The International Research Foundation
for English Language Education

Title of Project:
Assessment Literacy of University English
Teachers in China: A Mixed Methods Study

Researcher:
Xu Yueting
University of Hong Kong
21859236@qq.com

Research Supervisor:
Prof. David Robert Carless
University of Hong Kong

Xu Yueting

Final Report

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

2

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

Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The professional knowledge base of science teaching* (Vol. 12, pp. 205-221). Dordrecht, The Netherlands: Springer.

Adie, L. (2013). The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice, 20*(1), 91-106.

Airasian, P. W. (1991). Perspectives on measurement instruction. *Educational Measurement: Issues and Practice, 10*(1), 13–16.

Airasian, P. W., & Russell, M. K. (2008). *Classroom assessment* (6th ed.). New York, NY: McGraw-Hill.

Ajjawi, R., & Boud, D. (2017). Researching feedback dialogue: An interactional analysis approach. *Assessment & Evaluation in Higher Education*, *42*(2), 252-265.

Ajzen, I. (2005). *Attitudes, personality and behavior* (2nd ed.). New York, NY: Open University Press.

Alkharusi, H. (2011). Psychometric properties of the teacher assessment literacy questionnaire for preservice teachers in Oman. *Procedia - Social and Behavioral Sciences, 29*, 1614-1624.

Alkharusi, H., Kazem, A. M., & Al-Musawai, A. (2011). Knowledge, skills, and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pacific Journal of Teacher Education, 39*(2), 113-123.

Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice, 20*(1), 20-34.

American Federation of Teachers, National Council on Measurement in Education, & National Education Association (AFT, NCME, & NEA). (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice, 9*(4), 30-32.

Bailey, K. M., & Brown, J. D. (1995). Language testing courses: What are they? In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 236-256). Cleveton, UK: Multilingual Matters.

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 93-114). New York, NY: Routledge.

Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. In H. Fives & M. G. Gill (Eds.), *International handbook of research on teacher beliefs* (pp. 284-300). New York, NY: Routledge.

Beauchamp, C., & Thomas, L. (2009). Understanding teacher identity: An overview of issues in the literature and implications for teacher education. *Cambridge Journal of Education, 39*(2), 175-189.

Beaumont, C., O'Doherty, M., & Shannon, L. (2011). Reconceptualising assessment feedback: A key to improving student learning? *Studies in Higher Education, 36*(6), 671–687.

Beijaard, D., Meijer, P. C., & Verloop, N. (2004). Reconsidering research on teachers' professional identity. *Teaching and Teacher Education, 20*(2), 107-128.

Biggs, J. (1996). Western misperceptions of the Confucian-heritage learning culture. In D. Watkins & J. Biggs (Eds.), *The Chinese learner: Culture, psychological, and contextual influences* (pp. 45-67). Hong Kong: CERC & ACER.

Bishop, J. H. (1992). Why U.S. students need incentives to learn. *Educational Leadership, 49*(6), 15-18.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappa, 92*(1), 81-90.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*(1), 8-21.

Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Success, stumbling blocks, and implications for professional development. *Teaching and Teacher Education, 13*(3), 259-278.

Boud, D. (2000). Sustainable assessment: Assessment for the learning society. *Studies in Continuing Education, 22*(2), 151-167.

Boud, D., Cohen, R., & Sampson, J. (2001). *Peer learning in higher education: Learning from and with each other*. London, UK: Routledge.

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education, 38*(6), 698-712.

Boud, M. H., & Hwang, K. K. (1986). The social psychology of Chinese people. In M. H. Boud (Ed.), *The psychology of the Chinese people* (pp. 213-266). Hong Kong: Oxford University Press.

Brookfield, S. D. (1995). *Becoming a critically reflective teacher.* San Francisco, CA: Jossey-Bass.

Brookhart, S. M. (1999). Teaching about communicating assessment results and grading. *Educational Measurement: Issues and Practice, 18*(1), 5–14.

Brookhart, S. M. (2002). What will teachers know about assessment, and how will that improve instruction. In R. W. Lizzitz & W. D. Schafer (Eds.), *Assessment in educational reform: Both means and ends* (pp. 2-17). Boston, MA: Allyn & Bacon.

Brookhart, S. M. (2010). Mixing it up: Combining sources of classroom achievement information for formative and summative purposes. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 279-298). New York, NY: Routledge.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice, 30*(1), 3-12.

Brookhart, S. M. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice, 20*(1), 69-90.

Brookhart, S. M., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A centry of grading research: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803-838.

