

6-1-2007

ESL/EFL Instructors' Beliefs about Assessment and Evaluation

W. Todd Rogers

University of Alberta, todd.rogers@ualberta.ca

Liyong Cheng

Queen's University, liyong.cheng@queensu.ca

Huiqin Hu

Data Recognition Corporation, hhu@datarecognitioncorp.com

Follow this and additional works at: <http://ir.lib.uwo.ca/cie-eci>

Recommended Citation

Rogers, W. Todd; Cheng, Liyong; and Hu, Huiqin (2007) "ESL/EFL Instructors' Beliefs about Assessment and Evaluation," *Canadian and International Education / Education canadienne et internationale*: Vol. 36: Iss. 1, Article 4.

Available at: <http://ir.lib.uwo.ca/cie-eci/vol36/iss1/4>

This Research paper/Rapport de recherche is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Canadian and International Education / Education canadienne et internationale by an authorized administrator of Scholarship@Western. For more information, please contact kmarshal@uwo.ca.

ESL/EFL Instructors' Beliefs about Assessment and Evaluation

W. Todd Rogers (University of Alberta)

Liyang Cheng (Queen's University)

Huiqin Hu (Data Recognition Corporation)

Abstract

The beliefs of 95 ESL/EFL instructors in Canada, 44 in Hong Kong, and 124 in Beijing about assessment and evaluation were examined with 32 questionnaire items. While the results revealed more similarities than the differences, among the instructors in the three contexts, the beliefs expressed by the instructors in the three contexts were somewhat mixed and, at times, contradictory. While the beliefs that assessment and evaluation were important for instruction and help improve student learning and the actual purposes of and uses of assessment and evaluation held by the instructors were positively related. The instructors' beliefs about how they conducted their assessments and evaluations, the time required for assessments and evaluations, and their understanding of and preparation for assessment and evaluation were only somewhat related to their actual assessment practices.

Résumé

Nous avons examiné les convictions sur les jugements et les évaluations de 95 instructeurs ELS/ELE au Canada, 44 à Hong Kong, et 124 à Pékin avec un questionnaire de 32 points. Alors que les résultats font apparaître plus de similarités que de différences parmi les instructeurs dans les trois milieux, la confiance exprimée par les instructeurs est plus ou moins mêlée et parfois contradictoire dans ces trois milieux. Pour ces instructeurs, il y a une corrélation positive entre leur confiance que les évaluations et les jugements sont importants pour l'enseignement et aident les élèves à améliorer leur apprentissage et leur confiance sur le but réel et l'utilisation des jugements et des évaluations. Mais la conviction sur la façon avec laquelle ils conduisent les jugements et les évaluations, le temps nécessaire pour les conduire et leur compréhension comme leur préparation des jugements et évaluations n'ont qu'une relation assez vague avec leurs pratiques des évaluations.

Acknowledgements

Support for the project was made possible in part through funds from Social Sciences and Humanities Research Council (SSHRC) of Canada. We would like to express our sincere thanks to Jo Lewkowicz (University of Hong Kong) and Yi'an Wu (Beijing Foreign Studies University) for coordinating the project in Hong Kong and Beijing. We would also like to thank Margaret Thomason and Xiaoying Wang for their contribution to different stages of this project as research assistants at Queen's University. Thanks also go to the many instructors

who took the time and effort in responding to our survey in the three settings. Without them, we would not have been able to complete the study

Introduction

Assessment and evaluation play a central and important role in teaching and learning. Every model of the teaching-learning process requires that school teachers and university instructors base their decisions – instructional, grading, and reporting – on the degree to which their students have progressed toward and achieved desired learning outcomes (Anderson, 1989, 1990; Educational Testing Services, 1995; Wilson, 2000). Consequently, teachers devote between a quarter and a third of their instructional time creating assessment instruments and observation procedures, marking, synthesizing results, and reporting (Rogers, 1991).

