
Title	Chinese award-winning tutors' perceptions and practices of classroom-based assessment
Author(s)	Zhou, Jiming and Deneen, Christopher Charles
Source	<i>Assessment & Evaluation in Higher Education</i> , 41(8), 1144-1158. http://dx.doi.org/10.1080/02602938.2015.1066306
Published by	Taylor & Francis (Routledge)

Copyright © 2015 Taylor & Francis

This is an Accepted Manuscript of an article published by Taylor & Francis in *Assessment & Evaluation in Higher Education* on 24/07/2015, available online: <http://www.tandfonline.com/10.1080/02602938.2015.1066306>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source.

Citation: Zhou, J., & Deneen, C. C. (2015). Chinese award-winning tutors' perceptions and practices of classroom-based assessment. *Assessment & Evaluation in Higher Education*, 41(8), 1144-1158. <http://dx.doi.org/10.1080/02602938.2015.1066306>

Citation: Zhou, J., & Deneen, C. C. (2015). Chinese award-winning tutors' perceptions and practices of classroom-based assessment. *Assessment & Evaluation in Higher Education*, 41(8), 1144-1158. DOI: 10.1080/02602938.2015.1066306

Complexities and Gaps: Chinese Award-Winning Tertiary Teachers' Perceptions and Practices of Assessment

Zhou, Jiming

Deneen, Christopher Charles

This study examines Chinese tertiary award-winning tutors' perceptions and reported practices of classroom-based assessment. 17 tutors in the final stage of a national university teaching contest were individually interviewed. An interview framework was developed using three process dimensions of assessment for learning (AfL). A sequential and iterative analysis of resulting data was conducted based on Miles and Huberman's (1999) protocols for qualitative analysis. Participants demonstrated a complex set of connections between perceptions and practices around issues of sharing standards, delivering feedback and response to external assessments. Results deviate from widely promoted principles of AfL and classroom-based assessment espoused in the international literature; in doing so, they challenge existing research and assumptions about the standardized and international nature of award-winning instructors' assessment practices. Findings are discussed in relationship to understanding best practices in tertiary assessment, especially in regards to emerging tertiary education markets. Practical implications for the further development of learning and assessment practices and theoretical implications for assessment theory are also discussed.

Keywords: classroom-based assessment; award-winning tutors; Chinese tertiary education

Two major educational advances in recent decades are the development of assessment for learning (AfL) and standards-based assessment (Brookhart 2011). Universities worldwide have made visible commitments to practices and policies strongly reflecting both. Scholarship and practice around these advances are now seen as integral to higher education identity, quality assurance and quality enhancement (Macdonald and Joughin 2009). As rapidly emerging higher education players, Chinese universities are focusing upon the potential for assessment to support student learning and demonstrate achievement according to intended, explicit standards of performance. This is motivated by national and institutional aspirations to connect to the international tertiary community and the desire for better student engagement in subjects deemed of critical importance, such as English-language education (Wu 2013).

Official commitment however neither predicts nor guarantees change in practice. Implementation of change may demonstrate complex patterns of adoption, modification and resistance (Carless 2011; Deneen and Boud 2014). In understanding realized assessment practice, it is essential to analyze these variations and patterns as well as the contextual issues that contribute to them.

Teachers' perceptions serve as a powerful mediating force for the translation of assessment policy into classroom practices; this is recognized through international research as well as research specific to China (e.g. Chen et al. 2013). Teachers' beliefs are found to be a more powerful mediator of pedagogical innovation than examinations in primary schools in southern China (Deng and Carless 2010). Among primary and secondary teachers in Hong Kong and southern China, perceptions that the core intention of assessment is to evaluate students were found to strongly and positively

correlate with the perception and intended enactment of assessment for students' improvement (Brown et al. 2009; 2011).

Further research is warranted, however. Brown et al.'s studies examine Chinese schoolteachers' perception of assessment in general, without addressing specific types of assessment (e.g., low-stakes classroom assessment vs. high-stakes external examinations). Also, universities are recognized as operating under quite different value systems and traditions from primary and secondary schools (Kember and McNaught 2007). In the Chinese system, high-stake university entrance examinations demarcate a schooling and school-culture boundary between pre-tertiary and tertiary education (Chen et al. 2013). This boundary suggests that while research carried out at primary and secondary levels may offer general understandings of assessment in the Chinese context, this cannot substitute for research at the tertiary level.

This study seeks to enhance understanding of Chinese national award-winning English language tutors¹' perceptions of classroom-based assessment, their descriptions of assessment practices and the relationships between their perceptions and practices. Award-winning tutors arguably embody the qualities of tertiary teaching valued by particular communities and acknowledged through evaluation processes. These embodiments are understood as having to be congruent with contextual values and requirements (Devlin and Samarawickrema 2010; Pratt, Kelly and Wang 1999). At the same time, award-winning tutors are seen as cleaving to an international consciousness of what constitutes quality that subsumes or transcends local contextual concerns (Kember and McNaught 2007). This suggests a complex, hierarchical and potentially contradictory set of embodiments around excellence in assessment. China is an emerging higher education market but it has millennia-old traditions of examinations; award-winning assessment may then be a locus of tension and contradiction for China as it seeks global recognition for tertiary excellence. Findings in this study may carry practical implications for further development of assessment practices and understanding assessment theory in different cultural contexts.

