Indonesian EFL Journal, Vol. 1(1) January 2015
ISSN 2252-7427

AISEE
The Association of Indonesian Scholars of English Education

AN INVESTIGATION OF CHINESE MIDDLE SCHOOL IN-SERVICE ENGLISH TEACHERS' ASSESSMENT LITERACY

Lin Dunlai
School of Foreign Languages and Literature; National Innovation Center for Assessment of Basic Education Quality, Beijing Normal University
E-mail: lindunlai@bnu.edu.cn

Su You
School of Humanities, Beijing University of Posts and Telecommunications
E-mail: suyou@bupt.edu.cn

APA Citation: Lin, D. & Su, Y. (2015). An investigation of Chinese middle school in-service English teachers' assessment literacy. *Indonesian EFL Journal 1*(1), 1-10

Received: 04-08-2014 Accepted: 10-10-2014 Published: 01-01-2015

Abstract: This paper reports an investigation into the status quo of assessment literacy of Chinese middle school in-service English teachers. Using tasks designed by Coombe et al. (2007), the study finds out that Chinese secondary English teachers have low levels of assessment literacy. They are not aware of such principles as authenticity, sensitivity issues about test content and self-assessment. Especially, they are extremely incompetent in understanding statistics about item analysis and distractor efficiency analysis. No significant difference was detected about teachers' assessment literacy in terms of teaching experience and whether they have taken assessment training courses in any forms. The authors call for a study into language assessment courses offered for secondary English teachers and enough attention paid to the relevance of language assessment training courses to classroom assessment.
Keywords: *assessment literacy, middle school in-service English teachers. China.*

INTRODUCTION

Teachers spend up to one third or even half of their career life in dealing with assessment (Stiggins, 1991a; 1999). It is no doubt that teachers play a pivotal part in classroom assessment. Besides, research has found that if assessment is applied appropriately, it would boost student learning (Black & Wiliam, 1998). The current beliefs about the prominent role of assessment in fostering learning, referred to as "assessment *for* learning" (Gipps, 1994; Broadfoot & Black, 2004), call for the knowledge required for conducting assessment activities, known as "assessment literacy".

The term "assessment literacy" was first coined by Rick Stiggins (1991b). He first described what assessment illiterates cannot do in Stiggins (1991b) and was more straightforward in delineating what assessment literate teachers can do in Stiggins (1995). Popham (2011) defines assessment literacy as "(it) consists of an individual's *understandings* of the *fundamental assessment concepts and procedures* deemed *likely to influence educational decisions*" (italics in original text). Language testing arena didn't take up the term until 2009 when Taylor (2009) called for the sharing of the language testing knowledge, skills and understanding in wider circles. She argues that "training for assessment literacy entails an appropriate balance of technical know-how, and understanding of principles, but all firmly contextualized within a sound understanding of the role and function of assessment within education and society" (ibid: 27). Inbar-Lourie (2008) also perceived language assessment literacy as encompassing layers of assessment literacy and language specific elements. In this article, the authors take a classroom-oriented layer of assessment literacy, which suggests that "in order to become literate in language assessment, one needs to attain knowledge in formative and summative testing and assessment methods, in interpreting student scores, in understanding the complexities of

validity and reliability including current tensions which question the application of traditional psychometric measures to teacher-based assessment" (Inbar-Laurie, 2013; Teasdale & Leung, 2000).

As a matter of fact, the knowledge base for assessment has been offered by official documents. The American Federation of Teachers (AFT), National Council on Measurement in Education (NCME) and National Education Association (NEA) identified seven components that formed the base required for performing assessment tasks. The seven components are (1) Teachers should be skilled in choosing assessment methods appropriate for instructional decisions; (2) Teachers should be skilled in developing assessment methods appropriate for instructional decisions; (3) Teachers should be skilled in administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods; (4) Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement; (5) Teachers should be skilled in developing valid pupil grading procedures that use pupil assessments; (6) Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators; (7) Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information (AFT, NCME, & NEA, 1990).

