, Cai, 2022). According to recent advancement in the IRC, the contributions of cognition (e.g. cognitive and metacognitive strategies, background

knowledge, critical thinking, system thinking, design thinking, etc.) or cognition-related factors (e.g. anxiety, self-concept, growth mindset, motivation regulation, etc.) to learning achievement fluctuate with the variation of their proficiency in the learning domain (e.g. reading, science, mathematics, etc.) (Cai, Chen, 2022; Cai & Kunnan, 2019, Cai, 2020; Cai, Yang, 2022; Cai, 2020; Wang et al., 2021). A common pattern of this fluctuation is smaller-larger-smaller across the low-, medium- and high-achieving groups.

Based on the finding that effects of formative assessment were only found in low- and medium-level achievers in reading assessment, reading teachers are recommended to identify the different learning needs and preferences of students with different language proficiency levels (Cai, 2020; Cai, 2020). Such information can be obtained through diagnostic assessment, observations, and teacher-student interactions, which can be used in devising formative assessment strategies that flexibly cater to students' differing learning needs and preferences (Shute, 2008).

Regarding covariate effects, girls in low- and medium-proficiency groups obtained statistically significant higher scores in reading, but the gain in reading was not significant with the high-proficiency group. Besides, SES was positively associated with reading in all three groups. The findings regarding the effects of gender (Cai, King, 2022; Chiu & McBride-Chang, 2006; Yan, Cai, 2021) and SES on reading (Bernardo et al., 2021; Chiu & McBride-Chang, 2006; Yeung et al., 2022) are consistent with the literature in general.

Gender was related to reading achievement at different levels; it was statistically significant in low- and medium-achievers. The effect of gender on reading decreased as students' reading achievement increased and the effect disappeared with the high-reading achievers. SES had a statistically significant relation to students' reading achievement and this relation seemed to be relatively stable across all groups. These results were largely consistent with findings of existing studies (OECD, 2010; Schleicher, 2019).

6. Conclusion

Cultural appropriateness is more of an important factor to be considered in explaining *why* formative assessment is not effective for some students rather than a factor to predict *whether* it is effective to some students. The analogy of this is like the influence of certain genes of the parents on their children. It's likely but not definite. The results of this study warrant some practices for implementing formative assessment strategies in low- and medium-level achievers, because methodologically we used latent variable to represent formative assessment strategies enacted by teachers, in this way measurement errors were reduced to the minimum. Relatively large sample size was used compared to classroom-based observations. More importantly, we identified the non-linear relationship between formative assessment strategies and reading achievement.

The design of this study had certain limitations that might be rectified in the future. Therefore, the results of this study might not be generalised to the implementation of formative assessment strategies in Hong Kong and other CHC societies.

First, we used students' report of teacher application of formative assessment strategies, which may not have reflected teachers' actual use of formative assessment strategies in classrooms. This restriction of students' report was deliberate in order to examine the students' perception of formative assessment strategies.

Second, previous studies indicated that teachers in CHC societies tended to adopt a restricted, teacher-centred form of formative assessment (Carless, 2011; Gillis et al., 2016), even though some teachers demonstrated exceptional formative assessment strategies (Thanh Pham & Renshaw, 2015; Xiao & Yang, 2019). Teachers' perception of formative assessment is likely to influence students' understanding and practice thus on their reading achievement (Afflerbach, 2016), it would be informative to conduct analyses on the PISA teacher questionnaire to compare differences in teachers' and students' perception of teachers' formative assessment strategies.

Third, only nine items of the students' questionnaire were included as predictors of formative assessment strategies, and although the nine items represent basic elements of formative assessment, there is not warrant for specifically generalising these results to formative assessment strategies, since the items only captured a narrow scope of formative assessment strategies, based on our critical synthesis of the general formative assessment literature (Black, 2015; Carless, 2011) as well as the reading assessment literature (Afflerbach et al., 2018; Jones et al., 2016; Li, 2016; Smagorinsky, 2009).