Literature review

Classroom-based assessment

This paper focuses specifically on classroom-based assessment. For brevity, the term 'assessment' will be used from this point on to mean classroom-based assessment. Any deviations from this use will be noted. Assessment can be understood as a formal to informal continuum, with specific tasks and intentions plotted at different points along this continuum (Rea-Dickins 2001). Summative, formal assessment may comprise the most visible grouping on this continuum; less formal activities designed to elicit evidence of learning and provide adjustments of understanding may represent broader range and greater frequency of occurrence (Davison and Leung 2009). Thus, a comprehensive understanding of assessment requires extending beyond the most visible elements into those frequent interactions that represent AfL intentions.

A useful, concise way of understanding AfL intentions is to focus on three dimensions: learners' current relationship to learning, the desired state to be achieved, and the means of reaching it (Assessment Reform Group 2002). Many assessment

¹ To avoid confusion, the term 'tutor' will be used to describe tertiary educators, as opposed to primary and secondary teachers.

strategies are developed and enacted around these three dimensions. For example, Wiliam (2010) suggests effective teacher classroom assessment strategies focus on a) clarifying learning intentions and sharing criteria for success; b) engineering effective classroom discussions, activities and tasks that elicit evidence of learning; and c) providing feedback that moves learners forward. Similarly, these three dimensions resonate with operationalizing feedback as a means of identifying and closing gaps (Sadler 2007). Thus, these dimensions provide a means of understanding the ways in which perceptions and practices on the assessment continuum may represent AfL intentions.

There are tensions, though, between intentions that assessment empower learners and facilitate lifelong learning, and increased emphasis on visible, standards-informed outcomes in higher education (Torrance 2007). This tension manifests in the classroom, specifically within negotiation of standards and setting of criteria. The benefits of establishing explicit, transparent criteria are well established in assessment and especially AfL literature (e.g. Assessment Reform Group 2002; Wiliam 2010). However, the standards contributing to these criteria have explicit and tacit dimensions (Sadler 1987). Explicit standards can be verbally described and captured within criteria and outcomes (O'Donovan, Price and Rust 2004). Tacit standards, however, may reside in the minds of assessors (Sadler 1987; 2007). Making these explicit may occur through iterative teacher-learner engagements (Crossouard 2010). Such engagements include communicating performance criteria to students and exemplifying some of the elements of quality assignments (Carless 2014).

Sadler argues that adoption of explicit criteria and criterion-based assessment may go too far, though reducing learning and its exhibition to 'pea-sized bits' (2007, 390). Codification of standards may have a reifying effect, shifting the goal of learning towards adopting set criteria and meeting benchmarks of external parties and away from growth of the learner (Boud and Falchikov 2006; Torrance 2007). Buhagiar (2007) argues self-referenced criteria would allow students to better regulate their learning; in doing so, these criteria would enhance engagement and motivation.

The relative importance of making standards and criteria tacit vs. explicit varies across contexts (O'Donovan et al. 2004). Primary teachers' preference for ambiguity of criteria was observed in Chinese Heritage Language classes (He 2001). In He's study, students were compelled to draw inferences regarding their teachers' references, intentions, dispositions and goals. He argues that preference for ambiguity emerges as a communicative pattern from those of higher status (e.g. teachers) in the classroom settings. Given the importance of standards and criteria to formative and summative intentions, it is worthwhile to explore the degree and ways in which tutors at a tertiary level negotiate standards and criteria.

Feedback

The purpose of feedback is to move learners forward by bridging the gap between current level of performance and a desired goal (Hattie and Timperley 2007; Sadler 1989; 2007). Beyond this understanding, the literature demonstrates variation and contradiction. There is evidence suggesting this gap would be best bridged by focusing on the process and outcome of a task rather than the learner's sense of self (Hattie and Timperley 2007). Feedback which takes into account learners' self-esteem to avoid negative impact is, however more likely to create the conditions for feedback to be dialogic, yielding greatest effect in gap closure (Buhagiar 2007; Carless 2012). Moreover, the role and importance of feedback in influencing students' sense of self is

understood to be of great importance in sustaining learning, progress and engagement (Pryor and Croussard 2008).

As with standards and criteria, perceptions and practices of feedback reflect cross-cultural tensions and contextual influences. Stobart (2012) suggests task-related feedback sits more comfortably in cultures that see the role of the teacher to instruct rather than care for the whole learner. In Chinese education, a hierarchical but harmonious, even familial relationship between teachers and students is accepted and valued (Hu 2002). This may manifest as Chinese teachers focusing feedback more on the learner than their Western counterparts would (Leung 2001). Chen et al. (2013) found Chinese tutors' feedback tended to be more oriented towards the individual than the task and skewed towards highly positive messages. When these different paradigms interact, misunderstandings can occur. Tian and Lowe (2013) in their study of Chinese postgraduates in UK universities found that process-oriented comments intended as formative feedback from Western tutors were often misinterpreted by Chinese postgraduates as summative judgments of their academic performance and of themselves.

