Check for updates

Problematising formative assessment in an undeveloped region of China: voices from practitioners

Qiuxian Chen[1] · Jiajin Zhang[1] · Liang Li[1]

Abstract

Despite the widely acknowledged pro-learning function of formative assessment and its wide adoption around the globe, the gaps between policy intention, interpretation and implementation remain a problem to be solved. While this problem is noted universally, it could be particularly serious in China, where Confucian Heritage Culture is deeply ingrained and education development is not quite balanced. This study, via interview data with English teachers and deans from eight universities in an undeveloped region of the Mid-western China, explores the overall environment for a formative assessment initiative that is currently in place. Data analysis reveals multiple issues, such as insufficient support, improper dissemination and ineffective training at the meso-level and the instructors' limited assessment ability, large class sizes and student's resistance at the micro-level. A conclusion is thus drawn that the overall environment in this region is by no means favourable for the effective implementation of formative assessment, and implications are derived for better realisation of assessment innovations in this and other undeveloped regions of China.

✉ Qiuxian Chen
chenqx@sxu.edu.cn

Jiajin Zhang
zhangjiajin1966@sxu.edu.cn

Liang Li
liliang@sxu.edu.cn

[1] Foreign Language School, Shanxi University, Xiaodian District, 92# Wucheng Road, Taiyuan 030006, Shanxi, China

1 Introduction

While the driving and shaping power of assessment on curriculum and teaching/ learning remains strong, this century has seen the scenario of educational assessment undertaking a huge change. Indeed, with formative assessment's[1] enabling potentials such as enhancing learning outcome, students' meta-cognitive abilities and participation (Black & Wiliam, 1998; Earl, 2012; Wanner & Palmer, 2018) being acknowledged, it has secured a legitimate position in many educational contexts around the globe (Berry & Adamson, 2011). Despite the warm embrace of formative assessment at the policy level, the realisation of its much claimed promises in practice has proved to be far from satisfactory (Laveault & Allal, 2016). As Black (2015 p. 161) sighed after a comprehensive review of the formative assessment initiatives and its implementation across nine countries; so far, it is still "an optimistic but incomplete vision".

Investigations into formative assessment and its implementation have revealed multiple reasons explaining this result. For example, practical constraints such as insufficient teacher training, time restrictions and large class size are widely noted (Torrance & Pryor, 2001). Similarly well documented are technical issues including teachers' mechanical uptake of procedures rather than going deep into its "spirits" (Marshall & Drummond, 2006), misconception of formative assessment philosophy as that for measurement (Hargreaves, 2005) or misuse of assessment information for summative purposes (Hume & Coll, 2009; Klenowski, 2009). Powerful hamstrung from assessment habits and mindsets (CMEC, 2005) or the system which prioritises accountability, standards and summative results (Deluca et al., 2012; Shute, 2008) and misalignment with learning theories (Baird et al., 2017) are evidenced as well. These problems in the Anglophone context are mostly attributed to the fact that the key dimensions of the change process (such as dissemination, agency, professional learning and impact) or the needs of the key communities involved are not taken full account (Assessment Reform Group[ARG], 2002). Comparatively, the implementation of formative assessment in contexts which have borrowed it across borders and cultures could be more challenging because "what works in one culture will work in another" (ARG, 2009 p.7). That is, a compatibility issue might arise due to different politics, policies and cultures in the situated contexts (Black & Wiliam, 2005; Flórez, 2014). Formative assessment initiatives in this kind of settings need to handle with extra hindrances of contextual roots, which otherwise would lead these innovations to a failure (Nguyen & Khairani, 2017; Pham & Renshaw, 2015). This study is an attempt to explore the implementing situations of a formative assessment initiative in an unprivileged locality of China. With a focus on the constrains and barriers in particular, it hopes to add to the nuanced understanding of

---

[1] We are aware of the differences between formative assessment and assessment for learning (AfL). Yet, since Black and Wiliam, two top experts from Assessment Reform Group (ARG) mostly use the two terms interchangeably, we chooses to use "formative assessment", and keeps "assessment for learning" only when original texts apply.

the complexities involved in the formative assessment as a borrowed initiative from a Chinese perspective.

2 Formative assessment problematised in the Chinese context

Soon after the turn of the century, China, like many other settings, has introduced formative assessment to its education of various levels via updated syllabi. The English as a foreign language (EFL) area, for instance, has witnessed formative assessment to be included in *Nine-year Compulsory Education English Curriculum Standards* (CMoE, 2011), *Full-time High School English Curriculum Requirements* (CMoE, 2003) and *College English Curriculum Requirements (CECR)* (CMoE, 2007). Over a decade's efforts to translate these initiatives to practice has revealed constraints from the Chinese educational system, history and ideologies.

Centralisation, stratification and selection are the main features of the Chinese educational system (Wang, 1996). The system is hierarchically structured with the Chinese Ministry of Education (CMoE) as the central policy-making and coordinating body at the top, and administrative organisations at regional and school levels responsible for implementation. Educational activities within the whole education system are mandated through national education policies, national curricula and teaching syllabi, appointed or recommended textbooks, and above all, the large-scale external examinations by a top-down approach (Wang, 1996). The past decade has seen policy endeavors to innovate the system by empowering authorities of local levels to interpret guidelines into an implementable curriculum that fits within a local context (OECD, 2016). The structure, however, remains largely untouched.

The system is stratified in that schools are streamed into a key and none-key structure at every level, mostly due to the limited resources and opportunities provided. More recently, with increased financial inputs and opportunities, policy efforts have been made to blur this division for the sake of equity (Wang, 2005). Some, however, claim that this division has not left, but merely faded (Zhang, 2017). The 211, 985 and the most recent Double-First-Class Program,[2] for example, might as well be an alternative to key/non-key stratification of universities since they receive the lions' share of government funding and resources. This stratification is seen as a major cause for the imbalanced educational development in China (Hu, 2003).

Selection and accountability functions of assessment are highlighted in the system, with large-scale external examinations such as *the Senior High School Entrance Examination (Zhongkao)* and *National College Entrance Examination (Gaokao)* as gate-keeping and screening devices for next and much narrower layer of the educational ladder. The system is hence "a steep pyramid" (Wang, 1996 p. 76) and highly competitive (Cheng & Curtis, 2009). Enrollment expansion for higher education over the last two decades seems to have made the pyramid less steep; insiders,

---

[2] 211, 985 and Double First Class are three programs that the Chinese government launched in late 1990s, early 2000s and 2010s respectively with enormous financial and policy input for the purpose of uplifting its higher education quality to the world class.

however, have argued that these changes are merely a shift from a focus on tertiary education opportunities to limited vacancies in elite universities (Guo & Wu, 2008; Zhang, 2017; Zhou, 2019). In other words, competition in the Chinese education system is as severe as ever.

This system and its mechanisms have deep roots in Chinese history. The kéjǔ (科举) system, a nationwide examination system, had been used from Han Dynasty (206 BC–220 AD) on to until the end of the Qíng (清) Dynasty (1644–1911) early twentieth century to select civil officials for administration purposes (Spolsky, 1995). This system has left the Chinese society with a tradition which emphasizes the role of examination in education and a high-stake uses of "one-off results" of assessment (Han & Yang, 2001 p. 5), which over time has developed an intensely testing-dominated assessment culture. This system, along with the assessment culture, was resumed in 1977 with Deng Xiaoping's restoring *the National College Entrance Examination* to select elites in service of rebuilding the devastated country, and has been in use since then.

Also coming down in a continuous line are the ideologies about education and assessment. A "utilitarian value" of education and a "pragmatic approach" to teaching/learning, for instance, are well noted in the context, as success in examination was usually associated with immense material benefits as well as upward mobility in social status in ancient times (Chen, 2016) and visible privileges in resources rationing and opportunities such as employment and advancement at contemporary China (Jin, 2014). Teacher authority, hierarchical teacher/student relationships and passive learners (Poole, 2016) are also rooted in the Chinese educational tradition. All these values, though not necessarily unique to this context and not without alterations or intermissions over time, still hold a "residual influence" on the mindsets and behaviours of the members within this cultural community (Carless & Lam, 2014; Kennedy, 2016).