But the real situation seems to lag too much behind. Many researchers have consistently found that teachers lack assessment literacy (Arter, 2001; Mertler, 2004; Mertle & Campbell, 2005; Popham, 2006; Wang, Wang & Huang, 2008; Lin, 2014), which makes it simply impossible to build up assessment culture. The methods that were adopted in the existing literature were mostly survey, which often includes the instrument developed by Plake, Impara & Fager (1993), which is a 35-multiple-choice test developed from the seven standards (AFT, NCME, & NEA, 1990), mentioned above. Plake and Impara (1997) carried out a national survey over 555 American teachers, and found "woefully" low levels of assessment competence. Campbell, Murphy and Holt (2002) applied the instrument on pre-service teachers and found incompetence in assessment. Mertler (2004) investigated 61 in-service and 101 pre-service secondary teachers in the US, and found incompetency in grading and interpreting test results. Zhang and Burry-Stock (1997) used the assessment inventory to find out 7 factors of assessment literacy. They compared teachers with different teaching experience and assessment training courses and found significant difference of assessment literacy perception both in terms of teaching experience and assessment training courses.

Empirical research into language assessment literacy is still rare (Lin & Wu, 2014). Existent research focuses on the knowledge base of language assessment literacy (e.g. Bailey & Brown, 1996; Brown & Bailey, 2008; Tsagari, 2011; Fulcher, 2012; Jeong, 2013). In Chinese context, Jin (2010) investigated 86 language testing courses for teacher preparation programs for universities across China and found a heavy focus on testing and measurement perspective rather than assessment perspective. As for secondary English teachers' assessment literacy, Lin (2014) found low levels of assessment literacy of the teachers with both a quantitative and qualitative design. In this study, the authors want to report a study on the secondary English teachers' assessment literacy based on a test. The research questions for the study are: (1) What is the status quo of assessment literacy of Chinese middle-school in-service English teachers? (2) Does teaching experience make a difference? (3) Does assessment training course make a difference?

METHOD

The sample

This study took the convenience sampling method due to the lack of resources. 39 middle school in-service English teachers (N=39) took part in the test. These teachers were attending a language assessment course as part of their Master of Education program. 33 (N=33) were female, and 6 (N=6) were male. They were from 14 provinces (or municipalities) (There are 34 provincial administrative zones in China). In terms of age, 22 were below 30 years old, 11 were at

the age of 31-35, and 6 were above 36 years old. As regards the teaching experience, 18 of them had been teaching for 2-5 years, 12 of them 6-10 years and 9 of them more than 11 years. 35 of them graduated from a teacher training program. As for language proficiency, 26 passed Test for English Majors Band 8 (TEM 8), which suggests very high English language proficiency. 14 of them were junior high school teachers and 25 taught senior high. As to class size, 6 of them taught classes with fewer than 30 students, 7 of them with 30-40 students, 10 of them with 41-50 students, and 16 of them with more than 50 students. As for these teachers' professional development in assessment, 20 of them never took language assessment courses in any form, 12 of them took a complete language assessment course, 7 of them got to know language assessment through lectures or language teaching methods course. As for their perception about the importance of knowledge about language testing, 34 of them thought knowledge about language assessment very important for an English teacher.

The instrument

The instrument for this study was adapted from a book written by Coombe, Folse and Hubley (2007). Altogether the instrument includes ten tasks. The first task tests the participants' knowledge about language assessment. It required the participants to select one best answer. An example is as follows:

It's the beginning of the semester, and you have a mixed-level class. You want to get an idea of the class's strengths and weaknesses before you plan your lessons. Which kind of test would give you the information you need?
A. Placement. *B. Diagnostic.*
C. Proficiency. *D. Aptitude.* *(Key: B)*

The second and third tasks present two scenarios of an English teacher conducting language testing in a semester. The participants were required to underline where there are some inappropriate language testing practices and briefly explain why.