The need to develop harmonious tutor-learner relationships has increasing relevance in western higher education contexts. National-scale surveys (e.g. the National Student Survey in UK, the Course Experience Questionnaire in Australia) show assessment and feedback are the least-satisfying aspects of higher education (Medland 2014). The NSS survey results show UK students' satisfaction with assessment and feedback represents a year-by-year improvement, since its first implementation in 2005. Increasing use of student experience surveys and the annually published league tables for universities play a major role in the growing marketization of higher education (Medland, 2014). Enhancing students' assessment experience is, therefore a priority for western universities, as is visibly celebrating practice and practitioners identified as 'award winning.'

Understanding the perceptions of award-winning tutors

Award-winning tutors are positioned both as exemplifying best principles and disseminating best practices (Kember 2009). They are also seen as being strongly connected to an international culture of excellence (Kember and McNaught 2007). In their study of Australian and Hong Kong universities, Kember and McNaught found award-winning tutors' perceptions of assessment appeared quite uniform and in alignment with internationally recognized models of practice. Both groups espoused that assessment should be consistent with 'desired learning outcomes and eventual student needs' (110). Interestingly, participants also indicated a belief that it was their responsibility to motivate students through encouragement, modeling enthusiasm, and providing a variety of active learning approaches. Researchers concluded that this apparent uniformity stemmed from identification with an international higher education culture that superseded contextual variation. They went on to make the bold claim that there was 'no evidence of any cultural disparity between East and West' (2) between the groups.

Award-winning tertiary educators appear to strongly value AfL. Bain (2011) observes tutors' preference for supporting learning rather than recoding performance in conducting and using assessment. Bain notes that traditional tutors following the performance-based approach tend to establish behavioral rules within the class and judge students according to how well they meet those rules. In contrast, award-winners in Bain's study adopted a learning-oriented approach—they used assessment to help students learn, not just for rating, ranking or motivation. Examinations were perceived

as ‘extensions of the kind of work that is already taking place in the course’ (162). Tutors established a continuum between the intellectual objectives of the course and those that the examination assessed, often supporting this continuum through sharing standards with students. This is in alignment with recognized, salutary models of assessment and with constructive alignment between curriculum, instruction and assessment (Biggs and Tang 2011).

Excellence in assessment demands the implementation of contextually appropriate practices. Context may vary and change according to student expectations and the potentially wide range of language, cultural and educational backgrounds of students (Devlin and Samarawickrema 2010). Discourses around assessment in western higher education contexts also require tutors to take into account assessment of learning, assessment for lifelong learning, and assessment for student experience (Medland 2014). These requirements are not incompatible with AfL purposes; they do, however increase complexity of intention, planning and execution, thus making excellence in assessment practices a varied, even contested space. Understanding the complexities of award winners’ best practices in assessment is, therefore an area needing further study. Given the relative absence of research into perceptions and practices of assessment within a Chinese higher education context, it is worthwhile, then to bring these elements together in an examination of award-winning Chinese university tutors’ perceptions and practices of assessment.

Method

Participants

Interviews were conducted with 17 of 24 finalist English tutors of the Chinese National Foreign Language Teaching Awards. Each finalist represented one province in China. To reach the finalist round, tutors progressed through a three-stage evaluation process. A key prerequisite for entry into the first stage was high normative ranking through formal student evaluations (top 10%). At the institutional (Stage 1) and provincial levels (Stage 2), tutors were evaluated on 20-minute micro-lessons. Based on these results, one finalist from each province moved to the national round (Stage 3). This round consisted of three parts: a paper-and-pencil examination on subject and pedagogical knowledge, delivery of a 20-minute micro-lesson, and a question and answer session focusing on candidates’ perceptions and practices. Experienced tutors and EFL researchers constituted the evaluation panels. At each stage, participants’ assessment practices were evaluated including in the micro-lessons.

Data collection

The first author conducted one-on-one interviews immediately following Stage 3. Each interview lasted approximately 30 minutes. Participants were individually interviewed in Chinese. The selection of the 17 for interview was a matter of availability (i.e. sample of convenience). Equivalent translation into English was achieved through initial translation, checking by a bilingual ESL colleague and a final discussion of meaning between the co-authors.

Interview structure was guided by an ethnographic technique (Spradley 1979). Sequence and wording of prompts were semi-structured rather than strictly constrained, allowing interviews to remain naturalistic. A clear framework was used, however, with topic areas congruent with the three process dimensions of AfL (Assessment Reform Group 2002). Specific items drew on the framework and relevant existing research.

Analysis

A sequential and iterative procedure was used based on Miles and Huberman's (1999) protocols for qualitative analysis. First, interviews were transcribed verbatim and participants were assigned identification codes T1- 17. The first author carried out initial analysis that yielded categories under two overarching themes: perceptions and practices. Key statements were translated by the first author and verified using equivalent translation protocols. The two researchers then verified and standardized codes, code definitions and relationships based on emergence of observable data patterns and evolving understanding.