This educational system, testing-dominant tradition and these embedded values are at sharp odds with formative assessment principles, and has made it "particularly difficult for formative assessment to be established" in the context (Carless & Lam, 2014 p.167). This difficulty is made evident by a multitude of studies. For instance, students are found to lack the awareness and ability to participate in assessment because of their habitual reliance on teachers' transmission in classroom (Su, 2012), care about the assessment results and tend to neglect learning and its formative process (Chen et al., 2014), or lack of assessment-related training (Wang, 2014). The objectivity of testing is still highly valued by teachers, students and parents alike, and seen as a fair means to achieve educational equity (Cheng & Curtis, 2009). Formal and informal summative assessment is frequently used as a means of motivating guiding instruction in classrooms (Kennedy et al., 2008). Secondary school teachers' assessment practices remain largely driven by textbooks, teaching experience and high-stakes examination; the curriculum standards which were designed as a means of formative assessment became "the garnish only" (Gu, 2014). Of course, there are reports about teachers adopting strategies to use summative results formatively and productively (Xiao, 2017). Yet, overall, teaching as well as assessment practices at the secondary setting particularly remains highly controlled by high-stakes examination (Chen & Brown, 2013).

In the higher educational context, where stakes are much lower (Chen et al., 2020), teachers' assessment practices are found to be "complex and situated" (Liu & Xu, 2017 p. 27). For instance, formative assessment prescribed in the CECR (CMoE, 2007) was misinterpreted as process assessment (i.e., to record students' performance in classroom activity, assignment and attendance during the term, and give a grade to be used in the final assessment of the course) at the school-level assessment policy and eventually enacted in classrooms (Chen et al., 2013; Chen, 2017; Huang, 2010; Wang & Wang, 2011). Authentic formative assessment strategies such as feedback, self- and peer assessment and standards sharing are used only to a limited degree (Guo & Xu, 2020). Moreover, their assessment practices were mediated by quite a few factors such as their prior assessment experience and hierarchical power relationship at workplace (Xu & Liu, 2009), and more vitally, their inadequate assessment literacy (Gu, 2014; Guo & Xu, 2020; Xu, 2013). Large classes, which are common in China, are reported to have posed challenges such as inadequate attention and reduced feedback opportunities for individual students (Xu & Harfitt, 2018). More importantly, assessment in the Chinese teachers' conception serves six competing and complementary purposes, ranging from the positively regarded students' development in personal qualities and academic abilities, to the more negatively viewed role of assessment for management and the external inspection of schools; they tend to deliver value-added benefits such as learning facilitation and personal development in addition to exam-oriented preparation (Brown & Gao, 2015). While the significance of developing teachers' formative-related assessment literacy as a conceptual mode is acknowledged (Xu & Brown, 2016), how to realise this aspiration is a big challenge for themselves as well as the trainers.

Fulmer et al. (2015) coined the term *contextual factors* to describe all the influences on teachers' assessment practice and derived a multi-level conceptual model from the ecological system theory. The macro-level of the model focuses on broad national and cultural influences such as national curriculum, cultural values and norms, and national or international policies; the meso-level mainly involves factors that are external to the classroom, yet directly influence it, particularly school-specific factors such as policies and support from school leadership, school climate for assessment and training and technical provisions; the micro and the most specific level encompasses influences from the immediate context of the classrooms, that is, teachers' and students' individual factors that might influence teachers' assessment practices in classrooms.

Liu and Xu (2017), upon a comprehensive review of the contextual constraints and relevant literature, pointed out that formative assessment in the Chinese context was problematic not only at the macro (cultural and system)-level, but also at the meso- (institutional/school) and micro (classroom)-levels. Considering the variations of the over 2000 institutions within the country, they advocated for a local perspective of understanding the complexities that implementing this imported initiative in the Chinese context could have involved. This study is accordingly designed as such to unveil the complex situations that formative assessment in a provincial locality of the Mid-western China is faced with data from practicing teachers and administrators. The focus, however, is mainly on the second and third levels for three reasons. The first is that the larger constraints in the Chinese and other CHC

contexts have been much discussed in international literature (Chen, 2016, 2017; Carless & Lam, 2014; Kennedy et al., 2008; Poole, 2016) and presented previously in this section; yet, challenges at the second and third levels are still under-explored (Fulmer et al., 2015), particularly in the Chinese context (Liu & Xu, 2017). Second, compared with macro-societal studies, a focal investigation into the assessment environment of a specific context leads more likely to a nuanced understanding of formative assessment in action (Cheng et al., 2015). Thirdly, as previous studies have largely taken place in the context of more economically and educationally developed parts of China and involved only one or two cases, a multiple-case study from an unprivileged region would undoubtedly add a missing piece to the puzzle about formative assessment in the Chinese context.

Teachers' and administrators' perspective is valued in this study because they, as the actual practitioners and key agents of assessment policy (Leung, 2014), have personally witnessed and experienced the change, and know well the ins and outs of the matter (Fullan, 2015); their voices, therefore, should be heard. The specific research question this study sought to address is as follows:

How is the formative assessment initiative afforded and constrained in Chinese universities of undeveloped localities?

3 Research context and method

3.1 Context of the study

As this study aims for a local perspective, it is necessary and essential to introduce its situated context. The disciplinary and educational context of this study is College English—the compulsory EFL course for undergraduate students who major in disciplines other than English in the over 2000 Chinese higher educational institutions. For the past four decades, assessment in this area is virtually equivalent to testing, with College English Test (CET-4/6), a large-scale standardised testing system dominating the scene (Jin & Sun, 2020). The CET-4/6 is used as an external benchmark for student English learning as well as teaching and school accountability; that is, schools ranking, teacher bonuses, students' degrees, certification, scholarship and even employment are closely associated with students' performance in it (Cheng, 2008; Jin, 2014).

The scenario was set off to a change with the issue of *College English Curriculum Requirements* (CECR) (CMoE, 2007), the unified national syllabus which explicitly advocated the incorporation of formative assessment into the College English assessment framework. A rationale to promote learning outcomes via alternative assessment was articulated as well. Acknowledging the imbalanced development of College English education in varied regions and universities, the CECR empowered individual universities to make up their curriculum arrangements according to their individualised circumstances. Well aware of the importance of the leadership of individual universities, the CMoE demanded necessary support be provided at local levels to ensure a smooth and effective implementation of the policy. It is also noteworthy that this initiative was accompanied by a formal press release from the

CMoE ([2005](#)), which encouraged the unpegging of high stakes from the CET-4/6. A large-scale report reveals that many universities have disconnected the CET-4/6 from degree certification and included a formative element in the College English assessment thereafter (Wang & Wang, [2011](#)). Now, after over a decade, the overall situations are supposed to be stable enough for a deep investigation.

Geographically, the present study was conducted in an inland province of Midwestern China. It is important to mention that, just like its disadvantaged economic status, local education overall (and higher education particularly) is in desperate need of improvement. This can be seen from the fact that no university in this province is included in the 985 program (top 39 universities) and only one university is enlisted among the 211 program (top 116), two national programs that the Chinese government launched to uplift its higher education to the world level with financial investments, resources and policy supports (Costa & Zha, [2020](#)). This situation has put local higher education in a rather disadvantaged position. For one, they receive limited funds and support from the national and local government, with which teachers and teaching are inadequately resourced in terms of professional development and scholarship. For another, local universities could only enroll students who score lower in the *National College Entrance Examination* than those admitted by the 985 or 211 universities (also referred to as key universities). The students' English proficiency level hence could be less than desirable. It also needs noting that this region, as the birthplace of the Chinese civilisation, is known for conventionality and conservativeness (Zhang, [2006](#)). That is, people there are largely reluctant to change. This background does not seem optimistic for the establishment of formative assessment. Yet, formative assessment, with its learning-wise potential, might just as well be a way out for this or this kind of region to invigorate and uplift its education. All these give more prominence to this study.