Notwithstanding the limitations, the re-conceptualisation of formative assessment and empirical evidence on the role of formative assessment in reading achievement among CHC students makes this study an important contribution to the field. The findings from this study should lend encouragement to recent efforts that seek to mitigate sociocultural biases in the implementation of formative assessment both in CHC classrooms and in other settings where students are in need of support in successful reading literacy development (Chen et al., 2020; Jones et al., 2016; Stahl et al., 2020).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The work was supported by The Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (Code: TP2018068) given to Professor Yuyang Cai.

Notes on contributors

Yuyang Cai, PhD is Professor of School of Languages and Founding Director of Centre for Language Education and Assessment Research (CLEAR), Shanghai University of International Business and Economics. His areas of interest include language testing, self-regulated learning, academic motivation and systems theory. He is an active user of advanced psychometric and multivariate models and also developed the statistic model of Multi-Layered Moderation Analysis (MLMA). He can be reached at sailor_cai@hotmail.com with details of his publications at ORCID.

Min Yang, PhD is an Assistant professor of Department of the Department of Curriculum and Instruction at The Education University of Hong Kong. As a teacher educator, she works with student teachers and supervise their teaching and projects in schools. Her publications and research interests are in pedagogies, and assessment and feedback strategies that support productive student learning.

Juanjuan Yao, PhD, is a lecture in EFL in School of Languages, Shanghai University of International Business and Economics. She does research in formative assessment and learning analytics. She can be reached at jjyao2002@126.com

ORCID

Yuyang Cai http://orcid.org/0000-0002-0320-4602

Data availability statement

The data that support the findings of this study are openly available in OECD PISA 2009 dataset at: http://www.oecd.org/pisa/pisaproducts/pisa2009database-downloadabledata.htm

References

Afflerbach, P. (2016). Reading assessment. *The Reading Teacher*, *69*(4), 413–419. https://doi.org/10.1002/trtr.1430

Afflerbach, P., Cho, B., Crassas, M., & Kim, J. (2018). Best practices in reading assessment. In N. K. Duke & J. R. Del Nero (Eds.), *Best practices in literacy instruction* (pp. 337–359). Guilford Press.

Anaya, L., & Zamarro, G. (2020). *The role of student effort on performance in PISA: Revisiting the gender gap in achievement*. Education Reform Faculty and Graduate Students Publications, Issue. https://scholarworks.uark.edu/edrepub/116

Andrade, H. L., & Heritage, M. (2018). *Using formative assessment to enhance learning, achievement, and academic self-regulation*. Routledge.

Asadullah, M. N., Perera, L. D. H., & Xiao, S. (2020, 09 01). Vietnam's extraordinary performance in the PISA assessment: A cultural explanation of an education paradox. *Journal of Policy Modeling*, *42*(5), 913–932. https://doi.org/10.1016/j.jpolmod.2020.02.007

Bernardo, A. B., Cai, Y., & King, R. B. (2021). Society□level social axiom moderates the association between growth mindset and achievement across cultures. *The British Journal of Educational Psychology*, *91*(4), e12411. https://doi.org/10.1111/bjep.12411

Berry, R. (2008). *Assessment for learning*. Hong Kong University Press.

Biggs, J. (1998). Assessment and classroom learning: A role for summative assessment? *Assessment in Education: Principles, Policy & Practice*, *5*(1), 103–110. https://doi.org/10.1080/0969595980050106

Black, P. Formative assessment – an optimistic but incomplete vision. (2015, 01 02). *Assessment in Education: Principles, Policy & Practice*, *22*(1), 161–177. https://doi.org/10.1080/0969594X.2014.999643

Black, P., & Wiliam, D. Assessment and classroom learning. (1998, 03 01). *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (Formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *92*(1), 81–90. https://doi.org/10.1177/003172171009200119

Box, C. (2019). *Formative assessment in United States classrooms*. Springer.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230–258. https://doi.org/10.1177/0049124192021002005

Cai, Y. (2020). *Examining the interaction among components of English for specific purposes ability in reading: The triple-decker model*. Peter Lang.