The fourth task is about techniques for multiple choice questioning. The participants were to name the defects in the multiple choice question setting. An example is as follows:

An architect is a person who does not _____.
a. design automobiles
b. design buildings
c. design houses
d. design offices
(suggested answer: 1. "design" should be provided in the stem to avoid repetition. 2. We normally do not define something as "not"—authenticity issue)

The fifth to eighth tasks provide a short reading, writing, listening and speaking test respectively. Participants were required to point out where a violation turns up regarding testing techniques.

The ninth task is a scenario about an English teacher's test preparation practice for students. Participants were required to underline where inappropriate and give some comments.

The tenth task is for participant to read an item analysis and distractor analysis report and interpret the results and suggest some improvements. Table 1 indicates the tasks and its contents.

Table 1. Ten tasks, its contents and question types

1	2	3	4	5	6	7	8	9	10
knowledge	Scenario	scenario	MC	reading	writing	listening	speaking	test pre- paration	statistics
MC	Open	open	open	open	open	open	open	open	open

The data collection process

The test was administered at the very beginning of the above-mentioned language assessment course to avoid any learning. In other words, the participants should rely on their prior knowledge to finish the test. The tasks were printed as booklets to avoiding misplacing of the test paper. Space was more than enough for the participants, so they can come up with ideas and write freely. Participants were notified that the test was only for research purpose so they should not hesitate to write down their response exactly as what they thought about the issues. The

test lasted for 2 hours. To avoid fatigue, a ten-minute break was called in. All the test papers were collected, coded and kept safely by the researchers.

The data processing method

The scoring of the test was done by the authors collaboratively. The marking scheme was developed based on Coombe et at. (2007)'s reference answers to the tasks and the participants' real performance. The authors worked out the indicators to serve as the basis for marking. The numbers of indicators for each task was shown in Table 2. Altogether 123 indicators were worked out. Each indicator was marked as 0, 1, except Task 2 and Task 4, where partial credit was necessary and the indicators were marked as 0, 0.5, 1. As Task 2 and Task 4 involve more judgment to be made, the researchers used double marking. The Pearson correlation was calculated, and the results were 0.945 and 0.943, respectively, which suggest very high agreement between the two researchers. The scores from the two researchers were averaged out to reach the final markings for Task 2 and Task 4.

Table 2. *Ten tasks and number of indicators for each task*

1	2	3	4	5	6	7	8	9	10
knowledge	Scenario	scenario	MC	reading	writing	listening	speaking	test pre- paration	statistics
10	16	15	13	20	10	15	10	8	6

FINDINGS AND DISCUSSION

Research Question 1: The status quo of assessment literacy

The reliability check of the test shows that the Cronbach α=0.828, showing very high level of internal consistency. As indicated in Figure 1, out of 123, the highest score is 65.25, the lowest is 15.5, the average score is 33.22, and the standard deviation is 9.397. The facility index of the test is 0.27, showing that the test is very difficult for these secondary teachers. Table 3 shows *IF* of each task.