3.2 A multiple-case study approach

This study adopted a multiple case study approach, as it intended to investigate indepth (Creswell, [2015](#)) the local environment for the implementation of formative assessment. To dig beyond what is happening to why it has happened, and make the study more robust (Creswell, [2015](#)), this study involved all the eight major universities in this inland province. The eight universities, including an agricultural university (abbreviated as AU), a comprehensive university (CU), an engineering university (EU), a finance and economic university (FU), a science university (SU), a medical university (MU), a normal university (NU) and a technology university (TU), are all well-established in the local province, situated in or near the provincial capital (Table [1](#)). Case selection as such is supposed to be able to reflect the overall situations for formative assessment in this province.

Like many others around the country, most of the eight universities (except AU) have incorporated a process assessment element of 10–30% to their College English frameworks in response to the CECR (CMoE, [2007](#)) and its formative assessment initiative (Chen, [2017](#)).

Table 1 Information of the eight universities

University	Location	History	Student population
AU	A small town near the provincial capital	Over 110 years	20,000
CU	Provincial capital	Over 120 years	15,000
EU	Suburb of the provincial capital	Over 60 years	20,000
FU	Provincial capital	Over 60 years	12,000
SU	Provincial capital	Over 100 years	40,000
MU	Provincial capital	Over 90 years	11,000
NU	Provincial capital	Over 60 years	11,000
TU	Provincial capital	Over 60 years	30,000

3.3 Data sources

The data for this paper were collected from two sources: (1) individual face-to-face interviews with the deans of College English education in the chosen universities. They were coded along with the university they were from. The Dean from AU, for example, were coded as AUD; and (2) focus group interviews with College English teachers from each of the eight universities. Altogether, 40 teachers were involved (Table 2). They were coded along with their universities and the order they appeared in the interviews. AUT1, for instance, is the first teacher who talked in the focus group interview of AU teachers.

The teacher participants were recommended by their deans so that they could fairly represent the College English teacher cohort in terms of gender, teaching experience and professional title. Their consent was sought before the interviews. Admittedly, this recommendation type of teacher interview participant is convenient, yet could be biased, for the Dean could have recommended the most impressive teachers in his/her mind. As can be seen from the table, six teachers were male, which left female teachers to account for 85% of the participants. This could be a reflection of the disproportionate amount of female teachers in the EFL field. In terms of professional titles, lecturers (32, 80%) took up the lion's share, while teachers of high (5, 12.5%) and low academic ranks (3, 7.5%) were in the minority. As professional title is often viewed an indicator for teacher quality in the Chinese context (Wang, 2021), the majority of middle- and low-ranking informants could in a sense mirror the unsatisfactory teacher quality in this region. Over two-thirds of the participants have taught College English for more than 10 years, the rest were comparatively young and inexperienced. These varieties were attained so that teachers' perspective could be documented to the best possible extent.

3.4 Data collection and analysis

The interviews were based on the interview schedule that Chen (2017) developed, but were modified to include questions about curriculum arrangements,

Table 2 Teacher Information (Note: F = female; M = male)

| University	Teacher	Age	Sex	Title	Teaching years	No. of student & class
AU	AUT1	34	F	Lecturer	11	120×3
	AUT2	34	M	Lecturer	10	120×4
	AUT3	36	F	Lecturer	12	120×4
	AUT4	36	F	Lecturer	14	120×4
	AUT5	40	F	Lecturer	12	120×3
CU	CUT1	33	M	Lecturer	6	50×4
	CUT2	32	M	Lecturer	9	60×2
	CUT3	35	F	Lecturer	11	50×3
	CUT4	30	F	Lecturer	5	50×2
	CUT5	47	F	Professor	23	55×2
EU	EUT1	29	F	Assistant	3	80×3
	EUT2	36	F	Lecturer	6	80×3
	EUT3	28	F	Assistant	2	80×3
	EUT4	43	F	Asso-Prof	15	80×2
	EUT5	35	F	Lecturer	11	80×3
	EUT6	36	F	Lecturer	10	80×2
FU	FUT1	58	F	Professor	30	55×3
	FUT2	38	F	Lecturer	14	60×3
	FUT3	40	F	Lecturer	18	55×3
	FUT4	40	F	Lecturer	17	60×3
	FUT5	56	M	Professor	22	55×3
MU	MUT1	41	F	Lecturer	12	50×3
	MUT2	36	F	Lecturer	12	60×2
	MUT3	33	F	Lecturer	9	60×2
	MUT4	36	F	Lecturer	10	60×2
NU	NUT1	32	F	Lecturer	10	40×4
	NUT2	37	F	Lecturer	6	40×4
	NUT3	35	F	Lecturer	6	40×4
	NUT4	37	M	Lecturer	14	40×6
	NUT5	30	M	Assistant	4	40×6
SU	SUT1	33	F	Lecturer	10	80×3
	SUT2	34	F	Lecturer	12	85×3
	SUT3	35	F	Lecturer	12	80×3
	SUT4	38	F	Lecturer	16	80×3
TU	TUT1	53	F	Asso-Prof	29	120×2
	TUT2	35	F	Lecturer	12	150×3
	TUT3	42	F	Lecturer	20	120×2
	TUT4	34	F	Lecturer	10	150×3
	TUT5	44	F	Lecturer	14	150×3
	TUT6	40	F	Lecturer	18	150×3
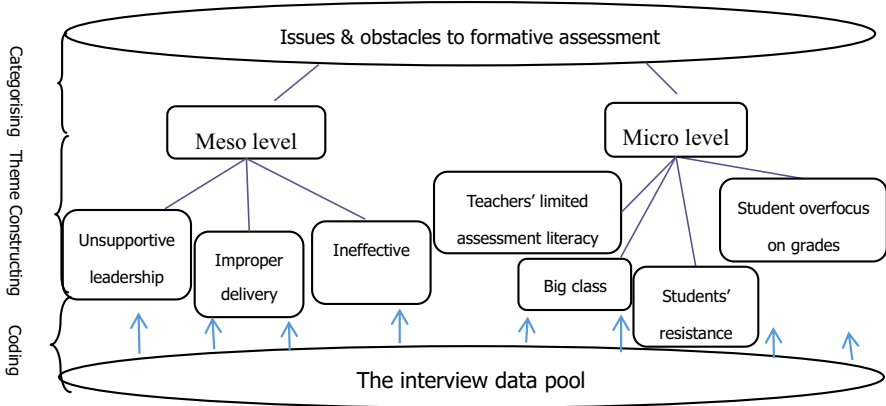
Fig. 1 Data analysis process

teacher allocation and target student population for a better understanding of the situated context. The interview questions particularly relevant to this study were about the affordances for and constraints to the implementation of formative assessment in local universities. The interviews were conducted in mandarin, the native language of the participants, to ensure effective communication. The interviews, which were around 16 h in total, were audio-recorded and transcribed. The transcription was conducted by two Master's degree candidates of translation, who were trained to transcribe beforehand. The transcription process was monitored and cross-checked by the researchers to ensure validity.

The prepared interview data were subjected to a categorical content analysis, which aims to explore meanings, themes and patterns from the text data source (Zhang & Wildemuth, 2009). Specifically, the data have been read several times and annotated by the researchers and put through an iterative condensing process. Careful open coding was followed by further clustering based on prior categories, that is, *problems and obstacles for formative assessment.* Inductive reasoning, constant examination and conferencing between the three researchers led to themes related to leadership, delivery, training, teachers' assessment literacy and class size. We have also looked beyond the phenomena for additional insights, which elicited themes related to students such as students' resistance and overemphasis on grades. The themes were further aggregated and categorised onto meso- and micro- levels. The coding and analytic process is illustrated with Fig. 1.

This process was by no means linear; rather, it is an iterative back-and-forth process, which involves categorising data bits, comparing, refining and further refining categories (Datt & Chetty, 2016). Double-checking and frequent discussions and reflections between each step helped to ensure the process trustworthy. All these were done to capture the meaning, themes and patterns manifest or latent in the interview texts so as to inform research questions this study sought to address.