Cai, Y. (2022). Island Ridge Curve: An invisible hand behind self-regulated learning. *SSRL SIG Newsletter (Spring 2022)*, *41*, 10–11. https://ssrlsite.files.wordpress.com/2022/05/ssrl_spring_2022_newsletter_final.pdf

Cai, Y., & Chen, H. (2022). The fluctuating effect of thinking on language performance: New evidence for the Island Ridge Curve. *Language assessment quarterly*, *19*(5), 465–479. https://doi.org/10.1080/15434303.2022.2080553

Cai, Y., King, R., & McInerney, D. M. (2022). The concurrent trajectories of utility value, metacognitive strategy use, and English achievement: A multivariate growth modeling analysis. *Journal of Experimental Education*, 1–22. https://doi.org/10.1080/00220973.2022.2053496

Cai, Y., & Kunnan, A. J. (2019). Detecting the language thresholds of the effect of background knowledge on a Language for Specific Purposes reading performance: A case of the island ridge curve. *Journal of English for Academic Purposes*, *42*, 100795. https://doi.org/10.1016/j.jeap.2019.100795

Cai, Y., & Kunnan, A. J. (2020). Mapping the fluctuating effect of strategy use ability on English reading performance for nursing students: A multi-layered moderation analysis approach. *Language Testing*, *37*(2), 280–304. https://doi.org/10.1177/0265532219893384

Cai, Y., & Yang, Y. (2022). The fluid relation between reading strategies and mathematics learning: A perspective of the Island Ridge Curve. *Learning and Individual Differences*, *98*, 102180. https://doi.org/10.1016/j.lindif.2022.102180

Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-Heritage settings*. Routledge.

Chen, J., Zhang, Y., & Hu, J. (2020). Synergistic effects of instruction and affect factors on high- and low-ability disparities in elementary students' reading literacy. *Reading and Writing*, *34*(1), 199–230. https://doi.org/10.1007/s11145-020-10070-0

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chiu, M. M., & McBride Chang, C. (2006). Gender, context, and reading: A comparison of students in 43 countries. *Scientific Studies of Reading*, *10*(4), 331–362. https://doi.org/10.1207/s1532799xssr1004_1

Duckor, B., & Holmberg, C. (2017). *Mastering formative assessment moves: 7 high-leverage practices to advance student learning*. ASCD.

Frønes, T. S., Rasmusson, M., & Bremholm, J. (2020). Equity and diversity in reading comprehension—a case study of PISA 2000–2018. In T. S. Frønes, A. Pettersen, J. Radišić, & N. Buchholtz (Eds.), *Equity, equality and diversity in the Nordic model of education* (pp. 305–335). Springer International Publishing. https://doi.org/10.1007/978-3-030-61648-9_12

Gillis, S., Polesel, J., & Wu, M. PISA Data: Raising concerns with its use in policy settings. (2016, 03 01). *The Australian Educational Researcher*, *43*(1), 131–146. https://doi.org/10.1007/s13384-015-0183-2

Greyling, W., Ahmad, F., & Wallace, N. (2020). *Reading assessment as developmental tracking: A Vygotskyan perspective*. In H. (30 June ed.).

Hattie, J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. Taylor & Francis.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hau, K. T., & Salili, F. (1990, 01 01). Examination result attribution, Expectancy and achievement goals among Chinese students in Hong Kong. *Educational Studies*, *16*(1), 17–31. https://doi.org/10.1080/0305569900160102

Hooley, D. S., & Thorpe, J. (2017). The effects of formative reading assessments closely linked to classroom texts on high school reading comprehension. *Educational Technology Research and Development*, *65*(5), 1215–1238. https://doi.org/10.1007/s11423-017-9514-5