Table 3 Ten tasks and *IF* of each task

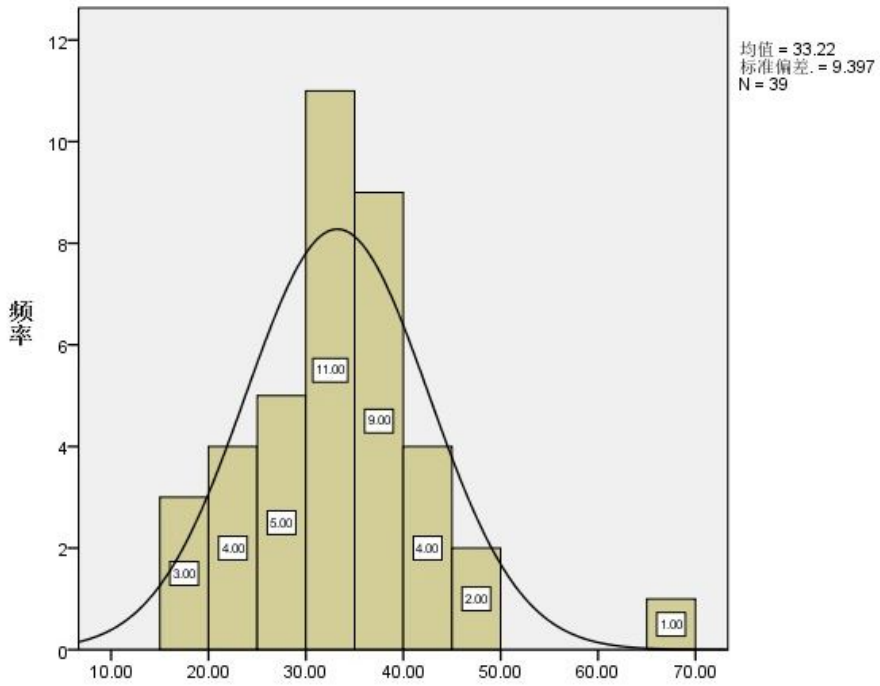
1	2	3	4	5	6	7	8	9	10
knowledge	Scenario	scenario	MC	reading	writing	listening	speaking	test pre- paration	statistics
0.36	0.29	0.56	0.24	0.17	0.26	0.12	0.36	0.35	0.06

As we can see in Table 3, six tasks (*IF*<0.3) are extremely difficult for the participants, with the final task—interpreting statistics—as the most difficult one. The techniques for making multiple choice questions are not mastered by the teachers, despite the ubiquity of MC questions in almost every kind of test. Compared with testing productive skills (writing and speaking), testing receptive skills (reading and listening) is more difficult for the teachers. Teachers are totally statistically illiterate for item analysis.

A comparison study was done between teachers with different length of teaching experience by ANOVA. The result shows that *F*-value is 0.262 (*p*=0.771), suggesting no significant difference between teachers of different length of teaching experience.

Another comparison study was done between teachers with different levels of assessment training by ANOVA. The result shows that *F*-value is 0.865 (*p*=0.468), suggesting no significant difference between teachers with different levels of assessment training experience.

Figure 1 Total score distribution of the test



Research questions 2 & 3: The role of teaching experience and assessment training course

Low levels of assessment literacy

Using mostly open-ended questions, this study resonates with other existent research (e.g. Arter, 2001; Mertler, 2004; Mertler & Campbell, 2005; Popham, 2006; Wang, Wang & Huang, 2008; Lin, 2014) that secondary English teachers lack a desirable level of assessment literacy. Pill and Harding (2013) conceptualized different levels of assessment literacy as illiteracy, nominal literacy, functional literacy, procedural and conceptual literacy and multidimensional literacy. Table 5 shows the meaning of the five levels of literacy.

Table 5. *Five levels of assessment literacy (Adapted from Pill & Harding, 2013: 383)*

Illiteracy	**ignorance of language assessment concepts**
Nominal literacy	Understanding that a specific term relates to assessment, but may indicate a misconception
Functional literacy	Sound understanding of basic terms and concepts
Procedural and conceptual literacy	Understanding central concepts of the field and using knowledge in practice
Multidimensional literacy	Knowledge extending beyond ordinary concepts including philosophical, historical and social dimensions of assessment

Judging from teachers' responses to the ten tasks and the above scale, the assessment literacy of Chinese middle school English teachers can be rated as somewhere between illiteracy to nominal literacy. The latest *National English Curriculum Standards* for compulsory education (Ministry of Education, or MoE, 2012) stipulates that teachers should make use of various kinds of assessment to evaluate students' development, including formative assessment and summative assessment. It includes exemplary assessment tasks and comments for teachers' reference, covering as many as 45 pages. Obviously, teachers lacking assessment literacy will hinder the sound implementation of the *Standards.* As Alderson (2011) argues, "testing is too important to be left to testers".