4 Findings and discussion

4.1 Issues at the meso-level

Analysis of the data set revealed three major problems at the meso-level: unsupportive leadership, improper policy delivery and largely ineffective training.

4.1.1 Unsupportive leadership

The importance of the supportive leadership for the policy change was well-perceived by the teacher interviewees.

> You know, it is virtually not possible for an individual teacher to do it by him/ herself alone... definitely needs the school administration to be aware of the significance of the initiative and support it, otherwise, you know...... (FUT6).

The support, however, was far from sufficient, which was, first of all, showcased with the neglected status of College English education in these universities. While this status was commonplace to all, it was especially serious in those of specialised areas such as medicine, technology and agriculture. MUT5's irony was representative: "you know, our university [Medical University] is a specialised one. For them (at the top), we are the 'side heresy in the door'…" That is, regardless of the emphasis of CMoE (2007) on the CECR and its potential importance to students, College English education in these specialised universities was a course of minor importance in the school authorities' eyes. This situation has resulted in managing challenges for leaders of bottom level, as was verified by TUD's bitter comment: "If the leaders attach importance to College English education and it[the CECR], it will be easier [for us] to manage". Indeed, a sense of bitterness towards the status of being ignored was frequently seen in the interview data.

The neglected status of College English was manifested in inadequate facility provisions. The poor teaching facilities at MU, for instance, were cynically remarked on by MUT1 as "primitive" and "slash and burn", and by MUT5 as "mostly on strike". For this reason, the MU teachers said, they mostly "gave up trying new things" including new technologies and the innovative formative assessment initiative. For another instance, AUD hoped that the "hardware" (up-to-date equipment) needed by teachers and teaching could be supplied so that teachers could be relieved of the demanding workload pressure and spare time and energy for the assessment change and other innovations. NUD, too, complained: "the top (university authority) pays neither due attention to College English nor due salary to the teachers; how could we expected the teachers to be enthusiastic about innovations and the alike?" All this indicates that unsupportive leadership at the school level has hindered the teachers' very intention of engaging with the formative assessment initiative.

At CU, the issue was escalated into a trust crisis. The dean and teachers proposed a higher proportion of process assessment to the administration so as to better motivate students; their proposal, however, was turned down because the administration

doubted the validity of teachers' methods for scoring student performance. The excerpt below was illustrative of teachers' reactions to this accusation.

> They are afraid that we would abuse our rights. I am so offended that I begin to think: "well, since you don't trust me, why should I bother to weigh between 9 points and 10 at all?" (CUT3)

In this sense, lack of trust has hurt not only teachers' feelings but also their actual approaches to assessment.

Overall, with school authorities neglecting College English education, not providing proper teaching facilities and lacking trust in College English teachers, the overall environment in this region is more negative than positive for the assessment policy change that CMoE launched via CECR. It needs noting, however, that while the facility issue can be solved with money, the dampened enthusiasm of teachers and resistance caused by distrust and unsupportive leadership is much harder to handle.

4.1.2 Improper delivery

Delivery-related issues surfaced in the data. First of all, school leaders, who lacked the knowledge of formative assessment, failed to disseminate the formative assessment initiative and principles properly to the administrators in the field. This was made clear by SUD when he said: "Even the top (the institutional policy-makers) themselves don't understand the principles and benefits of formative assessment policies; not to mention explaining to us." As a result, the deans did not seem to have attained a proper understanding of formative assessment. EDU, for instance, frankly admitted his limitations in this regard:

> One difficulty is our [limited] understanding of assessment and formative assessment in particular. [We do not know] how on earth this kind of assessment is more advantageous in bettering our English teaching in general. Another thing is whether it is mandatory. If yes, we have no choice but to do it. But if no, we prefer not to.

EUD certainly knew what was needed for a top-down initiative to be put into practice—benefits and rationale explained to practitioners and leadership which, however, seemed absent in his and other cases this study investigated (TU and MU).

The delivery of the policy and its rationale to the teachers was questionable as well. The teachers were told to record students' performance in classroom participation, attendance, and assignments, and grade accordingly for the purposes of more attention to the process (see in Chen, 2017), that is, to practice process assessment in classrooms. It is not surprising to find teachers puzzled. CUT1, for instance, critiqued: "As far as I know, both the institutions and teachers are at lost about how to link their teaching to these changes. The link is missing, for now at least". That is to say, both deans and teachers were not provided the guidance needed to realise this policy change in classroom.

The CECR, while advocating the use of formative assessment, had empowered individual universities to "formulate their own policies according to their own actual conditions" (CMoE, 2007 p.1). The data, however, seem to show that these universities failed to take up the power. The reasons for this warrant further investigation, yet it is most likely that the policy-makers at the institutional level did not have the needed expertise to do so; or they were not used to exercising this power. After all, the individualisation of CECR was the first change of this magnitude in the history of College English (Hu, 2004). Yet, with the policy not properly delivered, institutional administrators not knowing what to do with the CECR and the formative assessment policy in particular, and teachers confused about the specific procedures and the underlying rationale, the chances for the realisation of formative assessment and its learning potential in the context were reduced to the minimal.

4.1.3 Ineffective training

To the following question: "Is there any assessment-related training provided to the teachers?", seven out of the eight deans gave a negative reply. FUD was the only one who said "yes". This "yes" was later turned out to be teaching contests and exchange activities at the campus or beyond, and training seminars provided by publishing houses during summer vocation. These trainings were not as effective as they were intended to be. To illustrate:

TUD: Only one or two days, like a whirlwind tour... of little practical value.
MUD: You cannot possibly learn much practical stuff.
CUT2: Most teachers use it as a chance for free tour...

This kind of training was provided by publishing houses as a bonus for using their published textbooks. Its ineffectiveness can be seen from above excerpts—very short, not quite useful, and not quite valued by teachers and the deans alike. A couple of teachers did mention assessment-related seminars they attended, though.

NUT4: It (assessment) was included in one session... I remember a university shared their experience. No more.
EUT6: A speaker did talk about how to assess, feedback and take students into consideration and the stuff… sorry, I forgot the details.

Listening to other university's experience seemed to have left some impression on NUT4 and EUT6, yet probably not sufficient to provoke changes in their assessment practices, particularly when details were forgotten. A teacher from CU mentioned the online training provided by experts from "above", which was organised by CMoE. However, only a few teachers attended due to limited seats. Besides, as assessment was only a "fraction" of the training content, the attendants ended up with "a rough understanding about it (formative assessment) … know they've got a concept like that" (CUT1). CUT1's conceptual understanding, similar to NUT4 and EUT6's impression, was again not enough to change teachers' understanding and practice regarding assessment.

Above analysis points to the conclusion that proper training for the CECR formative assessment initiative was either absent or ineffective, in other words, far from enough to help develop the teachers' assessment literacy and effect change in their classroom practice. This would further reduce the chances for effective formative assessment in this context.

4.2 Issues at the micro level

Micro problems out of the data analysis include teachers' limited assessment literacy, big class sizes resulting from teacher shortages, and a few student-related issues such as students' reluctance to participate and their over-attention to assessment results.

4.2.1 Teachers' limited assessment literacy

With proper training missing, it is not surprising to find the teachers limited in assessment literacy. Indeed, five out of the eight teacher groups asked the researcher to explain, before (AU, MU, EU) or amid (TU, FU) the interview, what formative assessment was. The worrying status of teachers' assessment literacy in the region was further verified in the interviews when a majority of the 40 teachers gave negative or nearly negative responses to the question: "what do you know about formative assessment?" Still, the teachers' assessment literacy varied and can be classified into three categories. The first category, which is largely illiterate, is demonstrated with the following excerpts:

> AUT1: I don't quite know.... all these technical terms in assessment domain... know really little.
> EUT2: We usually don't use it in teaching...the sight of the word (formative assessment) makes me dumb mentally.