Brown, G. T. L. (2008a). Assessment literacy training and teachers' conceptions of assessment. In C. Rubie-Davies & C. Rawlinson (Eds.), *Challenging thinking about teaching and learning* (pp. 285-302). New York, NY: Nova Science.

Brown, G. T. L. (2008b). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York, NY: Nova Science.

Brown, G. T. L., Irving, E., & Keegan, P. (2014). *An introduction to educational assessment, measurement, and evaluation: Improving the quality of teacher-based assessment.* Auckland, New Zealand: Dunmore Publishing Limited.

Brown, G. T. L., & Remesal, A. (2012). Prospective teachers' conceptions of

assessment: A cross-cultural comparison. *The Spanish Journal of Psychology*, *15*(1), 75–89.

Brown, J. D., & Bailey, K. M. (2008). Language testing courses: What are they in 2007? *Language Testing, 25*(3), 349-384.

Buck, G. A., & Trauth-Nare, A. E. (2009). Preparing teachers to make the formative assessment process integral to science teaching and learning. *Journal of Science Teacher Education, 20*(5), 475–494.

Buzzelli, C. A., & Johnson, B. (2002). *The moral dimension of teaching: Language, power, and cuture in classroom interaction*. New York, NY: Routledge.

Bybee, R., W. (1997). *Achieving scientific literacy: From purposes to practices*. Portsmouth, NH: Heinemann.

Campbell, C. (2013). Research on teacher comptence in classroom assessment. In J. H. McMillan (Ed.), *The Sage handbook of research on classroom assessment* (pp. 71-84). Los Angeles, CA: Sage.

Campbell, C., & Collins, V. L. (2007). Identifying essential topics in general and special education introductory assessment textbooks. *Educational Measurement: Issues and Practice, 26*(1), 9–18.

Campbell, C., & Evans, J. A. (2000). Investigation of preservice teachers' classroom assessment practices during student teaching. *The Journal of Educational Research, 93*(6), 350–355.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education, 31*(2), 219-233.

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

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

Carless, D. (2015). *Excellence in university assessment: Learning from award-winning practice*. London, UK: Routledge.

Carruthers, C., McCarron, B., Bolan, P., Devine, A., McMahon-Beattie, U., & Burns, A. (2015). "I like the sound of that"–An evaluation of providing audio feedback via the virtual

learning environment for summative assessment. *Assessment & Evaluation in Higher Education*, *40*(3), 352-370.

Cartney, P. (2010). Exploring the use of peer assessment as a vihecle for closing the gap between feedback given and feedback used. *Assessment & Evaluation in Higher Education*, *35*(5), 551-564.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative anlaysis.* London, UK: Sage.

Chen, J., & Brown, G. T. L. (2013). High-stakes examination preparation that controls teaching: Chinese perspective teachers' conceptions of excellent teaching and assessment. *Journal of Education for Teaching: International Research and pedagogy, 39*(5), 541-556.

Chen, Q., Kettle, M., Klenowski, V., & May, L. (2013). Interpretations of formative assessment in the teaching of English at two Chinese universities: A socioculturual perspective. *Assessment & Evaluation in Higher Education, 38*(7), 831-846.

Chen, Q., May, L., Klenowski, V., & Kettle, M. (2014). The enactment of formative assessment in English language classrooms in two Chinese universities: Teachers and student responses. *Assessment in Education: Principles, Policy & Practice, 21*(3), 271-285.

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

Cheng, L., Roger, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practice: Purposes, methods, and procedures. *Language Testing, 21*(3), 360-389.

Cheng, L., & Sun, Y. (2015). Teachers' grading decision making: Multiple influencing factors and methods. *Language Assessment Quarterly, 12*(2), 213–233.

Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 1-3). New York, NY: Routledge.

CMoE. (2007). *College English curriculum requirements* [*daxue yingyu kecheng yaoqiu*]. Beijing, China: Foreign Language Teaching and Research Press.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher, 23*(7), 13-20.

Cobb, P., Wood, T., & Yackel, E. (1990). Classrooms as learning environments for teachers and researchers. In R. B. Davis, C. A. Mayer, & N. Noddings (Eds.), *Constructivist views on the teaching and learning of mathematics* (pp. 125-146). Reston, VA: National Council of Teachers of Mathematics.

Cochran-Smith, M., & Lytle, S. L. (1999). Relationships of knowledge and practice: Teacher learning in communities. *Review of Research in Education, 24*(1), 249-305.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). London, UK: Routledge.