Here, the authors want to point out some prominent issues in testing that the

participants were not aware of. The first issue is authenticity in language assessment. Authenticity is defined as "the degree of correspondence of the characteristics of a given language test task to the features of a TLU (Target Language Use) task" (Bachman & Palmer, 1996:23). In Task 2, the scenario describes a novice teacher asking students to write an essay about the use of modal verbs: "Write a 300-word essay on the meanings of modal verbs and their stylistic uses. Give examples and be specific." In Task 6, students are asked to describe a petrol pump. In Task 7, the listening material is about an introduction to a made-up place. All these tasks violate the principle of authenticity. But none of the participants in this study commented on this principle in this test. The second issue is about fairness review. McNamara and Roever (2006: 129) points out that to avoid or reduce differential item functioning (DIF), test makers use "sensitivity review" at the early stages of test creation. This should also be relevant in classroom context. In Task 5, the reading passage is about alligators attacking and claiming people's life. Only one participant pointed out that this text may arouse negative feeling in students and should be avoided in classroom test. Another issue the authors want to point out is that teachers do not attach importance to students' self-assessment. In Task 9, a teacher says "I usually don't like students to mark their own papers", only 3 teachers thought that students should be given opportunities to self-assess their performances and make appropriate adjustment of their study.

Anxiety for statistics

In this study, we found that statistics for item analysis is extremely difficult for those in-service secondary English teachers. 7 of the participants didn't write any answer in this part, while 16 of them indicated clearly that they didn't know the answer. The other participants gave wild guessing. Typical answers were "The *IF* is 0.77. It's too hard"; "The discrimination index is 0.61. It's too low and needs to be dropped". We know from Ebel and Frisbie (1991: 232) that as regards index of discrimination, "0.4 and up suggests very good items; 0.30-0.39 suggests reasonably good but possibly subject to improvement; 0.20-0.29 suggests marginal items, usually needing and being subject to improvement; and below 0.19 suggests poor items to be rejected or improved by revision". The participants were not able to pinpoint the distractive power of different options and make suggestions for improvements.

As so many participants (23 out of 39) claimed ignorance in statistics, we find it urgent to look into the issue of statistics anxiety, which is defined by Brown (2013:353) as " a complex of behaviors, including uneasiness, trepidation, nervousness, and even debilitating fear, that may occur in some students when they are confronted with studying or using statistics". Despite statistics anxiety, Brown (2012) believes that though Classical Test Theory has its disadvantages, it is still "sufficiently accurate, easy-to-learn, and practical to continue in use for years to come in real (especially local) testing situations" (p.334). He further points out that classical item analysis and distractor efficiency analysis will continue to provide useful feedback to item writers and test developers about their items and their test specifications.

So it is crucial and practical to equip English teachers with Classical Test Theory. In this regard, Brown's (2013) proposal of *need-to-know approach* is appropriate for developing teachers, i.e., teacher trainers should analyze what needs to be acquired for classroom teachers, and make statistical instruction accessible and manageable for the secondary teachers.

Teaching experience and assessment training

In this study, no significant difference was detected about teachers' assessment literacy as regards to teaching experience and whether they have taken any form of assessment training course. This is in contradiction to the study carried out by Zhang and Burry-Stock (1997). But the problem is Zhang and Burry-Stock's research used teachers' self-perception rather than a test like the current study, which may be influenced by teachers' over-confidence in their own assessment literacy (cf. Wise & Lukin, 1993).

Here it is very important to examine the nature of assessment literacy. Wang et al.

(2008) takes assessment literacy as a part of the package of pedagogical content knowledge (PCK), which has been introduced as an element of the knowledge base for teaching (Schulman, 1986). Based on Cochran, DeRuiter and King (1993), who thought Schulman's concept of PCK did not account for teachers' initiative in developing their own PCK and put forward pedagogical content knowing (PCKg) instead, Lin (2014) conceptualizes assessment literacy as a part of PCKg. It both emphasizes pre-service development on language assessment and in-service development. What's more important, it calls for teachers' reflection on the scene and take into consideration the local social-cultural environment. Assessment literacy should be acquired by doing assessment.