This kind of response applied to quite a number of teachers (12 out of 40), particularly those in EU and AU where no changes to assessment policy were made. The teachers did not see the need to know much about assessment beyond what they were required to do and what they previously experienced.

The second category of teachers knew the term in name, yet not its meaning. See the excerpts below:

> TUT4: I received my master degree overseas; I know the term formative assessment. I have seen it somewhere, but I did not go deep.
> CUT5: I did hear about it, but I have no idea how to talk about it.

This group of teachers was not few either (14 out of 40). They, out of their overseas learning experience or maybe abovementioned training or else, had heard about the term, yet did not seem to go further than a nominal knowledge about it.

A few teachers (8 out of 40), of the third category, indicated that they had some basic knowledge of formative assessment due to their majors or research reasons:

> FUT5: We (pointing to FUT3) studied Second Language Acquisition in graduate school. Assessment was part of the curriculum.
> SUT3: I learned a bit as a graduate student... stuff like formative assessment in syllabus and text book evaluation.
> NUT2: I read something two years ago when I tried to write a paper. Yes, I know a little bit about it.

The few teachers of this category seemed to have developed assessment literacy from a formal or self-initiated learning experience, and applied their expertise in their classrooms. As later elaboration on their assessment experience revealed, their assessment literacy had enabled NUT2 to use peer assessment with group presentation activities in class, and SUT3 and her research team to conduct a research project which focused on the application of peer assessment. They had shed a changing light to the assessment scenario of their universities and this province at large.

Though the assessment literacy degree of the third category needs further evidence to clarify, there is no doubt that that the first two categories falls into the lowest "illiteracy" or "nominal" stages of the assessment literacy scale (Pill and Harding, 2013). With most teachers' assessment literacy not upgraded yet, large-scale change to the classroom assessment in this region could not be expected.

4.2.2 Big class size

Analysis of the collected data showed that class size in all the eight universities was big as a result of teacher shortage, though in varying degrees. In AU and TU, for example, the average class size came to 120 and 150 students, respectively. With each teacher taking up two to four classes, the teacher/student ratio in the two universities reached up to 1:360–480 and 1:300–450. The situations in other universities were better, with average class size of 80 at EU and SU; 60 at CU and FU (Table 2). The impact of big class sizes on classroom assessment was strongly felt by the teachers. To illustrate:

> TUT6: ... the fundamental problem is too many students. I have many ideas [about assessment], just cannot put it into practice...
> AUT1: The classes are too big...you can hardly do anything, really too many students.
> EUT2: Given the big class size, we can only communicate with those who sit in the front seats. Individual feedback? Totally impossible.

Overwhelmingly, large class sizes have constrained teachers' attempts to innovate their pedagogy and assessment practices in the classroom. NU, aware of this problem, has limited their class size to 40 students, which, however, raised another issue—teachers' workload increased to 4–6 classes and 16–24 h per week—the heaviest among all the teachers involved in this study. This has made it extremely difficult for the teachers because they could spare little or no time for their own career development or updating their knowledge of new teaching practises such as assessment literacy. In this sense, the class size has become another obstacle to the

implementation of formative assessment at AU, NU and other selected universities in this region.

4.2.3 Students' resistance to participate

Quite a number of students, according to teacher interviewees, were reluctant to participate in classroom activities, which was part of the process assessment. It was further revealed that this issue was particularly salient with those of low English proficiency. MUT5 made this point very clear: "some class has got good students, and did pretty well [in group discussion and peer assessment]; yet, if students' English level is not good, this kind of activity did not work at all." This was echoed in FU, when FUT4 said: "[In my third-tier class] very few students volunteer to participate in classroom activities; sometimes it is totally silent, no response at all." A student majoring in dance (whose English was particularly poor) in NUT3's class even came to her and pled: "please don't call me to answer questions in class." That is, students, whose English proficiency was substandard, found it difficult to involve themselves in activities teachers organised in class. Some of them resorted to silence, while others uttered it explicitly. This participation issue, on the other hand, was attributed by teachers such as EUT2 to students' personal character: "you know, some students are timid, and dare not speak in class". Yet, FUT5 perceived it a cultural phenomenon: "you see, like most Chinese, students tend to be shy; it's hard to make them express themselves [in public]." Regardless of the inclusion of classroom participation within assessment framework, passive rather than active engagement in classroom activities is still commonplace to students. Whether it be lack of English proficiency, personal character or cultural disposition, it is certain that students were not quite motivated as they were supposed to be, and this passivity was preventing students from experiencing the benefits of formative assessment.

Data further revealed that students' reluctance to participate was found to be accompanied by their over-reliance on their teacher, which was elaborated upon by EUT4:

> Their dependency on teacher is noticeable; they hope teacher to show them the way to go and guide them along the way, so that they can follow step by step. They are used to this ever since primary school on. If you leave all these to them, they will write something like: "this teacher is not responsible" when evaluating teachers.

These students, like those in Su (2012), were unaware of the need to, unwilling to, or unable to take responsibility for their learning. The passivity and over-reliance of students is surely another obstacle to overcome for formative assessment to happen for real in this region.

4.2.4 Students' over-attention to grades

Students were also described by their teachers as over-focused on assessment and its results. This was demonstrated in their obsession with CET-4/6, which was currently disconnected with their certification. According to MUT1, "CET-4/6 is not

mandatory at all. Yet, students couldn't put their minds at ease if they don't make it". They even "strongly asked" teachers to use class-time to prepare them for CET-4/6; otherwise, they "refused to come" (MUT1, MUT3, TUT5). "Quite utilitarian!" was the teachers' unanimous comment. Students' attention on CET-4/6 was not limited to MU; rather, it is quite commonplace across the eight universities, because their performance on the test remained linked with other benefits such as awards and employment opportunities. MU students seemed quite obvious in this regard, most likely due to their demanding specialties, which had left with them limited time for a subject like English. Still, there was no denying of their utilitarian orientation towards assessment results. On another occasion, SUT2 gave examples of students' coming to her for higher grades for achievement tests with reasons such as "scholarship", "studying overseas" and the like. Put together, students' over-attention to grades seems associated with the uses that the grades were put to. With stakes like awards, employment, scholarship and studying overseas still there., it is irrational to expect students to divert their orientation away from testing and grades.

Their concern for grades was extended to that of their peers'. In a peer assessment experience, SUT4 gave students writing criteria and asked them to assess each others' writing tasks, and noticed that:

> They are capable of evaluating; however, out of concern for face or their relationship, they won`t grade under 60 points even though he knows the article is not well-written... haha! You can just feel how they think.

Wondering if this concern might lead to unfair assessment results, SUT4 raised doubts on the validity of peer assessment. NUT2 shared the same doubt, which, however, arose from an incident in her class. She told of a student, who confidently gave himself 100 points when he had the chance to assess his own presentation. NUT2, knowing his performance did not deserve that much, doubted the feasibility of empowering students to assess themselves and their peers. Students' lack of assessment literacy in self/peer- assessment or lack of criteria to refer to could be the reason; yet, students' obsession with assessment results and caring more about face or relationship than fairness and objective judgement put the teachers to struggle balancing between the curriculum requirements and students' needs. On the other hand, these behaviours conflict with the principles of formative assessment, which prioritises learners and their use of criteria to regulate the learning process (Chen, 2016; Carless & Lam, 2014), posing further obstacles for the translation of formative assessment from rhetoric to reality in this chosen context.

5 Discussion as related to the research question

Education is an ecological system in which "multiple" and "nested" subsystems act and interact to effect changes (OECD, 2019). Educational reforms need to follow well-planned procedures for effective implementation and sustainable development (ARG, 2009). Formative assessment-related innovations, for their revolutionary nature, are more complicated because they demand stakeholders to have a transformed understanding of assessment from principles to procedures, and be well

provided throughout the process (ARG, 2009). Any missing link in the process and in the system might lead to the futility or failure of the reforms. The above analysis of data, while identifying multiple issues, demonstrates that the implementation of formative assessment initiative that the CMoE (2007) issued via the CECR are poorly afforded and well constrained in this unprivileged region of China.

