Motivation for the Research
Teacher assessment literacy (AL), generally understood as teacher competencies regarding the assessment of student learning, has been increasingly recognized as an essential component of teacher professionalism. While recent studies show its general inadequacy across educational settings and geographical contexts, teacher AL is still an underexplored area of research, especially among university English teachers in China who teach the largest group of adult English learners in the world. To address this gap, this mixed methods study explored what, how, and why issues of teacher AL through both a quantitative survey test with a large sample of university English teachers (N=891) and qualitative case studies with three carefully-selected teachers (Linda, Rosa, and May, pseudonyms) working in three universities of different tiers in China.

Research Questions
Two main research questions guide this study:
RQ1: How assessment literate are university English teachers in China?
RQ2: How is assessment literacy enacted in the practices of a carefully selected sample of three teachers?
Together, answers to these questions seek to uncover the nature of teacher AL on both empirical and theoretical grounds.

Research Methodology
Given that the study sought for answers to the what, how and why issues of teacher AL, a mixed methods research design was employed (Creswell, 2003). Specifically, it followed a sequential explanatory model (p. 215) with a quantitative survey test followed by qualitative case studies.

First, an overall understanding of the status of teacher AL was obtained through a survey test of the University English Teacher Assessment Literacy Questionnaire (UETALQ) adapted from the Teacher Assessment Literacy Questionnaire (TALQ) (Plake, Impara, & Fager, 1993). By survey test, I mean that this instrument intends to measure teachers’ mastery of assessment principles in eight areas with 24 dichotomously scored items, although the original TALQ was called a “questionnaire.” Drawing upon data from a large sample of university English teachers across China (N=891), the results answer RQ 1 by
reporting the psychometric properties of the UETALQ, generalizing the current AL levels, and identifying any demographic features that have an impact on teachers’ AL performance (Xu & Brown, 2017).

Second, based on the results from the survey tests, three university English teachers (Linda, Rosa, May, all pseudonyms) were selected to be case studies. The case studies drew upon data from semester-long classroom observations, individual teacher interviews, and documents and sought to gain an in-depth and nuanced understanding of each teacher’s AL in practice and to identify significant personal and contextual factors that have impacted the development of their AL.

The empirical understandings of teacher AL, from both the quantitative and qualitative studies, were then put together to prepare the groundwork for a new conceptual framework of teacher AL in practice.

Summary of Findings

RQ 1 is answered in terms of the psychometric properties of the UETALQ, the general AL level of university English teachers in China, as well as the impact of the demographic characteristics on teacher AL performance. The psychometric analysis of the UETALQ using 3 parameter logistic item response theory (3PL IRT) and confirmatory factor analysis (CFA) led to a reduced one-factor model of teacher AL with ten valid items. Results of 3PL scores based on the valid items indicate that the current AL level of university English teachers in China is fairly basic, covering competency areas such as aligning tasks to instructional goals, objective scoring of assessment tasks, clarifying the purposes of assessments, engaging students in assessment practices, valid grading, accurate interpretations of scale scores, and identifying unethical assessment practices. In addition, most of the demographic characteristics have little influence on teacher AL performance (i.e., gender, years of teaching, and professional title), except for the interaction effects of two groups of features (i.e., university * region; university * qualification). The limited effects of years of teaching and professional title may indicate that AL development is a relatively independent entity that does not necessarily grow when one teaches longer or moves up on a professional trajectory; or it may indicate that AL is not included as part of teacher qualification, and, therefore, one’s AL development does not synchronize with other professional characteristics. Additionally, the significant interaction effect of two groups of demographic features suggests that university level might be a potentially important contextual factor for teacher AL.