Definition of Belief

It has been suggested that teachers' and instructors' beliefs about assessment and evaluation can directly affect how they design and implement their student assessments and evaluations and how they interpret the results. In the research literature regarding teachers' beliefs, various terms (e.g., attitudes, opinions, perceptions, and rules of practice) have been used. The term *belief*, as used in this study, corresponds to the definition proposed and used by Pajares (1992): “an individual’s judgment of the truth or falsity of a proposition” (p. 316). In the present study, the propositions comprise the functions and forms of assessment and evaluation as perceived by university instructors in their classrooms within the contexts of English as a Second Language and English as a Foreign Language (ESL/EFL). Few would argue that the beliefs held by teachers and instructors influence their perceptions, which, in turn, affect their behaviors in their classrooms (see review by Pajares, 1992). Given this, teachers and instructors who possess more favourable or positive beliefs about assessment and evaluation are likely to embrace these activities as part of their instruction and regularly use a variety of assessment procedures, while teachers and instructors who possess less favourable or negative beliefs are likely to approach these activities in less positive ways with a more restricted set of activities.

As indicated above, different terms have been used in the study of teachers' beliefs. In the literature reviewed next, different terms, most frequently ‘attitude’, were used, and these terms have been retained when presenting the findings. However, as used, attitude is equivalent to belief in this study.

Beliefs about Assessment and Evaluation

Gullickson (1984) surveyed a stratified random sample of 391 3rd-, 7th-, and 10th-grade teachers in South Dakota. He reported that the teachers who possessed more positive attitudes toward assessment and evaluation tended to agree that classroom tests “increase student effort, affect student self-concept, create competition, improve student interaction, and in general improve the learning environment” (p.

247). In their longitudinal cross-sectional study, Green and Stager (1986) found that teachers' attitudes toward classroom testing became more positive with increasing grade in their random sample of 555 teachers in Wyoming. The teachers with positive attitudes toward classroom testing indicated that classroom tests and testing effectively motivated students, directed learning, allowed identification of student problems, and provided the information needed to communicate with and report to parents. Further, Green and Stager found evidence to suggest that negative prior experiences with tests and the test results that teachers received when they were students in school may have helped shape the attitudes the teachers held.

Quilter and Chester (1998) conducted two studies in which they investigated the attitudes teachers held toward assessment and evaluation. In their first study, 168 teachers currently enrolled in a master's degree program at Eastern Michigan University provided their perceptions about assessment and evaluation from the perspectives of a student (past experience) and as a teacher (current experience). The results of this first study showed that the majority of the teachers tended to be more positive about classroom assessment and evaluation than about standardized testing. Further, the teachers who were positive about standardized testing tended to see less value in alternative forms of assessment, while the teachers who were more positive about classroom assessment tended to be more accepting of alternative forms of assessment. As with Green and Stager (1986), Quilter and Chester found that past personal experiences with measurement and evaluation and current views of assessment and evaluation were highly interrelated, suggesting that the perceptions teachers have about assessment and evaluation are influenced somewhat by their own past experiences as students. Quilter and Chester (1998) concluded their first study by stating that "Teachers must see value in it [assessment] if they are to buy into it" (p. 11) and suggested that it is just as important that teachers develop positive perceptions of measurement and evaluation as it is to equip them with the techniques of measuring and evaluating student learning. In their second study, Quilter and Gallini (2000) investigated the relationship between teachers' knowledge of and attitudes toward assessment and evaluation. They administered a questionnaire consisting of questions related to assessment literacy and attitude toward (i) standardized testing, (ii) classroom assessment, and (iii) alternative forms of assessment to a group of 117 in-service elementary and secondary teachers. They found that the teachers' experiences with standardized testing and classroom assessment were positively related, respectively, to their current attitudes toward standardized testing and classroom assessment and negatively related to their current attitudes toward alternative assessment forms. Further, in agreement with the findings of their first study, the results showed that personal experiences with testing played a more important role in understanding teachers' current attitudes toward assessment than the educational assessment course they took during their teacher education program.

In contrast to the studies cited above, which were conducted in the United States, Blok, Otter, and Roeleveld (2002) examined the assessment and

evaluation practices used by 617 teachers and 214 school principals in The Netherlands and their attitudes toward student assessment and evaluation. They found that the teachers and principals preferred curriculum-embedded tests and grading of work sheets more than norm-referenced tests, valued reporting to parents and supporting learning and teaching more than accountability, and favorably viewed the quality of their own assessments. In an Australian study, Breen, Barratt-Pugh, Derewianka, House, Hudson, Lumley, and Rohl (1997) conducted a national study in which they investigated how teachers interpreted and used national and state assessment frameworks to assess young children's development of English as a second language. They reported that the teachers placed a high value on finding continuity and consistency between their own assessments and the new frameworks. However, teachers in states with imposed statewide assessment procedures expressed resistance, and almost all of the teachers were concerned about the workload involved in using the frameworks. Teachers who were positive about the frameworks tended to be those who were well supported by their school communities and specialist advisers.