The meso- or institutional level saw constraints such as unsupportive leadership and financial provisions, improperly delivered policy and insufficient and ineffective training. These issues demonstrate explicitly missing or weak links all through the implementation processes (ARG, 2009). Firstly, leadership at the local level is crucial to the success and failure of assessment reform in that it decides to a large extent whether the right conditions, structures (Spilliane, 2006), and more importantly, school climates (Hallinger, 2009; Scott et al., 2016) are set up for the change. An unsupportive leadership like this is detrimental to the teachers' agency in initiating and acting out innovative assessment. Distrust in teachers could entail more severe consequences, because educational change is virtually impossible in low-trust settings, especially when it comes to assessment reforms (Louis, 2007). Worse still, trust and positive feeling, once destroyed, are of little chance to be rebuilt (Carless, 2009). Given that "without effective educational leadership, little educational change will happen, and still less of it will be sustained over time" (Leithwood et al., 1999 p. viii; Fullan & Kirtman, 2019), these support-related issues, along with a consequently negative climate for change, could most probably screen out a big share of the possibility of effective formative assessment in these universities.

It also needs to be acknowledged that the reasons for these meso-level issues may not necessarily be exclusively technical. Rather, contextual factors such as the unprivileged positions of these universities and the disadvantaged localities of this region may, in part at least, account for most of these meso-level issues. Improper facility and training provisions, for example, could have been traced to the limited funding (Zhang, 2017) of these non-key universities from the central government and the economically developing provincial government. And the conservative dispositions of the local people (Zhang, 2006) could be part of the reason for leaders' reserved responses to the formative assessment initiative. These contextual factors, along with those mentioned above, have hindered the enactment of formative assessment at the school and administrative level in this locality.

The second, dissemination, a key link of policy implementation process, is particularly important when the policy adopts a top-down approach (ARG, 2009). If this process could not be ensured, certain things about the policy change including the rationale for the change, what to do, and how to do it, would not be able to properly delivered to lower-level policy-makers and practitioners (Fullan, 2015). It is safe to say that the dissemination of formative assessment, as revealed above, has failed to fulfill its designed functions; and hence, has left both the deans and teachers in these universities lost and confused. These findings also imply a possible failure to take up and make good use of the power that the CECR has been empowered to the institutional level (CMoE, 2007).

Finally, adequate training is a prerequisite and necessary condition for the implementation of all top-down educational innovations (Fullan, 2015). Formative assessment initiatives like the CECR need in-depth and continuous training

to ensure professional learning and conceptual change to happen because a paradigm shift is involved (Xu & Brown, 2017). The ineffective training provided to the teachers involved in this study was impossible to affect transformation in their understanding and assessment practice. With dissemination, training and agency—the three major links of the implementation process (ARG, 2009) all going wrong—the chances of effective enactment of the CECR formative assessment initiative in this region are very slim.

The effective implementation of formative assessment requires a classroom environment with teachers professionally trained in assessment, and students actively engaged and responsible for their learning (ARG, 2002; Black et al., 2003). Above data analysis, however, reveals a classroom with teachers overloaded and mostly illiterate in formative assessment except for a few, and students passive, habitually reliant and utilitarian. Teachers' assessment literacy is critical to "the success of educational assessment and even the overall quality of education" (Xu & Brown, 2017 p. 133). Without adequate assessment literacy, teachers not knowing what to do or how to do it (Black et al., 2003), a change to classroom assessment practices is not likely to happen (Taylor, 2009). While students' utilitarian approach to assessment results seems to echo their ancestors in the imperial kéjǔ times (Han & Yang, 2001), data in this study have revealed that their valuing of examination results is more closely linked with the practical uses to that the results are put. Even though the unpegging of degree with CET-4/6 have reduced some of the stakes, employment opportunities, awards and scholarships in academic or societal settings are realities they cannot afford to ignore (Chen et al., 2020). Also revealed in this study is that students' passivity and reluctance to participate in classrooms, rather than being a stereotyped cultural trait of CHC or Chinese learners in particular (Chen, 2016; Carless & Lam, 2014), could come from a variety of other reasons such as their habitual reliance on teachers (Su, 2012), their personal character, their limited assessment literacy (Wang, 2014) and more saliently, their low proficiency level.

These micro level issues again seem to associate with the disadvantaged development of this region and the unprivileged status of these universities. Limited funding and resources could mean limited professional development opportunities for local teachers, and opportunities to go beyond the Chinese learning and assessment regime and be exposed to a different learning mode and assessment culture are even less. Only one teacher in the interview mentioned her oversea study experience. This does not necessarily mean out of the 40 teachers, only one has academic experience overseas, since teachers' overseas experience was not covered in the interview. Yet, it is almost certain that their opportunities are definitely not comparable to their counterparts in elite universities in the country. Students' low English proficiency might again be a result of these universities' non-elite status, which can only enroll students of grades much lower than those elite universities. Teachers' and students' prior learning and assessment experience, which is mostly local, and their ideologies, which are more traditional than open, could play a part too. Overall, a micro environment like this goes totally against the conditions required for formative assessment (ARG, 2009), and leaves little hope for change which is intended in the CECR formative assessment initiative (CMoE, 2007).

It is hard to predict, when the factors of three levels are put together, how assessment is to enacted in day-to-day classroom practice, and how the teachers will perform within all these boundaries and balance between various tensions in their situated context (Xu & Brown, 2016). Nonetheless, it is safe to conclude that the overall environment for the realisation of formative assessment potentials in this region is by no means favourable and has addressed the research question, "How is the formative assessment initiative afforded and constrained in the Chinese universities of undeveloped regions?" in a negative way. To be more exact, this region does not yet seem to have the necessary soil needed for formative assessment to take roots and bloom, at least for now. This conclusion goes against the vision that the CECR formative assessment initiative (CMoE, 2007) has wanted for College English assessment and education in general, and provides more food for thought about the top-down approach that has been proved ineffective and even detrimental on many occasions (Skedsmo & Huber, 2019).

However, it needs noting that these issues are not all present at each university. Also, some elements of the bright side are visible as well—some teachers (NUT2, SUT3, SUT4, MUT4) who endeavoured to try out innovative pedagogy to empower and engage students in classroom activities, and some deans (NUD, MUD, TUD) clearly knew where they currently were and what was needed to move forward. More importantly, some delightful changes have taken place in the past few years to College English education in China. Firstly, *China's Standards of English Language Ability* (CMoE, 2018) was developed and formally issued, which means that the assessment criteria, an essential condition for formative assessment, are now ready for use. Secondly, *The College English Teaching Guidelines (CETG)* has been issued recently. This new syllabus explicitly demands the "balanced use" of external/internal test, formative/summative assessment and qualitative/quantitative assessment for the best "curriculum enhancement purpose" (CMoE, 2020 p. 25). Compared with the CECR (CMoE, 2007), which advocated the incorporation of formative assessment into the College English assessment framework, this is a more fully considered and further advanced step. This syllabus also especially emphasizes that the significance of EFL education to tertiary students, and demands the guarantee of human, material and financial resources as well as teacher training. Hopefully, these measures could attract more attention from funding organisations and institutional authorities. Thirdly, CMoE has sponsored several large-scale projects on English teacher assessment literacy development in the past few years. Moreover, top universities in China such as Beijing Normal University and South-China Normal University have tried to provide face-to-face or online assessment literacy sessions. Indeed, this past year, thanks to the COVID-19 pandemic, has seen a flood of online conferences and seminars organised by privileged universities, publishing companies, or organisations, which were mostly free and accessible to well-informed teachers. Some of the sessions were assessment-related. In addition, over a decade's practice of process assessment in the College English area has shaken the originally dominant position of summative testing and redirected the assessment orientation towards the process, to some extent at least. It can be said that the overall assessment environment in this area is changing. The prospects of formative assessment in this region are hence not necessarily all bleak.