Colnerud, G. (1997). Ethical conflicts in teaching. *Teaching and Teacher Education, 13*, 627-635.

Cornish, L., & Jenkins, K. A. (2012). Encouraging teacher development through embedding reflective practice in assessment. *Asia-Pacific Journal of Teacher Education, 40*(2), 159-170.

Cortazzi, M., & Jin, J. (2006). Changing practices in Chinese cultures of learning. *Language, Culture & Curriculum, 19*(1), 5-20.

Cowie, B., Cooper, B., & Ussher, B. (2014). Developing an identity as a teacher-assessor: Three student teacher case studies. *Assessment Matters, 7*, 64–89.

Creswell, J. (2003). *Research design: Qualitative, auantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Creswell, J., & Miller, D. (2000). Determining validity in qualitative inquiry. *Theory into Practice, 39*(3), 124-130.

Creswell, J., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 209–240). Thousand Oaks, CA: Sage.

Crossman, J. (2007). The role of relationships and emotions in student perceptions of learning and assessment. *Higher Education Research and Development, 26*(3), 313-327.

Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing, 25*(3), 327-347.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Publications.

DeLuca, C. (2012). Preparing teachers for the age of accountability: Toward a framework for assessment education. *Action in Teacher Education, 34*(5/6), 576-591.

DeLuca, C., & Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education, 64*(4), 356-372.

DeLuca, C., Chavez, T., Bellara, A., & Cao, C. (2013). Pedagogies for preservice assessment education: Supporting teacher candidates' assessment literacy development. *The Teacher Educator, 48*(2), 128-142.

DeLuca, C., Chavez, T., & Cao, C. (2013). Establishing a foundation for valid teacher judgement on student learning: The role of pre-service assessment education. *Assessment in Education: Principles, Policy & Practice, 20*(1), 107-126.

DeLuca, C., Johnson, S. (2017). Developing assessment capable teachers in this age of accountability. *Assessment in Education: Principles, Policy & Practice, 24*(2), 121-126.

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice, 17*(4), 419-438.

DeLuca, C., Klinger, D., Pyper, J., & Woods, J. (2015). Instructional rounds as a professional learning model for systemic implementation of assessment for learning. *Assessment in Education: Principles, Policy & Practice, 22*(1), 122-139.

DeLuca, C., Klinger, D., Searle, M., & Shulha, L. (2010). Developing a curriculum for assessment education. *Assessment Matters, 2*, 133-155.

DeLuca, C., & Lam, C. Y. (2014). Preparing teachers for assessment within diverse classrooms: An analysis of teacher candidates' conceptualizations. *Teacher Education Quarterly, 41*(3), 3-24.

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016a). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, *28*(3), 251-272.

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016b). Approaches to classroom assessment inventory: A new instrument to support teacher assesment literacy. *Educational Assessment, 21*(4), 248-266.

DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2016). Teachers' approaches to classroom assessment: A large-scale survey. *Assessment in Education: Principles, Policy & Practice*, 1–21. DOI: 10.1080/0969594X.2016.1244514

Deneen, C. C., & Brown, G. T. L. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education, 3*(1), 1-14. DOI: 10.1080/2331186x.2016.1225380.

Denzin, N. K. (2014). Writing and/as analysis or performing the world. In U. Flick (Ed.). *The Sage handbook of qualitative data analysis* (pp. 569-584). Thousand Oaks, CA: Sage.

Dewey, J. (1933). *How we think*. Buffalo, NY: Prometheus Books.

Dillman, D. A. (1991). The design and administration of mail surveys. *Annual Review of Sociology, 17*, 225-249.

Ell, F., Hill, M. F., & Grudnoff, L. (2012). Finding out more about teacher candidates' prior knowledge: Implications for teacher educators. *Asia-Pacific Journal of Teacher Education, 40*(1), 55-65.

Engelsen, K. S., & Smith, K. (2014). Assessment literacy. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *The enabling power of assessment: Designing assessment for quality learning* (pp. 140-162). New York, NY: Springer.

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research, 83*(1), 70-120.

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavior Psychology, 70*(1), 113-136.

Fan, Y.-C., Wang, T.-H., & Wang, K.-H. (2011). A Web-based model for developing assessment literacy of secondary in-service teachers. *Computers & Education, 57*(2), 1727-1740.

Feldman, K. A. (1997). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Eds.), *Effective teaching in higher education: Research and practice* (pp. 93-143). New York, NY: Agathon Press.