As can be seen from the above review, teachers' beliefs play a role in teachers' assessment practices. Reported in the present paper are the results of an investigation of the beliefs about the value of assessment and evaluation held by ESL/EFL instructors in Canada, Hong Kong, and Beijing. The results were obtained as part of a three-year study (Cheng, Rogers, & Hu, 2004) designed to examine the assessment and evaluation practices used by ESL/EFL instructors at universities located in Canada, Hong Kong, and Beijing as well as their beliefs about assessment and evaluation. A review of literature in the second/foreign language education revealed limited studies on the beliefs about or perceptions of the value of assessment and evaluation held by university ESL/EFL instructors. Therefore, the purposes of this portion of the study were to answer the following questions: a) what are the ESL/EFL instructors' beliefs about the value of assessment and evaluation? b) what differences, if any, are there in the beliefs of the instructors in the three settings? and c) are the differences found, if any, related to the assessment and evaluation purposes held and practices used by the instructors in the three ESL/EFL contexts?

Method

For the purposes of the present study, assessment was defined as the process of collecting information about a student to aid in making an evaluation about the progress of a student. Evaluation was defined as the interpretation of results of an assessment, which describes the worth or merit of a student's performance in relation to a set of learner outcomes or standards of performance. These definitions were provided to the ESL/EFL instructors in a cover letter attached to the survey questionnaire.

Survey Questionnaire

The questions asked in the survey questionnaire were presented in five major sections. The first three sections corresponded to the major activities of classroom

assessment and evaluation. These sections were: Purposes of Assessment and Evaluation; Assessment Methods (used to assess ESL/EFL students in reading, writing, and speaking/listening); and Assessment Procedures (sources of and procedures used to construct assessment instruments; procedures and form of feedback provided to students during and at the end of the course; and estimated percentage of instructional time spent on assessment and evaluation). The fourth section contained questions developed to obtain information about the beliefs the instructors held about the value of assessment and evaluation and the fifth section contained bio-demographic questions (gender, age, educational qualifications, preparation for teaching, teaching experience, and current teaching load). The items included in the first three sections were based on the *Code of Fair Testing Practices for Education* (1988), *Standards for Teacher Competence in Educational Assessment of Students* (1990), *Report of the Task Force on Testing Standards (TFTS; 1995)*, and *Principles for Fair Student Assessment Practices of Education in Canada* (Rogers, 1993).

The items included in the fourth section of the questionnaire considered in the present paper were developed to gain an understanding of what teachers saw as the value of assessment and evaluation, the different methods they chose to assess their students, and their own preparedness and understanding of assessment and evaluation. These questionnaire items were developed based on the belief scale used by Gullickson (1984). Thirty-two statements were included in this section of the questionnaire (see Appendix). These items explored the beliefs of the ESL/EFL instructors in relation to their beliefs about enhancing instruction and student learning; beliefs about item formats, classroom assessment procedures, and time to prepare assessment; beliefs about standardized testing; and beliefs about instructors' understanding of and preparation for assessment and evaluation. The survey respondents were asked to indicate the degree to which they agreed with each statement using a five-point (1 = strongly disagree, ..., 5 = strongly agree) Likert scale.

The questionnaire was individually pilot-tested with 4 to 5 respondents in each of the three contexts to check that the items were clear and understandable. Based on the comments received, minor revisions with wording were made and a glossary of assessment and evaluation terms was added to ensure a common understanding of these two terms. Approximately 40-60 minutes was required to complete the questionnaire.

Samples

The ESL/EFL instructors at seven universities located in Alberta, British Columbia, and Ontario in Canada, seven universities in Hong Kong, and 11 universities in Beijing, China were invited to participate in the survey. The three samples represented, respectively, instructors in three different ESL/EFL instructional contexts – English-dominant, Bilingual (English and Cantonese), and Mandarin-dominant – representing a continuum of ESL and EFL instructional contexts (Dubin & Olshtain, 1986). These three contexts were chosen to determine if differences existed among the assessment practices used in the three

contexts given differences exist in instructional focus and context. The focus of instruction in Canada is, for the main part, on preparing students to enter an English speaking university, while in Hong Kong the focus is on students completing studies in English as a subject and in Beijing the focus is on preparing students to meet the English requirements for graduation. Previous research (e.g., see McMillan & Workman, 1998) suggested that there seems to be a relationship between the instructional context and teachers'/instructors' assessment beliefs and practices. Further, as pointed out above, the instructional goals differed among these three contexts.