Limitations

17 tutors form a small scope of interview participants. However, the contest's sample structure offsets this limitation somewhat, in that participants from 17 different provinces in China participated. By taking advantage of the contest structure, the study significantly expands the geographic scope of any prior research into tertiary assessment in China.

Assessment practices were self-reported by participants. This paper does not, however extend any claims beyond the inherent limitations of self-report. Participants were also in the midst of a complex evaluative and discursive process focusing on teaching and assessment. Interviews were conducted immediately after competition sessions had ended, on the premise that this would maximize the likelihood of this process yielding rich, legitimate dialogue on assessment perceptions and practices.

Findings

Tutors' statements were organized into two a priori categories: perceptions and practices (see Table 1). Assessment was generally perceived as improving student learning, especially for motivating and developing better learners. Tutors' perceptions fall into a continuum regarding use of examinations in classrooms and the relationship between assessment practices and external high-stakes testing. Several tutors fully embraced using examinations in classrooms; two tutors firmly criticized them as a 'primitive' means of evaluation; the majority perceived examinations as one choice among various assessment tasks, and in-class test preparation as a necessary trade-off to help students obtain higher scores. Reported assessment practices include delivering feedback that highlighted student achievement while posing low threat to self-esteem, withholding achievement-related criteria from students and articulating/ sharing non-achievement criteria.

Perceptions	Practices
<ul style="list-style-type: none">• AfL (homogeneous)<ul style="list-style-type: none">○ For motivation○ For engagement○ Developing better learners• AoL (heterogeneous)• In relationship to external examinations<ul style="list-style-type: none">○ Misalignment○ Accommodation; 'the game'	<ul style="list-style-type: none">• Feedback<ul style="list-style-type: none">○ Limitations○ Value○ Limiting threat to self-esteem• Criteria and standards<ul style="list-style-type: none">○ Concealment (degree of)○ Achievement versus non-achievement

Table 1. Coding results

(re)Interpretations of AfL

Participants described using a broad range of assessment practices. Some examples include 'several small quizzes scattering along a semester' (T2), 'students' daily English news commentary in pair work' (T3), 'peer evaluation of writing works or oral presentation' (T10 and T13), and 'student designing posters which highlights the semester's learning' (T6).

The 17 participants uniformly expressed the perception that the main purpose of assessment was to enhance learning (albeit at different points in the interview). However, only four tutors went on to articulate this in a manner congruent with adopted framework of the study. These tutors discussed how assessment could cultivate students' awareness of criteria (T11) and how this awareness might inform students of the gap between their present state of learning and the learning objectives (T13).

By contrast, twelve participants described AfL in terms of regulating student behavior, motivation, engagement and effort. A common thread among the 12 was using positive, formative feedback as a tool for whole class engagement. Positive recognition of achievement is a familiar, research-supported use of feedback for individual learners; T1 reinterpreted the purpose of such recognition as global, though: 'to give all students a sense of achievement.' She went on to suggest this had a significant effect on student engagement and motivation: 'they'll feel they get the tutor's attention, and they'll be more engaged in their study.' This common perception influenced the way some tutors planned their practices:

Some students are good at writing and others speaking, so only diversified assessment activities can give all of them a sense of achievement. (T10)

I'd like to design assessment tasks with different difficulties, usually on three scales. I allocate the tasks to students with different proficiency, of course in an implicit way. I want to let every student feel confident and comfortable when presenting his/her works. (T12)

Eight tutors emphasized assessment as a behavior-regulating mechanism, with intimations of surveillance. They gave word dictation to students (T5), checked the notes student had made in class (T17), graded or gave scores to their writing assignments (T15), with the mutual purpose to 'make sure that students are making efforts' (T3). The grades or marks produced in assessment activities were described as 'spurring students forward for better performance' (T2).

Five tutors talked about how they used assessment practices in the classroom to develop students holistically as better learners. As T9 explains:

I hope my students develop in an all-around way, rather than only learning language knowledge and skills. (T9)

A variety of assessment practices were discussed as achieving this purpose. For example, T1 implemented group activities in which students rotated as group leaders. She explained: 'my students are pre-service teachers, and they need to learn how to distribute appropriate work to team members.' T6 expressed that various assessment practices used within a semester convey the message to students that 'learning does not mean the two-week cramming before their final exam'. T9 helped his students work out their own 'progress plans' at the beginning of a semester. 'Student self-assessment', in T9's words, 'should be combined with this progress plan, rather than used in isolation in some fragmented language ability exercises'. Tutors who expressed belief in assessment

for developing better learners emphasized the importance of building up trusting relationships with students:

Firstly the tutor must get students' trust. Students will then be willing to reveal their weaknesses to you. Only after the trust relationship has been established can the tutor further nurture their confidence and monitor their achievements. (T9)