Feng, A. (2009). English in China: Convergence and divergence in policy and practice. *AILA Review, 22*, 85-102.

Fives, H., & Buehl, M. M. (2012). Spring cleaning for the "messy" construct of teachers' beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Individual differences and cultural and contextual factors* (Vol. 2, pp. 471-499). Washington, DC: American Psychological Association.

Fleer, M. (2015). Developing an assessment pedagogy: The tensions and struggles in re-theorising assessment from a cultural-historical perspective. *Assessment in Education: Principles, Policy & Practice, 22*(2), 224-246.

Flick, U. (2009). *An introduction to qualitative research* (4th ed.). London, UK: Sage.

Flick, U. (2014). Mapping the field. In U. Flick (Ed.), *The Sage Handbook of Qualitative Data Analysis* (pp. 1-18). Los Angeles, CA: Sage.

Fontana, A., & Frey, J. (2005). The interview: From neutral stance to political involvement. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 695-727). Thousand Oaks, CA: Sage.

Forsberg, E., & Wermke, W. (2012). Knowledge sources and autonomy: German and Swedish teachers' continuing professional development of assessment knowledge. *Professional Development in Education, 38*(5), 741-758.

Frey, N., & Fisher, D. (2009). Using common formative assessments as a source of professional development in an urban American elementary school. *Teaching and Teacher Education, 25*(5), 674-680.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly, 9*(2), 113-132.

Fulmer, G. W., Lee, I. C., & Tan, K. H. (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.

Gall, M. D., Gall, J. P., & Borg, W. R. (2006). *Educational research: An introduction* (8th ed.). Boston, MA: Allyn & Bacon.

Gao, X. (2017). Questioning the identity turn in language teacher (educator) research. In G. Barkhuizen (Ed.), *Reflections on language teacher identity research* (pp.189-195). New York, NY: Routledge.

Gil, J. (2016). English language education policies in the People's Republic of China. In R. Kirkpatrick (Ed.), *English language education policy in Asia* (pp. 49-90). New York, NY: Springer.

Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2015). The impact of the feedback source on developing oral presentation competence. *Studies in Higher Education*. Advance Online Publication. DOI: 10.1080/03075079.2015.1117064

Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2016). Fostering oral presentation performance: Does the quality of feedback differ when provided by the teacher, peers or peers guided by tutor? *Assessment & Evaluation in Higher Education.* Advance Online Publication. DOI: 10.1080/02602938.2016.1212984.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded Theory: Strategies for qualitative research*. Chicago, IL: Aldine Publishing Company.

Glesne, C., & Peshkin, A. (1992). *Becoming qualitative researchers: An introduction*. New York, NY: Longman.

Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice, 33*(2), 14-18.

Gottheiner, D. M., & Siegel, M. A. (2012). Experienced middle school science teachers' assessment literacy: Investigating knowledge of students' conceptions in genetics and ways to shape instruction. *Journal of Science Teacher Education, 23*(5), 531-557.

Gould, J., & Day, P. (2013). Hearing you loud and clear: Student perspectives of audio feedback in higher education. *Assessment & Evaluation in Higher Education, 38*(5), 554-566.

Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education, 21*(6), 607-621.

Green, T. F. (1971). *The activities of teaching*. New York, NY: McGraw-Hill.

Greenberg, J., & Walsh, K. (2012). *What teacher preparation programs teach about K-12 assessment: A review.* Washington, DC: National Council on Teacher Quality.

Gu, P. 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: Principles, Policy & Practice, 21*(3), 286-305.

Gullickson, A. R. (1993). Matching measurement instruction to classroom-based evaluation: Perceived discrepancies, needs, and challenges. In S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp. 1-25). Lincoln, Neb: Buro Institute of Mental Measurement, University of Nebraska-Lincoln.

Gunn, A. C., & Gilmore, A. (2014). Early childhood initial teacher education students' learning about assessment, *Assessment Matters, 7*, 24–38.

Guskey, T. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education, 4*(1), 63-69.

Guskey, T. (2011). Five obstacles to grading reform. *Educational Leadership, 69*, 17-21.

Harris, L. R., & Brown, G. T. L. (2009). The complexities of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice, 16*(3), 365–381.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Hendry, G. D., Armstrong, S., & Bromberger, N. (2012). Implementing standards-based assessment effectively: Incorporating discussion of exemplars into classroom teaching. *Assessment & Evaluation in Higher Education, 37*(2), 149–161.

Hennessy, C., & Forrester, G. (2014). Developing a framework for effective audio feedback: A case study. *Assessment & Evaluation in Higher Education, 39*(7), 777–789.