To make sure teachers develop desirable assessment competence, teacher preparation programs and teacher certifying institutions should pay more attention to developing teachers' assessment literacy. Early study showed that assessment training was largely neglected in teacher preparation programs (Noll, 1955; Schafer & Lissitz, 1987; Gullickson,1984; Stiggins & Conklin, 1988,1989; Wise & Lukin,1993). These studies were carried out in American context and in an early time, but the situation seems to have not improved too much in China based on the authors' preliminary research on pre-service language assessment courses. There is no published research in Chinese context except Jin (2010). But Jin's research was about teacher preparation programs for university English Teachers. A study is urgently needed about language assessment courses of teacher preparation programs for secondary teachers.

Another issue that should have people's attention is the relevance of language assessment course for secondary teachers. Jin's (2010) study found a heavy focus on testing rather than classroom assessment. As early as 1991, Stiggins (1991a) called people's attention to the mismatch between assessment training and classroom uses of assessment. He put forward a 30-contact-hour assessment training framework. As it is still relevant today, we quote it here with some amendments to suit language assessment. Session 1 is to make teachers aware of the meaning of quality assessment and why it is so critical to students' well-being. Session 2 is to show the importance of designing assessments with a clear vision of the achievement targets. Session 3 offers instruction in the design and use of paper and pencil assessment instruments. Session 4 addresses the assessment of the four different skills. Session 5 illustrates the use of observation and professional judgment as classroom assessments. Session 6 takes the writing assessment example and shows how it can be expanded to provide a methodology that can be applied to the observation and judgment of any achievement-related behavior or product. Session 7 deals with the assessment of affect. Session 8 is to develop sound grading practices. Session 9 addresses the norm-referenced standardized achievement tests. Session 10 is to come back to quality of assessment and common pitfalls.

CONCLUSION

This study shows that in Chinese context, middle school in-service English teachers lack language assessment literacy that is implicated by the *National English Curriculum Standards* (MoE, 2012). There is no significant difference of level of assessment literacy between teachers with different length of teaching experience and teachers with different levels of assessment training, which shows that language assessment literacy does not grow with more teaching, neither will it grow with simply taking a short assessment training course. A combination of pre-service and in-service teacher development of assessment literacy is expected. Assessment training materials and methods should be aligned to teachers' classroom practice to render for beneficial harvests.

References

Alderson, J. C. (2011). Testing is too important to be left to testers. In J. C. Alderson (Ed.), *A Lifetime of Language Testing* (pp. 219-239). Shanghai: Shanghai Foreign Language Education Press.

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*. Washington, DC: Buros Institute.

Arter, J. (2001). Learning teams for classroom assessment literacy. *NASSP Bulletin 85*(621), 53-65.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing*