There was significant variation among participants in the use of examinations in classrooms. Those who endorsed assessment for regulating student behavior appeared more willing to embrace examinations as effective assessment tasks. Five tutors clearly indicated that internal examinations were necessary. T15 believed 'exams in class are the most efficient way to push students and improve their learning efficiency.' T15 connected this to Chinese culture:

Students in China have been accustomed to the exam-oriented culture. They still care most about exams, like in their secondary school days. (T15)

By contrast, two tutors excluded examinations as AfL strategies: 'I don't think exams have anything to do with education' (T8), 'I told my students that this course is for discussion and idea exchange rather than for passing exams' (T11). T11 suggested that examinations are too 'primitive' to understand student progress. Neither T8 nor T11 endorsed AfL as a behavior-regulating mechanism. The majority of tutor perceptions fell between the two ends of this continuum, perceiving use of examinations as having some potential role, but with that role largely defined as a regulation mechanism. As T1 explained, exams 'keep track of students' studying so that they won't become too relaxed and sloppy.'

Assessment and external examinations

Compared with the open-minded system of classroom assessment, TEM (Test for English Majors) is a different system, very confined and standardized. We are seeking a balance between the two systems. (T10)

There are rules to get high marks! As a tutor, I can do nothing but reveal the rules to my students. (T5)

Eight tutors rejected that there was a positive relationship between their assessment and external testing. T2 stated: 'it is the daily accumulation rather than the mock test practices that will lead you to success.'

This rejection did not free tutors from the exigencies of test preparation. All but two participants felt obligated to use assessment for the purpose of enhancing students' scores on external tests. External testing was described as 'a game with specific rules' (T10). Assessment tasks were organized to prepare students for playing 'the game'. This use of assessment for test preparation was perceived as 'a necessary trade-off' (T3) in the exam-oriented context. Corresponding practices ranged. T5 required her students to cover testing-relevant topics in their daily presentations, with content addressing grammar rules or listening strategies. T6 divided her whole class into seven study groups, with group members exchanging learning strategies and group leaders keeping track of the test preparation of their members.

Feedback for dual purposes

All participants articulated the importance of feedback, but described sparse provision of it. Tutors all described their written feedback as composed mainly of grades and simple comments, phrased positively such as ‘I agree’, and ‘wonderful’ (T8). Five tutors attributed brevity of feedback to heavy workloads and intensified research pressure. Six tutors noted they supplemented written feedback with verbal feedback. Feedback was commonly described as a means of recognizing achievement and fostering effort:

You must give students feedback... or they'll feel their efforts are in vain and will give up. Feedback enables them to be more active and competitive. (T10)

Despite the apparent paradox of valuing yet limiting feedback, all participants believed that their feedback practices had positive impact. A prevalent theme in describing this impact was limiting threats to student self-esteem. T7 stated he would give relatively high marks and positive feedback for students' first assignments, and then became less lenient with his interpretation of the students' achievement. By doing so, he wanted to ‘push his low-achieving students to keep making effort’. T4 discussed her verbal feedback as a process whereby she would first praise students' ideas before commenting on their language usage. T11 labeled negative feedback as ‘judgmental’. She indicated the belief that judgment would frighten students away from committing to assessment tasks. The following quotation further illustrates this perspective:

My personal suggestion is not to directly point out students' errors in class. Instead, we should encourage them and positively influence them so that they can be better engaged into classroom activities. (T9)

For shy students who do not actively involve themselves into classroom activities, I encourage them and ask if it's because my questions are too stupid. For out-going students who are good at speaking but weak at writing, I suggest that if their writing were as good as their speaking, that would be perfect. It is a compliment that helps students improve. (T10)

These tutors were satisfied with the effect of this high-praise and low-criticism approach. For example, T10 believed her ‘mild’ and ‘humble’ feedback with low-level threat to student self-esteem worked well:

Some shy students would ask for my help after class by saying: ‘I'm not good at this. Could you help me?’ (T10)

Sharing of criteria

More than half the participants (10 tutors) expressed aversion to making achievement-related standards and criteria explicit to students prior to completion of tasks. As T1 explained, ‘some criteria are the tutors' and we don't need to make them explicit’. Five participants discussed the deliberate practice of withholding criteria relating to task process and performance. T12 expressed concern that revealing achievement criteria might ‘give students the feeling that the tutor requires something different from each student and that they are labeled by tutor’. In discussing her decision not to reveal criteria, T1 noted, ‘the point of an assessment task is to make students feel comfortable and inspired’. T1 went on to suggest that ambiguity of standards was conducive to this.

This is an especially interesting finding, as these practices and their intended outcomes echo some of the disagreement within western-originating literature on standards and criteria. These findings also seem incongruous with understandings of best practices among tertiary award-winners' use of criteria, regardless of cultural context (e.g. Bain, 2011; Kember and McNaught 2007).

These 10 tutors all expressed concern that if the criteria and benchmarks of the task were too clear, it would bore high-achieving students and intimidate low-achieving students. T2 and T5 expressed the concern that early disclosure of task criteria may stifle creativity. They preferred withholding criteria until they provided feedback on students' work. T2 described this practice as 'hiding the rubric in the feedback'. Only two tutors said they would share their achievement criteria with students prior to task completion.