Herman, J., Osmundson, E., Dai, Y., Ringstaff, C., & Timms, M. (2015). Investigating the dynamics of formative assessment: Relationships between teacher knowledge, assessment practice and learning. *Assessment in Education: Principles, Policy & Practice, 22*(3), 344-367.

Hill, M. F., Cowie, B., Gilmore, A., & Smith, L. F. (2010). Preparing assessment-capable teachers: What should pre-service teachers know and be able to do? *Assessment Matters, 2*, 44-64.

Hill, M. F., Ell, F., Grudnoff, L., Haigh, M., Cochran-smith, M., Chang, W., & Ludlow, L. (2016). Assessment for equity: Learning how to use evidence to scaffold learning and improve teaching. *Assessment in Education: Principles, Policy & Practice*, 1–20. DOI: 10.1080/0969594X.2016.1253541

Hill, M. F., Ell, F., Grudnoff, L., & Limbrick, L. (2014). Practise what you preach: Initial teacher education students learning about assessment. *Assessment Matters, 7*, 90–112.

Hill, M. F., Gunn, A., Cowie, B., Smith, L. F., & Gilmore, A. (2014). Preparing primary and early childhood initial teacher education students to use assessment in teaching. *Assessment Matters, 7*, 4–23.

Hoover, N. R. (2009). *A descriptive study of teachers' instructional use of student assessment data* (Unpublished doctoral dissertation). Virginia Commonwealth University, Richmond, VA.

Howley, M. D., Howley, A., Henning, J. E., Gilla, M. B., & Weade, G. (2013). Intersecting domains of assessment knowledge: School typologies based on interviews with secondary teachers. *Educational Assessment, 18*(1), 26-48.

Hoyle, R. H. (2008). Latent variable models of social research data. In P. Alasuutari, L. Bickman, & J. Brannen (Eds.), *The Sage handbook of social research methods* (pp. 395-413). Los Angeles, CA: Sage.

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.

Hu, G. (2005). Using peer review with Chinese ESL stduent writers. *Language Teaching Research, 9*(3), 321-342.

Ice, P., Curtis, R., Phillips, P., & Wells, J. (2007). Using asynchronous audio feedback to enhance teaching presence and students' sense of community. *Journal of Asynchronous Learning Networks, 11*(2), 3-25.

Impara, J. C., Divine, K. P., Bruce, F. A., Liverman, M. R., & Gay, A. (1991). Does interpretive test score information help teachers? *Educational Measurement: Issues and Practice, 10*(4), 16-18.

Impara, J. C., & Plake, B. S. (1995). Comparing counselors', school administrators', and teachers' knowledge in student assessment. *Measurement & Evaluation in Counseling & Development, 28*(2), 78-87.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3), 385-402.

Inbar-Lourie, O. (2013). Language assessment literacy. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2923-2931). Oxford, UK: Blackwell.

International Language Testing Association (ILTA). (2001). *Code of ethics*. Retrived March 3rd, 2015, from: http://www.iltaonline.com/images/pdfs/ILTA_Code.pdf

International Language Testing Association. (2007). *Guidelines for practice*. Retrived March 3rd, 2015 from: http://www.iltaonline.com/images/pdfs/ilta_guidelines.pdf

Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing, 30*(3), 345-362.

Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing, 27*(4), 555-584.

Johnson, V. E. (2003). *Grade inflation: A crisis in college education*. New York, NY: Springer.

Joint Committee on Standards for Education Evaluation (JCSEE). (2015). *Classroom assessment standards for PreK-12 teachers* [Kindle ed.]. Retrieved March 8th, 2015 from: http://www.amazon.com/Classroom-Assessment-Standards-PreK-12-Teachers-ebook/dp/B00V6C9RVO.

Kagan, D. M. (1990). Ways of evaluation teacher cognition: Inferences concerning the Goldilocks Principle. *Review of Educational Research, 60*(3), 419-469.

Kagan, D. M. (1992). Professional growth among preservice and beginning teachers.

Review of Educational Research, 62(2), 129–169.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, CT: Praeger.

King, J. D. (2010). *Criterion-referenced assessment literacy of educators* (Unpublished doctoral dissertation). University of Southern Mississippi, Hattiesburg, MS.

Koh, K. H. (2011). Improving teachers' assessment literacy through professioal development. *Teaching Education, 22*(3), 255-276.