In each of these contexts, ESL/EFL instructors at each university that formally offered ESL or EFL programs were sent questionnaires and self-addressed envelopes. Four researchers coordinated the study – one in the Western part and one in the Eastern part of Canada, one in Hong Kong, and one in Beijing. Wherever possible, meetings were held with the co-coordinator and faculty at each university to explain the purposes of the study and to answer any questions prior to the distribution of the questionnaires.

Altogether 461 questionnaires – 191 in Canada, 140 in Hong Kong, and 130 in Beijing – were distributed. Of this number, 95 (49.8%) were returned in Canada, 44 (31.4%) in Hong Kong, and 124 (95.3%) in Beijing. The difference in response rates is mainly attributable to the way in which each researcher coordinated the survey. For example, the survey questionnaire was administered on site during an EFL teacher training session in Beijing, administered on-site or left with the coordinator of the ESL program at the Canadian universities, and mailed to the instructors in the Hong Kong universities. Further, the lower response rate in Hong Kong may be attributed to fact that the instructors in Hong Kong had previously been surveyed several times during the same year.

Data Analysis

The responses to the survey questionnaire were entered into a computer file with 100% verification by a bonded data entry firm. Examination for missing item level data revealed that the number of missing responses for each item was two or less in each context. The mean item score by context was thus imputed for each missing response.

Two analyses were conducted at the item level (Note 1). First, a 3 x 5 (group-by-item response) contingency table analysis was performed for each item to examine differences in the distributions of responses from the respondents in the three samples across the five-points of the response scale. Second, a multivariate-univariate analysis was conducted to assess whether there were differences in the general elevation of the distributions of responses of the respondents in the three samples. The MANOVA revealed that there were significant differences in mean belief among the three settings (Wilk's $\Lambda = 0.258$; $F_{10,456} = 0.426$; $p < 0.05$).

Following Hummel and Sligo (1971) and Finn (1974), a one-way ANOVA, employing the Browne-Forsythe (Brown & Forsythe, 1974) test statistic given the lack of homogeneity of variance for some of the items and the unequal sample sizes, was then conducted for each item. Simultaneous pair-wise multiple-

comparisons tests employing Tamhane's (1979) procedure for samples of unequal size were completed for each item for which a significant difference was indicated. The analyses were completed using Version 14.0, Statistical Package for the Social Sciences (SPSS, 2006).

Given the lack of previous studies in which the beliefs of university level ESL/EFL instructors about assessment and evaluation were studied, the 0.05 level of significance was adopted. Significant pair-wise contrasts among the three settings were claimed only if the differences reflected the transitivity property (Note 2).

Results

Description of Samples

As shown in Table 1, (page 56) the percentage of male instructors in Hong Kong (36.4%) was greater than the percentages of male instructors in Canada (14.7%) and Beijing (14.5%). The instructors in Beijing were younger than the instructors in Canada and Hong Kong (83.8% vs. 34.7% and 34.1% below 41 years of age). While all but 3 instructors in Canada and 5 instructors in Beijing possessed a university degree, there were differences in the highest degree attained: all the instructors in Hong Kong possessed a masters or doctoral degree in comparison to 59.0% in Canada and 37.9% in Beijing. Lastly, there were differences in the instructors' preparation for the measurement and evaluation of their students. While 45.2% of the instructors in Canada reported they had completed a full course, 34.1% and 16.1% of the instructors in Hong Kong and Beijing, respectively, had completed such a course; 41.0% of the instructors in Canada reported that the topics of measurement and evaluation had been discussed in other courses they had taken while 59.1% of the instructors in Hong Kong and 26.6% of the instructors in Beijing attended such courses. Further, while slightly more than 2 out of 5 instructors in Beijing reported that they had no education in the areas of measurement and evaluation, less than 1 in 10 instructors in Canada and Hong Kong indicated they had no training.