RQ 2 is answered in terms of the three teachers’ respective AL in practice and the important personal and contextual factors that exert an influence on teacher AL. Findings from the case studies show that the three teachers displayed different stronger and weaker dimensions within AL. May was skillful at handling feedback processes yet unconfident in grading; Rosa was innovative in using technology in assessments yet struggled with quantifying student performance; May had great strategies for assessing class performance and giving quizzes and was committed to preparing students for the tests. Across these cases, commonalities include the multi-dimensionality, situatedness, and dynamics of teacher AL in practice, with significant influences from the teachers’ institutional contexts and individual conceptions of assessment. As such, each teacher’s AL is enacted in practice as a wide range of compromises made among tensions between their own conceptions of assessment and their socio-cultural and institutional contexts.

Based upon these findings, major arguments proposed by this study are as follows. First, institutional contexts play an important part in shaping teacher AL through directing the assessment purposes, prescribing the priority of teachers’ professional work, and allocating
resources and support. On the continuum of assessment purposes, the institutional policy influences which end the teacher’s AL is leaning toward: formative or summative. More importantly, the prescribed work priority may support or constrain teachers’ commitment to assessment. Research-intensive universities that prioritize research output over other responsibilities in teachers’ professional agenda compromise teachers commitment to assessment practices. Additionally, these case studies have identified the following resources as critically important in facilitating teacher AL development: well-designed assessment plan with explicit rationales and grading criteria; a standard-based checklist of principles for test item writing; and mechanism for professional conversations over moderation of assessment standards. Unfortunately, these are missing in all three institutional contexts. Arguably, at the very core of these absent elements is the negotiability of these resources and the empowerment of teachers as professionals.

Second, influences of teacher conceptions of assessment on AL, as the most important personal factor, are specified in this study. By confirming the “knowledge filtering” function of teacher conceptions in the cognitive dimension (Kagan, 1992), data from this study unravel how emotional dimensions of teacher conceptions of assessment influence teacher AL enactment. Teachers’ positive or negative emotional responses associated with assessment subconsciously shape the process of decision-making in many ways, including choosing (or not) certain assessment tasks, administering assessment with personal preferences, and attaching different levels of importance to tests. The double-channel influences of cognitive and emotional dimensions highlight a need for critical reflection and effective management of teacher emotion in assessment practice.

Third, teacher AL is multidimensional and situated. A teacher’s AL could be strong in one dimension while being weak in another and is dependent partly on individual styles, strengths, and conceptions of learning, as well as on how a teacher develops dimensions of AL formally or informally within his/her institutional contexts. A teacher’s AL in practice is not fixed but is a dynamic, developmental, and ongoing process that is constantly negotiated within tensions between conceptions of assessment and the micro- and macro-contexts. To achieve a balance among these dimensions within teacher AL, teachers need support from both teacher educators and assessment specialists to identify strengths and weaknesses within their AL dimensions and to develop accordingly in order to improve their day-to-day practice.

Fourth, teacher feedback literacy as one dimension of teacher AL has three layers of meaning. The first is related to the conventional understanding that teachers should be well versed in delivering quality feedback to help move students from their current levels to the desired goal. The second element refers to competencies of constructing dialogic feedback contexts through relational bonding and cognitive reinforcement. The third entails an enabling construct that seeds student feedback literacy through cognitive scaffolding and social-affective support.

Fifth, teachers’ grading competency, as another important dimension of teacher AL, which encompasses both cognitive and ethical aspects. Cognitively, it requires teachers to consider legitimate constructs of grading in relation to objectives of learning goals and assessment tasks, as well as the relevance of these factors to student learning. It also requires teachers to make grading decisions in a consistent and transparent way. To have their voices and professional judgments respected, teachers are advised to present their understanding of valid grading through research-based evidence. Ethically, teachers need to follow codes of practice whenever available,
understand and negotiate power relations with other stakeholders, and critically reflect upon the institutional conformity adapted to the particular context. The cognitively and ethically bounded nature of grading suggests that teachers need to integrate all of these perspectives and make each grading decision afresh with criticality and compelling justifications.