in practice. Oxford: Oxford University Press.
Bailey, K. M., & Brown, J. D. (1996). Language testing courses: What are they? In A. Cumming & R. Berwick (Eds.), *Validation in Language Testing* (pp. 236–256). London, UK: Multilingual Matters.
Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice 5*(1), 7-74.
Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of *Assessment in Education*. *Assessment in Education 11*(1), 7-27.
Brown, J. D. (2012). Classical test theory. In G. Fulcher & F. Davidson (Eds.), *The Routledge Handbook of Language Testing* (pp. 323-335). London and New York: Routledge.
Brown, J. D. (2013b). Teaching statistics in language testing courses. *Language Assessment Quarterly 10*(3), 351-369.
Brown, J. D., & Bailey, K. M. (2008). Language testing courses: What are they in 2007? *Language Testing 25*(3), 349-383.
Campbell, C., Murphy, J. A., & Holt, J. K. (2002). Psychometric analysis of an assessment literacy instrument: Applicability to pre-service teachers. Paper presented at the Mid-Western Educational Research Association, Columbus, OH.
Cochran, K. F., DeRuiter, J. A., & King, R. A. (1993). Pedagogical content knowing: An integrative model for teacher preparation. *Journal of Teacher Education 44*(4), 263-272.
Coombe, C., Folse, K. S. & Hubley, N. (2007). *A practical guide to assessing English language learners*. Ann Arbor: University of Michigan Press.
Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th Ed). Englewood Cliffs, NJ: Prentice Hall.
Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly 9*(2), 113-132.
Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London, England: Falmer Press.
Gullickson, A. R. (1984). Teacher perspectives of their instructional use of tests. *Journal of Educational Researc*h *77*(4), 244-248.
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). Wiley-Blackwell.
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.
Lin, D. (2014). *A study on Chinese middle school English teachers' assessment literacy* [Unpublished Doctoral Dissertation]. Beijing: Beijing Normal University.
Lin, D., & W. Z. (2014). New development of research on language assessment literacy. *Modern Foreign Languages 37*(5), 711-720.
McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Oxford, England: Blackwell.
Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education 33*(1), 49-64.
Mertler, C. A., & Campbell, C. (2005). Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory. Paper presented at the annual meeting of the American Educational Research Association, Quebec, Canada.
Ministry of Education, (2012). *National English Curriculum Standards (2011 ed.).* Beijing: Beijing Normal University Press.
Noll, V. H. (1955). Requirements in educational measurement for prospective teachers. *School and Society 80*, 88-91.
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. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp.53-68). London: Academic Press.
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.
Popham, W. J. (2006). Needed: A dose of assessment literacy. *Educational Leadership 63*(6), 84-85.
Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator 46*(4), 265-273.
Schafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel: recommendations and reality. *Journal of Teacher Education 38*, 57-63.
Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher 15*(2), 4-14.
Stiggins, R. J. (1991a). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice 10*(1), 7-12.
Stiggins, R. J. (1991b). Assessment literacy. *The Phi Delta Kappan 72*(7), 534-539.
Stiggins, R. J. (1995). Assessment literacy for the 21st century. *The 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., & Conklin, N. F. (1988). *Teacher training in assessment*. Portland, OR: Northwest Regional Educational Laboratory.
Stiggins, R. J., & Conklin, N. F. (1989). Teacher training in assessment. *Paper presented at the annual meeting of the National Council on Measurement in Education*, San Francisco.
Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics 29*, 21–36.
Teasdale, A., & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing 17*(2), 163-184.

Tsagari, D. (2011). Investigating the 'assessment literacy' of EFL state school teachers. In D. Tsagari & I. Csepes, (Eds.), *Classroom-Based Language Assessment* (169-190). Frankfurt am Main: Peter Lang.

Wang, T. H., Wang, K. H., & Huang, S. C. (2008). Designing a web-based assessment environment for improving pre-service teacher assessment literacy. *Computers & Education*, 51, 448-462.

Wise, S. L., & Lukin, L. E. (1993). Measurement training in Nebraska teacher education programs. In S. L. Wise (Ed.), *Teacher Training in Measurement and Assessment Skills* (pp.187-202). Lincoln, NE: Buros Institute of Mental Measurements.

Zhang, Z., & Burry-Stock, J. (1997). Assessment practices inventory: A multivariate analysis of teachers' perceived assessment competence. *Paper presented at the Annual Meeting of the National Council on Measurement in Education, March 1997, Chicago,* IL.

Berapa Harga Dalam Celana?

An Australian man visited a supermarket in Indonesia. He had just started learning Bahasa Indonesia, so he could not speak very fluently. He wanted to buy some underwear so asked the shop-assistant in Indonesian, "*Berapa harga dalam celana*? (How much for underneath the pants?). This guy thought that the Indonesian language had the same structure as English! Of course the shop-assistant laughed and the Australian man had to ask him for clarification. Now he understands about the structure of Indonesian! (Alfons Arsai, source: http://www.ialf.edu/dpdf/april05page1.html)