Hopfenbeck, T. N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J., & Baird, J. -A. Lessons learned from PISA: A systematic review of peer-reviewed articles on the programme for international

student assessment. (2018, 05 04). *Scandinavian Journal of Educational Research*, *62*(3), 333–353. https://doi.org/10.1080/00313831.2016.1258726

Jones, J. S., Conradi, K., & Amendum, S. J. (2016). Matching interventions to reading needs: A case for differentiation. *The Reading Teacher*, *70*(3), 307–316. https://doi.org/10.1002/trtr.1513

Kennedy, K. J. (2016). Exploring the influence of culture on assessment: The case of teachers' conceptions of assessment in Confucian-heritage societies. In G. Brown & L. Harris (Eds.), *Handbook of human factors and social conditions of assessment* (pp. 404–419). Routledge.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta□analysis and a call for research. *Educational Measurement: Issues and Practice*, *30*(4), 28–37. https://doi.org/10.1111/j.1745-3992.2011.00220.x

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, *30*(4), 28–37. https://doi.org/10.1111/j.1745-3992.2011.00220.x

Klenowski, V., & Wyatt-Smith, C. (2014). *Assessment for education: Standards, judgement and moderation*. Sage.

Lam, R. (2016). Implementing assessment for learning in a Confucian context: The case of Hong Kong 2004–14. *The Sage Handbook of Curriculum, Pedagogy and Assessment*, *2*, 756–771. https://doi.org/10.4135/9781473921405

Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, *33*(2), 1–17. https://doi.org/10.1080/08957347.2020.1732383

Li, H. (2016). How is formative assessment related to students' reading achievement? Findings from PISA 2009. *Assessment in Education: Principles, Policy & Practice*, *23*(4), 473–494. https://doi.org/10.1080/0969594X.2016.1139543

Mueller, R. O., & Hancock, G. R. (2019). Structural equation modeling. In G. R. Hancock, L. M. Stapleton, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (2nd ed., pp. 445–456). Routledge.

Muthén, L. K., & Muthén, B. Q. (19982018). Mplus 8.2 [Computer software]. Muthén & Muthén.

Nalipay, M. J. N., Cai, Y., & King, R. B. (2020). Why do girls do better in reading than boys? How parental emotional contagion explains gender differences in reading achievement. *Psychology in the Schools*, *57*(2), 310–319. https://doi.org/10.1002/pits.22330

Ning, B., Van Damme, J., Gielen, S., Vanlaar, G., & Van den Noortgate, W. (2016, 09 02). What makes the difference in reading achievement? Comparisons between Finland and Shanghai. *Scandinavian Journal of Educational Research*, *60*(5), 515–537. https://doi.org/10.1080/00313831.2015.1062413

OECD. (2009). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science*. OECD Publishing. https://doi.org/10.1787/9789264062658-en

OECD. (2010). *PISA 2009 results: What students know and can do: Student performance in reading, mathematics and science (volume I)*. OECD Publishing. https://doi.org/10.1787/9789264091450-en

Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment [Article]. *Oxford Review of Education*, *34*(1), 1–20. https://doi.org/10.1080/03054980701476386

Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling* (2nd ed.). Taylor and Francis Group. https://doi.org/10.4324/9780203930687

Rubie Davies, C. (2014). *Becoming a high expectation teacher: Raising the bar*. Routledge. https://doi.org/10.4324/9781315761251

Schleicher, A. (2019). *PISA 2018: Insights and Interpretations*. OECD Publishing. https://doi.org/10.1016/j.system.2019.01.004

Shore, J. R., Wolf, M. -K., & Heritage, M. (2016). A case study of formative assessment to support teaching of reading comprehension for English learners. *Journal of Educational Research and Innovation*, *5*(2), 4. https://digscholarship.unco.edu/cgi/viewcontent.cgi?article=1029&context=jeri

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Smagorinsky, P. (2009). The cultural practice of reading and the standardized assessment of reading instruction: When incommensurate worlds collide. *Educational Researcher*, *38*(7), 522–527. https://doi.org/10.3102/0013189x09347583

Stahl, K. A. D., Flanigan, K., & McKenna, M. C. (2020). *Assessment for reading instruction*. Guilford Publications.