6 Implications

The above findings and conclusion provide rich implications for the formative assessment initiative that the CMoE launched via the CECR (2007) and the coming CETG to be put into practice for real and its intent be better realised in this and other undeveloped regions. For one, more national and/or local funds are to be allocated to these disadvantaged universities so that classrooms may be properly equipped, and teachers be reasonably paid and professionally updated. For another, the policy is to be further developed at both the top and at the institutional levels so that teachers have specific and practical procedures to follow, and know the rationale for so-doing. Also, the policy needs to be delivered well so that enactors at every level understand the awaiting benefits of formative assessment and how to achieve them. For this purpose, professional and sustained training to administrators, deans, teachers and students is essential, so that the "missing link" between policy and practice could be provided. The training to the institutional policy-makers could be pivotal and vital because the training is supposed to enable their effective uptake of the power that national policy has given to their hands. Given the relatively conservative mindset of the local people (Zhang, 2006), training in this region probably needs to be "differentiated and situated" (DeLuca et al., 2019) and substantial to ensure "professional learning" to accrue in their assessment literacy and their identity as evaluators reconstructed (Xu & Brown, 2016). In addition, measures such as recruiting more teachers, exempting students of high English proficiency, and making good use of online resources could be taken to reduce class size and teachers' workload, so that they have time and energy to update their professional knowledge repertoires. Moreover, a supportive leadership is needed so that necessary conditions, sound structure, and a trusting and empowering institutional culture be established for the teachers and deans to try out their understanding of these new ideas. More importantly, a bottom-up rather than top-down approach is needed so that those involved may be proactive rather than reactive in response, and policy-makers could forge clarity out of all the complexities and feed back into the policy-practice circle (Fullan & Kirtman, 2019). All these are crucial for this or this kind of local settings to catch up and for a balanced and equitable development of English education nationwide to be achieved.

Availability of data and material (data transparency).
    The data and its backup were stored in the authors' PC and other digital storing devices. The data were declared in ethical approval to be shared by researchers involved only. Editors or reviewers, whom we consider as involved, can have access to the data upon request.

Declarations

Conflict of interest The authors declare no competing interests.

References

Springer.

ARG. (2009). *Changing assessment practice: Process, principles and standards*. Retrieved from http://www.assessment-reform-group.org/ARIA%20English.pdf.

Assessment Reform Group[ARG]. (2002). Assessment for learning: 10 principles. University of Cambridge Press.

Baird, J.-A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: Fields apart? *Assessment in Education, 24*(3), 317–350.

Berry, R. & Adamson, B. (2011). *Assessment reform in education*: *Policy and practice.* Springer.

Black, P. (2015). Formative assessment – An optimistic but incomplete vision. *Assessment in Education, 22*(1), 161–177.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 1*, 7–74.

Black, P., & Wiliam, D. (2005). Lessons from around the world: How policies, politics and cultures constrain and afford assessment practices. *The Curriculum Journal, 16*(2), 249–261. https://doi.org/10.1080/09585170500136218

Black, P., Harrison, C. Lee, B. Marshall & D. Wiliam. (2003). *Assessment for learning: Putting it into practice*. Open University Press.

Brown, G. T. L., & Gao, L. (2015). Chinese teachers' conceptions of assessment for and of learning: Six competing and complementary purposes. *Cogent Education, 2*(1), 993836. https://doi.org/10.1080/2331186X.2014.993836

Carless, D. (2009). Trust, distrust and their impact on assessment reform. *Assessment & Evaluation in Higher Education, 34*(1), 79–89.

Carless, D. & Lam, C. (2014). Developing assessment for productive learning in Confucian-influenced settings. In C. Wyatt-Smith, V. Klenowski & P. Colbert (Eds.), Designing Assessment for Quality Learning (pp. 167-182). Dordrecht: Springer.*Designing assessment for quality learning*, 167–179.

Chen, J., & Brown, G. T. L. (2013). High-stakes examination preparation that controls teaching: Chinese prospective teachers' conceptions of excellent teaching and assessment. *Journal of Education for Teaching, 39*(5), 541–556. https://doi.org/10.1080/02607476.2013.836338

Chen, Q. (2016). The cultural appropriateness and appropriation of formative assessment in the Chinese context. *Journal of Shanxi University, 39*(3), 80–90.

Chen, Q. (2017). Localized Representation of Formative Assessment in China: A Regional Study from a Sociocultural Perspective. *Frontiers of Education in China, 12*(1), 75–97.

Chen, Q., et al. (2013). Interpretations of formative assessment in the teaching of English at two Chinese universities: a sociocultural perspective. *Assessment & Evaluation in Higher Education, 38*(7), 831–846.

Chen, Q., et al. (2014). The enactment of formative assessment in English language classrooms in two Chinese universities: teacher and student responses. *Assessment in Education: Principles, Policy & Practice, 21*(3), 271–285.

Chen, Q., Hao, C., & Xiao, Y. (2020). When testing stakes are no longer high: Impact on the Chinese College English learners and their learning. *Language Testing in Asia, 10*(1), 6. https://doi.org/10.1186/s40468-020-00102-5

Cheng, L. (2008). The key to success: English language testing in China. *Language Testing, 25*(1), 15–37.

Cheng, L., Wu, Y., & Liu, X. (2015). Chinese university students' perceptions of assessment tasks and classroom assessment environment. *Language Testing in Asia, 5*(1), 13. https://doi.org/10.1186/s40468-015-0020-6

Cheng, L. & Curtis, A. 2009. The realities of English language assessment and the Chinese learner in China and beyond. In L. Cheng & A. Curtis (ed.) *English language assessment and the Chinese learner.* Routledge.

CMoE. (2003). *Full-time high school English curriculum requirements*. Beijing: Beijing Normal University Press.

CMoE. (2005). *Second press conference.* Retrieved from http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/moe_746/200506/9506.html (in Chinese)

CMoE. (2011). *Nine-year compulsory education English curriculum standards*.: Beijing Normal University Press.

CMoE. (2018). *China's standards of English language ability*. Higher Education Press.

CMoE. (2020). *Teaching guide for College English*. Higher Education Press.

CMoE. (2007). *College English curriculum requirements*. Foreign Language Teaching and Research Press.

Costa, D., & Zha, Q. (2020). Chinese Higher Education: The role of the economy and Projects 211/985 for system expansion. *Ensaio Avaliação e Políticas Públicas em Educação*, *28*(4), 1–24. https://doi.org/10.1590/S0104-40362020002802657

Council of Ministers of Education, Canada (CMEC). (2005). OECD study on enhancing.learning through formative assessment and the expansion of teacher repertoires. Retrieved from http://www.cmec.ca/Publications/Lists/Publications/Attachments/78/OECD_Formative.en.pdf

Creswell, J. W. (2015). *Educational research: planning, conducting and evaluating quantitative and qualitative research* (6th ed.). Pearson Education.

Datt, S. and Chetty, P. (2016). *8-step procedure to conduct qualitative content analysis in a research*. [online] Project Guru. Retrieved from https://www.projectguru.in/qualitative-content-analysis-research. Accessed 5 Apr 2021.

DeLuca, C., Luu, K., Sun, Y., & Klinger, D. A. (2012). Assessment for learning in the classroom: Barriers to implementation and possibilities for teacher professional learning. *Assessment Matters, 4*, 5–29.

DeLuca, C., Coombs, A., MacGregor, S., & Rasooli, A. (2019). Toward a differential and situated view of assessment literacy: Studying teachers' responses to classroom assessment scenarios. *Frontiers in Education, 4*(94), 1–10.

Earl, L. M. (2012). *Assessment as learning: Using classroom assessment to maximize student learning* London: Sage.

Flórez, P. M. T. (2014). Systems, ideologies and history: A three-dimensional absence in the study of assessment reform processes. *Assessment in Education: Principles, Policy & Practice, 22*(1), 3–26. https://doi.org/10.1080/0969594x.2014.943153

Fullan, M. & Kirtman, L. (2019). *Coherent school leadership: Forging clarity from complexity*. Virginia: ASCD.