Sixth, a basic AL level is not sufficient for Chinese university English teachers to take on their enormous assessment responsibilities, either for accountability or learning purposes. Although their inadequate formal assessment training, lack of professional standards, and the absence of AL in recruitment criteria can be blamed, teacher AL can be enhanced through learning from experience and colleagues. This suggests that it is potentially fruitful to embed assessment training into informal learning opportunities. In addition, a valid AL measure needs to be appropriate to the specific socio-cultural and educational contexts in which it is used and needs to incorporate generic assessment knowledge, which is applicable to all contexts, and contextually-grounded and discipline-specific assessment principles and values.

Finally, teacher AL needs to be reconceptualized into teacher assessment literacy in practice (TALiP) as it is not a static knowledge base, but rather a dynamic, situated, and developing system. Theoretical input is one source for teachers to acquire assessment principles, and yet it is filtered by the teacher’s conceptions of assessment developed from his/her prior experiences of assessment as a learner and teacher. Meanwhile, the socio-cultural and institutional contexts set requirements, directions, and boundaries for teacher AL in practice. Teacher AL is enacted in practice as a series of compromises that are made as a result of tensions between one’s conceptions of assessment and macro- and micro-contexts. To make better compromises, teacher learning is the impetus, and teachers need to constantly construct their identity as “assessors.”

**Implications**

As the first large-scale empirical investigation of AL among university English teachers in China, this study contributes to the scholarship of AL by offering a new conceptual framework of teacher assessment literacy in practice (TALiP) and a working definition of TALiP (Xu & Brown, 2016). The TALiP framework moves our understandings of teacher AL beyond a focus on the knowledge base to considerations of a situated, dynamic, and evolving system in which teachers constantly make compromises among competing tensions. The framework connects important mediating factors of teacher AL, and suggests that improvement of teacher AL is a systematic enterprise that depends on forming a virtuous cycle of TALiP driven by teacher learning and identity construction as assessors. This framework’s value also lies in the fact that it can be used as an operationalized model, as each component within TALiP serving as a point of entry for conducting AL research.

In addition, this study contributes to AL research with original insights about important influential factors and inherent dimensions of teacher AL. It first specifies how institutional contexts and teacher conceptions of assessment as salient contextual and personal factors shape teacher AL and how their interactions create tensions that lead to various compromises that teachers make. It then expounds upon two important inherent dimensions of teacher AL: feedback literacy and grading competency. It fills the gap of teacher feedback literacy research by addressing two underexplored components: constructing a context for dialogic feedback and enabling student feedback literacy (Xu & Carless, 2016). It also contributes to the existing literature by highlighting cognitive and ethical dimensions as two parallel considerations for teacher grading competency, which can be used as an analytical framework for making sense of teacher grading competency in complex cognitively- and ethically-bounded contexts.
Finally, this study offers a number of implications for policy, professional development, and teacher practice. Implications for creating policy include the development of standards for AL, the inclusion of AL in teacher licensure and accreditation, and the allocation of resources for the enhancement of AL. At the level of professional development, this study calls for diagnostic understandings of individual teachers’ AL strengths and weaknesses, reflective modules for teachers to scrutinize their conceptions of assessment derived from their prior assessment experiences, closer attention to workplace-based assessment learning, and emotional support for teachers’ assessment practice. At the level of teacher practice, this study provides a number of ready-to-use suggestions to help teachers improve the effectiveness of their assessment practices.

References


Pajares, M. F., & Graham, L. (1998). Formalist thinking and language arts instruction: Teachers’ and
students’ beliefs about truth and caring in the teaching conversation. Teaching and Teacher
Education, 14(8), 855-870.


Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.), (2001). Knowing what students know:
The science and design of educational assessment. Washington, DC: National Academy
Press.

parliamentary inquiry. Language Testing, 30(3), 381-402.


dilemmas in classroom assessment. Teaching and Teacher Education, 25, 778-782.


Practice, 48(1), 4-11.

Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator’s confession. The
Teacher Educator, 46(4), 265-273.

Studies in Higher Education, 36(8), 879-896.


Educator, 36, 115-131.

Journal of Educational Measurement, 46, 1–18.

(Ed.), The Sage handbook of research on classroom assessment (pp. 145-163). Los
Angeles, CA: Sage.