Thanh Pham, T. H., & Renshaw, P. (2015). Formative assessment in Confucian Heritage Culture classrooms: Activity theory analysis of tensions, contradictions and hybrid practices. *Assessment & Evaluation in Higher Education*, *40*(1), 45–59. https://doi.org/10.1080/02602938.2014.886325

Torrance, H. (2012). Formative assessment at the crossroads: Conformative, deformative and transformative assessment. *Oxford Review of Education*, *38*(3), 323–342. https://doi.org/10.1080/03054985.2012.689693

Wang, C., Cai, Y., Zhao, M., & You, X. (2021). Disentangling the relation between motivation regulation strategy and writing performance: A perspective of the Island Ridge Curve. *Foreign Languages World (Chinese)*, *204*(3), xxx. https://www.cnki.com.cn/Article/CJFDTotal-WYJY202103007.htm

Wiliam, D. Assessment for learning: Meeting the challenge of implementation. (2018, 11 02). *Assessment in Education: Principles, Policy & Practice*, *25*(6), 682–685. https://doi.org/10.1080/0969594X.2017.1401526

Wixson, K. K. (2017). An interactive view of reading comprehension: Implications for assessment. *Language, Speech, and Hearing Services in Schools*, *48*(2), 77–83. https://doi.org/10.1044/2017_LSHSS-16-0030

Xiao, Y. Y., & Yang, M. (2019). Formative assessment and self-regulated learning: How formative assessment supports students' self-regulation in English language learning. *System*, *81*, 39–49. https://doi.org/10.1016/j.system.2019.01.004

Xie, Q., King, R., & Cai, Y. (2022). Emotional contagion: A cross-cultural exploration of how teachers' enjoyment facilitates achievement via students' enjoyment. *Current Psychology*. https://doi.org/10.1007/s12144-022-02878-6

Yan, Z., & Brown, G. T. L. (2021, 03 01). Assessment for learning in the Hong Kong assessment reform: A case of policy borrowing. *Studies in Educational Evaluation*, *68*, 100985. https://doi.org/10.1016/j.stueduc.2021.100985

Yan, J., & Cai, Y. (2021). Teachers' Instruction of Reading Strategies and Primary School Students' Reading Literacy: An Approach of Multilevel Structural Equation Modelling. *Reading & Writing Quarterly*, *38*(2), 1–17. https://doi.org/10.1080/10573569.2021.1923100

Yang, M., Yang, L., & Song, B. R. (2021). Interplay between students' perceived utility, accountability, self-efficacy and social awareness when engaged withr peer feedback: A qualitative interpretation. In Z. Yan & L. Yang (Eds.), *Assessment as Learning Maximising Opportunities for Student Learning and Achievement* (pp. 79–97). Springer.

Yan, Z., King, R. B., & Haw, J. Y. (2021). Formative assessment, growth mindset, and achievement: Examining their relations in the East and the West. *Assessment in Education: Principles, Policy & Practice*, *28*(5–6), 676–702. https://doi.org/10.1080/0969594x.2021.1988510

Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L., & Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*, *28*(3), 1–33. https://doi.org/10.1080/0969594X.2021.1884042

Yeung, S. S., King, R. B., Nalipay, M. J. N., & Cai, Y. (2022). Exploring the interplay between socioeconomic status and reading achievement: An expectancy□value perspective. *The British Journal of Educational Psychology*, *92*(3), e12495. https://doi.org/10.1111/bjep.12495

Zhu, G. (2020). A prism of the educational utopia: The East Asian Educational Model, reference society, and reciprocal learning. *Discourse: Studies in the Cultural Politics of Education*, 1–15. https://doi.org/10.1080/01596306.2020.1714547