Koh, K., Carol-Ann Burke, L. E., Luke, A., Gong, W., & Tan, C. (2017). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*. DOI: 10.1177/1362168816684366

Kostal, J. W., Kuncel, N. R., & Sackett, P. R. (2016). Grade inflation marches on: Grade increase from the 1990s to 2000s. *Educational Measurement: Issues and Practice, 35*(1), 11-20.

Kuzel, A. J. (1992). Sampling in qualitative inquiry. In B. F. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (Vol. 3, pp. 31-44). Newburry Park, CA: Sage.

Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing, 32*(2), 169-197.

Leahy, S., & Wiliam, D. (2012). From teachers to schools: Scaling up professional development for formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 49-72). London, UK: Sage.

Leighton, J. P., Gokiert, R. J., Cor, M. K., & Heffernan, C. (2010). Teacher beliefs about the cognitive diagnostic information of classroom- versus large-scale tests: Implications for assesment literacy. *Assessment in Education: Principles, Policy & Practice, 17*(1), 7-21.

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

Levy-Vered, A., & Alhija, F. N.-A. (2015). Modelling beginning teachers' assessment literacy: The contribution of training, self-efficacy, and conceptions of assessment. *Educational Research and Evaluation, 21*(5-6), 378-406.

Li, M. H. (2012). English curriculum in higher education in China for non-English majors. In J. Ruan & C. B. Leung (Eds.), *Perspectives on teaching and learning English literacy in China* (pp. 105-114). Dordrecht, The Netherelands: Springer.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverley Hills, CA: Sage.

Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279-290.

Liu, J., & Xu, Y. (in press). Assessment for learning in English language classrooms in China: Contexts, problems, and solutions. In B. Zou, D. Nunan, & H. Reinders (Eds.), *Innovations in language learning and teaching: The case of China*. Basingstoke, UK: Palgrave MacMillan.

Lomax, R. G. (1996). On becoming assessment literate: An initial look at pre-service teachers' beliefs and practices. *The Teacher Educator, 31*(4), 292-303.

Looney, A., Cumming, J., Van der Kleij, F., & Harris, K. (2017). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice,* 1–26. DOI: 10.1080/0969594X.2016.1268090

Lortie, D. (1975). *Schoolteacher: A sociological study*. London, UK: University of Chicago Press.

Love, D. A., & Kotchen, M. J. (2010). Grades, course evaluations, and academic inventives. *Eastern Economic Journal, 36*, 151-163.

Lukin, L. E., Bandalos, D. L., Eckhout, T. J., & Mickelson, K. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice, 23*(2), 26-32.

Lunt, T., & Curran, J. (2016).'Are you listening please?' The advantages of electronic audio feedback compared to written feedback. *Assessment & Evaluation in Higher Education, 35*(7), 759-769.

Lyon, E. G. (2013). Conceptualizing and exemplifying science teachers' assessment expertise. *International Journal of Science Education, 35*(7), 1208-1229.

Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualizations. *Teaching and Teacher Education, 20*(5), 523–535.

Marvasti, A. B. (2014). Analysing observations. In U. Flick (Ed.), *The Sage handbook of qualitative data analysis* (pp. 354–366). Los Angeles, CA: Sage.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). Old Tappan, NJ: MacMillan.

McMillan, J. H. (2001). *Essential assessment concepts for teachers and administrators*. Thousand Oaks, CA: Corwin Press, Inc.

McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice, 22*(4), 34–43.

McMillan, J. H. (2013). Why we need research on classroom assessment. In J. H. McMillan (Ed.) *The Sage handbook of research on classroom assessment* (pp. 1-14). Los Angeles, CA: Sage.

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.

Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education, 33*(1), 49-64.

Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools, 12*(2), 101-113.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Murchison, J. M. (2010). *Ethnography essentials: Designing, conducting and presenting your research*. San Francisco, CA: Jossey-Bass.

Nepal, K. P. (2012). An approach to assign individual marks from a team mark: The case of Australian grading system at universities. *Assessment & Evaluation in Higher Education, 37*(5), 555–562.

Nichols, S. L., & Harris, L. R. (2016). Accountability assessment's effects on teachers and schools. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of social and human conditions in assessment* (pp.40-56). New York, NY: Routledge.

Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education, 35*(5), 501-517.

Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback. *Studies in Higher Education, 31*(2), 199-218.

Nida, E. A., & Taber, C. (1969). *The theory and practice of translation*. Leiden, The Netherlands: E. J. Brill.

Nitko, A. J., & Brookhart, S. M. (2011). *Educational assessment of students* (6th ed.). Boston, MA: Pearson Education, Inc.