By contrast, nearly all participants spoke positively about sharing non-achievement criteria. These included expectations of effort, activeness and class attendance. These criteria were described as 'belonging to students' (T1). Although not achievement criteria, these were seen as highly important. Four participants indicated that activeness and degree of involvement were recorded by the tutors and counted into students' final grades. Thus, while these criteria did not directly represent standards of achievement, they played a role in the ultimate recording of achievement, students' grades. At the same time, participants admitted that these criteria were 'subjective and casual' (T5). T1 explained:

Observation of students' activeness is an assessment, too. If students don't want to involve themselves in activities, their lack of motivation will undermine their learning efficacy. If they have motivation to participate, they will benefit a lot from the preparation process.

Discussion and Conclusions

Understanding Chinese award-winning tutors' perceptions and practices

Findings in this study extend existing understandings of assessment in Chinese tertiary education context. Compared with Chinese schoolteachers' conceptions of assessment (Brown et al. 2011), university award-winners in this study did not highly associate the accountability purpose of assessment with learning improvement. They suggested the value of tertiary education lay in learning and learner development. The perception that they had the obligation to teach students the 'game rules' of external testing seemed to emerge from their care for learners, rather than from endorsement of external examinations as meaningful assessment. Various forms of assessment were enacted in classrooms as a means to reconcile AfL priorities with external examinations.

Results of this study suggest that the commonly held stereotypes that Chinese tutors are examination-oriented or narrow in their definitions of assessment are oversimplifications, to the point of being demonstrably false. Most participants framed the learning-enhancement function of assessment through non-achievement dimensions—how assessment within the classroom motivates and engages students emotionally and behaviorally. This extends pre-tertiary findings that formative functions and feedback may be valued for reasons outside the immediate concern of increasing academic performance (Chen et al. 2013; Hu 2002; Leung 2001). Ambiguity of achievement-related criteria, articulation of non-achievement criteria, and limiting threats to student self-esteem in tutors' feedback emerged as patterns of practice. These

patterns echo Stobart's (2012) and Buhagiar's (2007) arguments that assessment may be valued as incentive and motivation to students in some contexts.

Chinese teacher's preference for ambiguity reported in He's study (2001) was found in tutors' strong preference for withholding achievement-related criteria. Findings also support that the relative importance of the tacit and explicit dimensions of criteria varies across contexts and cultures (O'Donovan et al. 2004). This resonates with Deng and Carless' (2010) argument that tutors' beliefs about the relationship between pedagogy and examinations are more significant than the nature of the assessment. That this was seen in a population of tertiary educators deemed by peers and students to embody excellence suggests that these characteristics receive significant and widespread endorsement.

Contrasting assessment excellence in different contexts

This study challenges the claim in existing research that excellence in assessment is a unitary, acultural construct (e.g. Bain, 2011; Kember and McNaught 2007). Participants demonstrated within-group variations in perceptions and reported practices. In addition, they demonstrated significant deviation from the largely western-derived construct of excellence in assessment practice. For example, tutors showed a tendency to conceal achievement-related standards; by contrast, they suggested non-achievement standards were the appropriate domain for dialogue with students. Participants also perceived feedback that made explicit the learning gap as 'judgment' that would discourage or disempower students. Consequently, participants avoided providing this feedback, instead relying heavily on highly positive messages that focused on the person as much or more than work product. These findings not only challenge unitary understandings of excellence in assessment, they stray quite far from the tenets of a research-established established framework for understanding assessment.

With these findings in mind, we suggest that constructing the concept of 'award-winning' assessment practice is a more value-laden, context-informed process than is commonly recognized. The contexts and values that surround and shape the constructs of 'excellence' and 'award-winning' also surround and shape stakeholder perceptions and practices of assessment. As explained in the methodology section of this paper, participants' journey through the award competition began with evaluations by students, and continued with evaluations by peers and external English language teaching experts. Thus, these findings cannot be readily dismissed as misjudgments on the part of a few stakeholders; we may instead have to re-evaluate the narrowing effect of assuming excellence in assessment practice is a unitary, acultural phenomenon.

In the absence of this re-evaluation, a tension emerges. There is an expressed desire in China for greater internationalization of higher education (Wu 2013). Disparity between Chinese understandings of excellence in assessment and espoused unifying principles in literature (most often derived from western contexts) will inevitably give rise to tensions around internationalization. It may be tempting to achieve resolution though asserting established standards for excellence, accompanied by externally derived frameworks and standards for teaching, learning and assessment. Doing so would be in keeping with recognized patterns of how audit culture and quality assurance are moving into Asia (Deneen and Boud 2014). This carries significant implications as China becomes increasingly concerned with recognition of excellence and university league table ranking. As the findings from this study suggest though, a unitary concept of excellence may not be accurate or applicable. Further, an equitable process of internationalization may require negotiating existing frameworks and standards with contextually situated ways and understandings; failing to do this would carry the threat

of internationalization as disempowering and hegemonizing for emerging higher education players.