Fullan, M. (2015). *The new meaning of educational change* (5th ed.) New York: Teacher's College Press.

Fulmer, G. W., Lee, I. C. H., & Tan, K. H. K. (2015). Multi-level model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy & Practice, 22*(4), 475–494. https://doi.org/10.1080/0969594X.2015.1017445

Gu, Y. (2014). The unbearable lightness of the curriculum: What drives the assessment practices of a teacher of English as a Foreign Language in a Chinese secondary school? *Assessment in Education, 21*(3), 286–305.

Guo, M. & Wu, X. (2008) School expansion and educational stratification in China, 1981–2006 Retrieved from https://www.academia.edu/2834176. Accessed 29 Apr 2021.

Guo, Q., & Xu, Y. (2020). Formative assessment use in university EFL writing instruction: A survey report from China. *Asia Pacific Journal of Education*. https://doi.org/10.1080/02188791.2020.1798737

Hallinger, P. (2009). *Leadership for the 21st century schools: From instructional leadership to leadership for learning*. Public Lecture Series, The Hong Kong Institute of Education, China.

Han, M., & Yang, X. (2001). Educational Assessment in China: Lessons from history and future prospects. *Assessment in Education, 8*(1), 5–10.

Hargreaves, E. (2005). Assessment for learning? Thinking outside the Black Box. *Cambridge Journal of Education, 35*(2), 213–224.

Hu, G. (2003). English language teaching in the People's Republic of China: Regional difference and contributing factors. *Journal of Multilingual and Multicultural Development, 24*, 290–318.

Hu, Z. L. (2004). The basic notions of "College English Curriculum Requirement": Individualization, collaboration, modulation and hypertextualisation. *Foreign Language Teaching and Research, 36*(5), 345–350.

Huang, H. (2010). Research on college english formative assessment under the multi-dimensional teaching model]. *Foreign Languages in China, 5*, 15–21.

Hume, A., & Coll, R. K. (2009). Assessment of learning, for learning, and as learning: New Zealand case studies. *Assessment in Education, 16*(3), 269–290.

Jin, Y. (2014). The limits of language tests and language testing: challenges and opportunities facing the College English Test. *In Coniam D. (Eds.), English Language Education and Assessment* (pp. 155–169). Singapore:

Jin, Y., & Sun, H. (2020). Forty years of the theory and practice of language testing in China: Retrospect and prospect. *Foreign Languages in China., 17*(04), 4–11.

Kennedy, K. J. (2016). Exploring the influence of culture on assessment: The case of teachers' conceptions of assessment in Confucian-heritage cultures. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 404–419). Routledge.

Kennedy, K. J., Chan, J. K. S., Ping, K. F., & Yu, W. M. (2008). Forms of assessment and their potential for enhancing learning: Conceptual and culture issues. *Educational Research for Policy and Practice, 3*, 197–207.

Klenowski, V. (2009). Assessment for Learning revisited: An Asia-Pacific perspective. *Assessment in Education, 16*(3), 263–268.

Laveault, D., & Allal, L. (Eds.). (2016). *Assessment for learning: Meeting the challenge of implementation*. Springer.

Leithwood, K., Jantzi, D., & Steinbach, R. (1999). *Changing leadership for changing times*. Open University Press.

Leung, C. (2014). Classroom-based assessment issues for language teacher education. In A. J. Kunnan (Ed.), vol. 3. *The companion to language assessment* (pp.1510–1519). Chichester, UK: Wiley Blackwell.

Liu, J. & Y. Xu. (2017). Assessment for learning in English language classrooms in China: Contexts, problems and solutions. In Reinders et al.(ed.) *Innovation in language learning and teaching* (pp. 17–37)*. London: Palgrave.

Louis, K. S. (2007). Trust and improvement in schools. *Journal of Educational Change, 8*(1), 1–24.

Marshall, B., & Drummond, M. J. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education, 2*, 133–149.

Nguyen, T., & Khairani, A. (2017). Reviewing the challenges of implementing formative assessment in Asia: The need for a professional development program. *Journal of Social Science Studies, 4*(1), 160–177.

OECD. (2016). *Education in China: A snapshot*. Retrieved from https://www.oecd.org/china/Education-in-China-a-snapshot.pdf. Accessed 29 Apr 2021.

OECD. (2019). *Future of Education and Skills 2030: Curriculum analysis*. Retrieved from https://www.oecd.org/education/2030-project/curriculum-analysis.pdf. Accessed 28 Apr 2021.

Pham, T. H. T., & Renshaw, P. (2015). Formative assessment in Confucian heritage culture classrooms: Activity theory analysis of tensions, contradictions and hydride practices. *Assessment & Evaluation in Higher Education., 40*(1), 45–59.

Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing, 30*(3), 381–402.

Poole, A. (2016). 'Complex teaching realities' and 'deep rooted cultural traditions': Barriers to the implementation and internalisation of formative assessment in China. *Cogent Education, 3*(1), 1156242.

Scott, S., Scott, D. E., & Webber, C. F. (2016). *Leadership of assessment, inclusion, and learning*. Springer International Publishing.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Skedsmo, G., & Huber, S. G. (2019). Top-down and bottom-up approaches to improve educational quality: Their intended and unintended consequences. *Educational Assessment, Evaluation and Accountability, 31*(1), 1–4.

Spilliane, J. P. (2006). *Distributed leadership*. Jossey-Bass.

Spolsky, B. (1995). *Measured words: The development of objective language testing*. Oxford University Press.

Su, W. (2012). Implementing formative assessment: Problems and solutions. *China Adult Education, 3*, 99–101.

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics, 29*, 21–36.

Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal., 27*(5), 615–631.

Wang, G. (1996). Educational assessment in China. *Assessment in Education: Principles, Policy & Practice, 3*(1), 75–89.

Wang, W. (2014). Students' perceptions of rubric-referenced peer feedback on EFL writing: A longitudinal inquiry. *Assessing Writing, 19*, 80–96.

Wang, S., & Wang, H. (2011). On the state of College English teaching in China and its future development. *Foreign Languages in China, 5*(4), 11–17.

Wang, X. (2005). Influences of the elementary education system on the equality of higher education *Heilongjiang Researches on Higher Education* (12), 17–19.

Wang, H. (2021). Breaking the "five-only" higher education evaluation: Difficulty, pain and breakthrough. *Journal of Chongqing University*, 1–10.

Wanner, T., & Palmer, E. (2018). Formative self-and peer assessment for improved student learning: The crucial factors of design, teacher participation and feedback. *Assessment & Evaluation in Higher Education, 43*(7), 1032–1047.

Xiao, Y. (2017). Formative assessment in a test-dominated context: How test practice can become more productive, *Language Assessment Quarterly*, *14*(4), 295–311.

Xu, Y. (2013). A review of classroom-based assessment literacy for EFL teachers. *Foreign Language Testing and Teaching, 04*, 42–50.

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*, 149–162.

Xu, Y., & Brown, G. T. L. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment, 6*(1), 133–158.

Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly, 43*(3), 493–513.

Xu, Y. and G. Harfitt (2019). Is assessment for learning feasible in large classes? Challenges and coping strategies from three case studies. *Asia-Pacific Journal of Teacher Education, 47*(5), 472–486.

Zhang, Y. (2017). Development trend of the college entrance examination over the past 40 years. *Journal of Shannxi Normal University, 4*, 1–12.

Zhang, Y. & Wildemuth, B. M. (2009). Qualitative analysis of content. In B. M. Wildemuth,(Ed.) *Applications of Social Research Methods to Questions in Information and Library Science* (p. 1–12)., Libraries Unlimited.

Zhang, Z. (2006) A cultural interpretation of Shanxi local people's characters *Qianjin* (07), 38–40.

Zhou, M. (2019). *Behind the Chinese obsession with elite universities*. Retrieved from https://news.cgtn.com/news/2019-07-17/Behind-the-Chinese-obsession-with-elite-universities-Ip9FwF8S1q/index.html. Accessed 29 Apr 2021.