O'Donovan, B., Rust, C., & Price, M. (2016). A scholarly approach to solving the feedback dilemma in practice. *Assessment & Evaluation in Higher Education, 41*(5), 938-949.

Orsmond, P., Merry, S., & Reiling, K. (2005). Biology students' utilization of tutor's formative feedback: A qualitative interview study. *Assessment & Evaluation in Higher Education, 30*(4), 369-386.

Orsmond, P., Maw, S. J., Park, J. R., Gomez, S., & Crook, A. C. (2013). Moving feedback forward: Theory to practice. *Assessment & Evaluation in Higher Education, 38*(2), 240–252.

O'Sullivan, R. G., & Chalnick, M. K. (1991). Measurement-related course work requirements for teacher certification and recertification. *Educational Measurement: Issues and Practice, 10*(1), 17-19, 23.

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.

Patri, M. (2002). The influence of peer feedback on self and peer-assessment of oral skills. *Language Testing, 19*(2), 109-131.

Patton, M. Q. (2002). *Qualitative research & evaluation methods*. Thousand Oaks, CA: Sage.

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.

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

Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice, 12*(4), 10-12, 39.

Pope, N., Green, S. K., Johnson, R. L., & Mitchell, M. (2009). Examining teacher ethical dilemmas in classroom assessment. *Teaching and Teacher Education, 25*, 778-782.

Popham, W. J. (1991). Appropriateness of teachers' test-preparation practices. *Educational Measurement: Issues and Practice, 10*(4), 12-15.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice, 48*(1), 4-11.

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

Price, M., Handley, K., & Millar, J. (2011). Feedback: Focusing attention on engagement. *Studies in Higher Education, 36*(8), 879-896.

Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect? *Assessment & Evaluation in Higher Education, 35*(3), 277-289.

Quilter, S. M., & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator, 36*, 115-131.

Randall, J., & Engelhard, G. (2009). Examining teacher grades using Rasch measurement theory. *Journal of Educational Measurement, 46*, 1–18.

Randel, B., & Clark, T. (2013). Measuring classroom assessment practices. In J. H. McMillan (Ed.), *The Sage handbook of research on classroom assessment* (pp. 145-163). Los Angeles, CA: Sage.

Raykov, T., & Marcoulides, G. A. (2007). *A first course in structural equation modeling.* New York, NY: Psychology Press.

Rojstaczer, S., & Healy, C. (2012). Where A is ordinary: The evoluation of American college and unviersity grading, 1940-2009. *Teachers College Record*, *114*(7), 1-23.

Rorsyth, R., Cullen, R., Ringan, N., & Stubbs, M. (2015). Supporting the development of assessment literacy of staff through institutional process change. *London Review of Education, 13*(3), 34-41.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36.

Rotheram, B. (2009). Sounds good: Using audio to give assessment feedback. *The Assessment, Learning and Teaching Journal, 7*(4), 22-24.

Rudner, L. M. (2012). PARAM Calibration Software for the 3 Parameter Logistic IRT Model [freeware] (Version 0.93). Retrieved on Dec 15, 2014, from http://echo.edres.org:8080/irt/param/

Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education, 28*(2), 147-164.

Sadler, D. R. (2009). Grade integrity and the representation of academic achievement. *Studies in Higher Education, 34*(7), 807–826.

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education, 35*(5), 535-550.

Sato, M., Wei, C. R., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal, 45*(3), 669-700.

Savin-Baden, M. (2010). The sound of feedback in higher education. *Learning, Media and Technology, 35*(1), 53–64.

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing, 30*(3), 309-327.

Schafer, W. D. (1991). Essential assessment skills in professional education of teachers. *Educational Measurement: Issues and Practice, 10*(1), 2-6,12.

Schafer, W. D. (1993). Assessment literacy for teachers. *Theory into Practice, 32*(2), 118-126.

Schafer, W. D., & Lizzitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. *Journal of Teacher Education, 38*(3), 57-63.

Schneider, M. C., & Randel, B. (2010). Research on characteristics of effective professional development programs for enhancing educators' skills in formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 251-276). New York, NY: Routledge.

Schreier, M. (2014). Qualitative content analysis. In U. Flick (Ed.), *The Sage handbook of qualitative data analysis* (pp. 170-183). Los Angeles, CA: Sage.

Schön, D. (1983). *The reflective practitioner: How professional think in action*. New York, NY: Basic Books.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–22.

Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education, 22*(4), 371-391.