Rethinking assumptions underlying pursuit of assessment excellence

Are current understandings of an international culture of excellence truly international? Are the accompanying expectations for assessment practice sufficiently pluralistic? Universities and nations increasingly seek recognition for excellence through international standards, metrics and league tables (Medland 2014). Throughout the Asia-Pacific region, nations and their tertiary institutions are attempting to develop characteristics of learning and assessment that may benefit their learners while earning reputation and league table ranking. More than any other aspect of the tertiary learning experience, assessment has the power to mediate internal and external processes of supporting and assuring quality (Deneen and Boud 2014). Will this process embrace new understandings or constrict them? Is the concept of international excellence a pluralising or hegemonizing force on emerging tertiary institutions and their nations?

Answering these questions is of seminal concern; as these new markets emerge with their varied contexts, these new players have the potential to broaden and pluralize global understanding of what constitutes best practices and excellence in assessment. This may have significant impact on the scope of research into best practices generally and best assessment practices, specifically. This can only occur, though if broadening perspective and pluralizing standards are welcome.

Within existing research, there is significant variation and even tension in establishing what constitutes best practices in key areas of classroom-based assessment, such as use of feedback, standards and criteria. Our findings reaffirm these variations and the importance of acknowledging them. For example, tutors' strategies of feedback are not just about making students happy; they are about how happiness might bring about a longer-term achievement benefit. Their emphasis on the non-achievement dimensions (e.g., motivation, curiosity, attitude, trust etc.) connects to the achievement dimension (e.g. revealing learning gaps, dialogic feedback), and subsequently into an international perspective, especially in the era of lifelong learning. These resonate with global aspirations to generate a classroom assessment environment that fosters self-esteem and motivation (cf. Buhagiar 2007; Pryor and Crossouard 2008). If we accept that the non-achievement elements are framed as supporting a broader educational purpose, we need consider whether these dimensions may legitimately fall on the spectrum of classroom-based assessment and thus require that we integrate them into frameworks for understanding assessment.

As Boud and Falchikov (2006), Torrance (2007), and Sadler (2007) have warned, overemphasis on visible outcomes and criteria might lead to a narrowing of quality of student learning experience and encourage students' criteria compliance. The desire for concealment or ambiguity of achievement standards among Chinese tutors suggests us to rethink the call for criteria transparency. It might be more fruitful to train tutors how to enact assessment in their classrooms in the spirit of congruency with the transparency principle, e.g. unpacking the unarticulated criteria in the minds of tutors throughout the tutor-learner joint participation in the evaluative activities. While this is beyond the scope of this inquiry, we suggest that this would be a robust trajectory for further research.

References

- Assessment Reform Group. 2002. *Assessment for Learning: 10 Principles*. Accessed Dec., 2014. <http://www.aaia.org.uk/content/uploads/2010/06/Assessment-for-Learning-10-principles.pdf>
- Bain, K. 2011. *What the Best College Teachers Do*. Massachusetts: Harvard University Press.
- Biggs, J., and C. Tang. 2011. *Teaching for Quality Learning at University*. 4th ed. New York: McGraw-Hill International.
- Boud, D., and N. Falchikov. 2006. "Aligning Assessment with Long-term Learning." *Assessment & Evaluation in Higher Education* 31(4): 399-413.
- Brookhart, S. M. 2011. "Educational Assessment Knowledge and Skills for Teachers." *Educational Measurement: Issues and Practice* 30 (1): 3-12.
- Brown, G. T. L., S. K. F. Hui, F. W. M. Yu, and K. J. Kennedy. 2011. "Teachers' Conceptions of Assessment in Chinese Contexts: A Tripartite Model of Accountability, Improvement, and Irrelevance." *International Journal of Educational Research* 50 (5): 307-320.
- Brown, G. T. L., K. J. Kennedy, P. K. Fok, J. K. S. Chan and W. M. Yu. 2009. "Assessment for Student Improvement: Understanding Hong Kong Teachers' Conceptions and Practices of Assessment." *Assessment in Education: Principles, Policy & Practice* 16 (3): 347-363.
- Buhagiar, M. A. 2007. "Classroom Assessment within the Alternative Assessment Paradigm: Revisiting the Territory." *Curriculum Journal* 18 (1): 39-56.
- Carless, D. 2011. *From Testing to Productive Student Learning: Implementing Formative Assessment in Confucian-heritage Settings*. New York: Routledge.
- Carless, D. 2012. "Trust and Its Role in Facilitating Dialogic Feedback." In *Feedback in Higher and Professional Education: Understanding It and Doing It Well*, edited by D. Boud and E. Molloy, 90-103. New York: Routledge.
- Carless, D. 2014. "Exploring Learning-oriented Assessment Processes." *Higher Education*. doi: 10.1007/s10734-014-9816-z
- Chen, Q, M. Kettle, V. Klenowski, and L. May. 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.
- Crossouard, B. 2010. "Reforms to Higher Education Assessment Reporting: Opportunities and Challenges." *Teaching in Higher Education* 15 (3): 247-258. doi: 10.1080/13562511003740809
- Davison, C., and C. Leung. 2009. "Current Issues in English Language Teacher-based Assessment." *TESOL Quarterly* 43 (3): 393-415.
- Deneen, C. C., and D. Boud. 2014. "Patterns of Resistance in Managing Assessment Change." *Assessment & Evaluation in Higher Education* 39 (5): 577-591.
- Deng, C., and D. Carless. 2010. "Examination Preparation or Effective Teaching: Conflicting Priorities in the Implementation of a Pedagogic Innovation." *Language Assessment Quarterly* 7 (4): 285-302. doi: 10.1080/15434303.2010.510899
- Devlin, M., and G. Samarawickrema. 2010. "The Criteria of Effective Teaching in a Changing Higher Education Context." *Higher Education Research & Development* 29 (2): 111-124. doi: 10.1080/07294360903244398
- Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81-112.
- He, A. W. 2001. "The Language of Ambiguity: Practices in Chinese Heritage Language Classes." *Discourse Studies* 3 (1): 75-96. doi: 10.1177/1461445601003001004
- Hu, G. 2002. "Potential Cultural Resistance to Pedagogical Imports: The Case of Communicative Language Teaching in China." *Language Culture and Curriculum* 15 (2): 93-105.
- Kember, D. 2009. "Promoting Student-centred Forms of Learning across an Entire University." *Higher Education* 58 (1): 1-13.