Silverman, D. (2006). *Interpreting qualitative data: Methods for analyzing talk, text, and interaction*. London, UK: Sage.

Smith, K. (2011). Professional development of teachers—A prerequisite for AfL to be successfully implemented in the classroom. *Studies in Educational Evaluation, 37*(1), 55-61.

Smith, L. F., & Galvin, R. (2014). Toward assessment readiness: An embedded approach in primary initial teacher education. *Assessment Matters, 7*, 39–63.

Smith, L. F., Hill, M. F., Cowie, B., & Gilmore, A. (2014). Preparing teachers to use the enabling power of assesment. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 418-445). Dordrecht, The Netherlands: Springer.

Stiggins, R. J. (1991a). Assessment literacy. *Phi Delta Kappan, 72*, 534-539.

Stiggins, R. J. (1991b). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice, 10*(1), 7-12.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan, 77*(3), 238-245.

Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice, 18*(1), 23-27.

Stiggins, R. J. (2010). Essential formative assessment competencies for teachers and school leaders. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 233-250). New York, NY: Taylor & Francis.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage.

Sun, Y., & Cheng, L. (2014). Teachers' grading practices: meaning and values assigned. *Assessment in Education: Principles, Policy & Practice, 21*(3), 326-343.

Swaffield, S. (2011). Getting to the heart of authentic Assessment for Learning. *Assessment in Education: Principles, Policy & Practice, 18*(4), 433-449.

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics, 29*, 21-26.

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing, 30*(3), 403-412.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp.127-146). New York, NY: Macmillan.

Tierney, R. D. (2013). Fairness in classroom assessment. In J. H. McMillan (Ed.), The *Sage handbook of research on classroom assessment* (pp. 125-144). Thousand Oaks, CA: Sage.

Tierney, R. D. (2015). Altered grades: A grey zone in the ethics of classroom assessment. *Assessment Matters, 8*, 5-30.

van den Berg, B. (2002). Teachers' meanings regarding educational practice. *Review of Educational Research, 72*, 577-625.

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of an European study. *Language Assessment Quarterly, 11*(4), 374-402.

Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education, 30*(3), 749-770.

Wang, H. (2009). A survey of college English teachers and teaching. *Foreign Language World, 4*, 6-13.

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

Wen, Q. (2012). Teaching English as an international language in mainland China. In A. Kirkpartrick & R. Sussex (Eds.), *English as an international language in Asia: Implications for language education* (pp.79-93). Dordrecht, The Netherlands: Springer.

Wiliam, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18-40). New York , NY: Routledge.

Willig, C. (2008). *Introducing qualitative research in psychology: Adventures in theory and method*. London, UK: Open University Press.

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualizing teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher, 40*(2), 241-256.

Wise, S. L., Lukin, L. E., & Roos, L. L. (1991). Teacher beliefs about training in testing and measurement. *Journal of Teacher Education, 42*(1), 37-42.

Woolfolk, H. A., Hoy, W. K., & Davis, H. A. (2009). Teachers' self-efficacy beliefs. In K. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 627-655). Mahwah, NJ: Lawrence Erlbaum.

Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice, 17*(1), 59-75.

Xie, H. (Ed.). (2011). *Education statistics yearbook of China*. Beijing, China: People's Education Publishing House.

Xu, Y. (2016). Teacher assessment planning within the context of university English language teaching (ELT) in China: Implications for assessment literacy. *Australia Review of Applied Linguistics, 39*(3), 233-254.

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 teachers assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment, 6*(1), 133-159.

Xu, Y., & Carless, D. (2016). 'Only true friends could be cruelly honest': Cognitive scaffolding and social-affective support in teacher feedback literacy. *Assessment & Evaluation in Higher Education*. DOI: 10.1080/02602938.2016.1226759.

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.

Yang, M., Badger, R., & Yu. Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing, 15*(3), 179-200.

Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education, 18*(3), 285-297.

Yin, R. Y. (2009). *Case study research: Design and methods* (4th ed.). Los Angeles, CA: Sage.

Yorke, M. (2003). Formative assesment in higher education: Moving towards theory and the enhancement of pedagogic practice. *Higher Education, 45*(4), 477-501.

Zeng, K. (1999). *Dragon Gate: Competition examinations and their consequences.* London, UK: Cassell.

Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education, 16*(4), 323-342.

Zheng, D. (2010). An investigation into the assessment literacy of secondary and primary school teachers: A report from Z province. *Global Education Review, 39*(2), 31-36.