- Kember, D., and C. McNaught. 2007. *Enhancing University Teaching: Lessons from Research into Award-winning Teachers*. New York: Routledge.
- Leung, F. 2001. "In Search of an East Asian Identity in Mathematics Education." *Educational Studies in Mathematics* 47 (1): 35-51.
- Macdonald, R., and G. Joughin 2009. "Changing Assessment in Higher Education: A Model in Support of Institution-wide Improvement." In *Assessment, Learning and Judgment in Higher Education*, edited by G. Joughin, 193–213. Dordrecht: Springer. doi:10.1007/978-1-4020-8905-3_11.
- Medland, E. 2014. "Assessment in Higher Education: Drivers, Barriers and Directions for Change in the UK." *Assessment & Evaluation in Higher Education*, DOI: 10.1080/02602938.2014.982072
- Miles, M. B. and A. M. Huberman. 1999. *Qualitative Research: An Expanded Sourcebook*. 2nd ed. Thousand Oaks CA: Sage Publishing.
- O'Donovan, B., M. Price, and C. Rust. 2004. "Know What I Mean? Enhancing Student Understanding of Assessment Standards and Criteria." *Teaching in Higher Education* 9 (3): 325-335. doi: 10.1080/1356251042000216642
- Pratt, D. D., M. Kelly, and W. S. S. Wong. 1999. "Chinese Conceptions of 'Effective Teaching' in Hong Kong: Towards Culturally Sensitive Evaluations of Teaching." *International Journal of Lifelong Education* 18 (4): 241- 258.
- Pryor, J. and B. Crossouard. 2008. "A Socio-cultural Theorisation of Formative Assessment." *Oxford Review of Education* 34(1): 1-20.
- Rea-Dickins, P. 2001. "Mirror, Mirror on the Wall: Identifying Processes of Classroom Assessment." *Language Testing* 18 (4): 429-462.
- Sadler, D. R. 1987. "Specifying and Promulgating Achievement Standards." *Oxford Review of Education* 13 (2): 191-209.
- Sadler, D. R. 1989. "Formative Assessment and the Design of Instructional Systems." *Instructional Science* 18 (2): 119-144.
- Sadler, D. R. 2007. "Perils in the Meticulous Specification of Goals and Assessment Criteria." *Assessment in Education: Principles, Policy and Practices* 14(3): 387-392.
- Spradley, J. P. 1979. *The Ethnographic Interview*. New York: Holt, Rinehart & Winston.
- Stobart, G. 2012. "Validity in Formative Assessment." In *Assessment and Learning*, edited by J. Gardner, 233-242. 2nd ed. Thousand oaks, California: Sage.
- Tian, M. and J. Lowe. 2013. "The Role of Feedback in Cross-cultural Learning: A Case Study of Chinese Taught Postgraduate Students in a UK University." *Assessment & Evaluation in Higher Education* 38(5): 580-598.
- Torrance, H. 2007. "Assessment as Learning? How the Use of Explicit Learning Objectives, Assessment Criteria and Feedback in Post-secondary Education and Training can Come to Dominate Learning." *Assessment in Education: Principles, Policy & Practice* 14 (3): 281-294.
- Wiliam, D. 2010. "The Role of Formative Assessment in Effective Learning Environments." In *The Nature of Learning: Using Research to Inspire Practice*, edited by H. Dumont, D. Istance, and F. Benavides, 135-159. OECD.
- Wu, D. 2013. *Introducing Writing across the Curriculum into China*. Berlin Heidelberg: Springer.