The instructors in Canada had more years of ESL/EFL teaching experience than the instructors in Hong Kong who in turn had more years of ESL/EFL teaching experience than the instructors in Beijing. The means of the number of years were, respectively, 18.6, 17.1, and 10.3 (Table 2, page 56). While more than 90% of the instructors in Hong Kong and in Beijing had full-time teaching appointments, approximately 75% of the instructors in Canada had full-time appointments. More than 90% of the instructors in Canada and Hong Kong taught all of their classes at the same university; in Beijing, slightly more than 80% taught their classes in the same university. The majority of courses taught in Hong Kong and in Beijing were university degree courses (93.2% and 93.5%); in contrast the courses in Canada were more evenly divided between degree (52.6%) and diploma/certificate courses (45.3%). Further, it appears that instructors in Beijing teach courses more at the undergraduate level (81.4%) than instructors in both Hong Kong (15.9%) and Canada (44.2%). Although the mean numbers and ranges of classes taught were similar across the 3 groups, the average

class size in Beijing was greatest (46 students) followed by Hong Kong (19 students) and Canada (15 students).

Comparative Analyses and Results

The descriptive analyses of the bio-demographic information provided by the respondents revealed that it was not possible to analyze the belief items in terms of the bio-demographic variables within each context due to insufficient sample sizes in some of the cells (e. g., see part-time instructors, Table 2, page 56). Consequently, the comparative analyses were restricted to comparing the three settings. Further, since the results of the chi-square analyses and the one-way ANOVAs agreed with one exception and given space limitations, only the ANOVA results are provided in this paper.

The items were grouped into six clusters based upon assessment and evaluation activities: beliefs about enhancing instruction and student learning (11 items); beliefs about item formats, classroom assessment procedures, and time to prepare assessments (14 items); beliefs about standardized testing (4 items); and beliefs about instructor understanding of and preparation for assessment and evaluation (3 items). The results are reported in Tables 3-6. The mean of the three samples together with an indication of how the three groups differed are provided for each item for which significant differences was found; a weighted mean for the three groups is used for the remaining items for which no significances were found. For these latter items, the ESL/EFL instructors held similar beliefs.

Beliefs about Enhancing Instruction and Student Learning

Table 3 (page 57) contains mean beliefs about enhancing instruction (Panel A) and student learning (Panel B). As shown in Panel A, overall the instructors in all three contexts tended to positively view the role that assessment and evaluation plays in the improvement of instruction and student learning. They strongly believed that assessment and evaluation results were important for instruction (weighted mean (wm) = 4.23) and helped them focus their teaching (wm = 3.78). However, they expressed mixed beliefs about how closely they aligned their instruction to student assessment results (wm = 3.31). The instructors in Canada and Beijing, more so than the instructors in Hong Kong, felt that the results of the assessments they conducted with their students revealed to them how well they had taught their students (mean = 3.44; 2.98, and 3.62 for Canada, Hong Kong, and Beijing, respectively).

Similar to the pattern for enhancing instruction, the instructors in the three contexts believed that assessment and evaluation provide valuable learning experiences for their students (wm = 4.01), affected student self-concept in important ways (wm = 3.94), improving ESL/EFL learning (wm = 3.90), motivated students to work harder (wm = 3.80), and to try their best (wm = 3.80) (see Panel B, Table 3, page 57). The instructors felt somewhat neutral about how well students liked to be assessed and evaluated (wm = 3.25). However, while the instructors in Beijing did not believe that assessment and evaluation created competition among students (m = 2.13), the instructors in Canada and Hong Kong

held neutral beliefs about student competition ($m = 2.97$ and 2.76). Taken together, these results suggest that the instructors in all three contexts tend to positively view the role that assessment and evaluation play in the improvement of instruction and student learning.

Beliefs about Item Types, Classroom Assessment Procedures, and Time to Prepare Assessments

As shown in Table 4, Panel A (page 58), the instructors in all three contexts believed that assessment procedures similar to real life experiences were better than paper-and-pencil assessments ($w_m = 3.91$) but, when all things were considered, that paper-and-pencil assessments were best for determining what students have learned ($w_m = 3.46$). They held neutral beliefs that speaking and listening assessment were better for assessing ESL/EFL learning compared with paper-and-pencil assessments ($w_m = 2.79$).

The instructors in all three contexts believed that frequently assessing their students ($w_m = 4.31$) with a variety of assessment methods ($w_m = 3.89$) will encourage learning (Panel B). They were relatively happy about the quality of their assessments ($w_m = 3.44$) and they strongly believed that the evaluation of students should be performance-based rather than based on comparison of students ($w_m = 4.07$). However, they held neutral beliefs about whether they evaluated their students more than other teachers ($w_m = 2.90$).

While the instructors in the three contexts held common beliefs about the items discussed above, they did hold different opinions about formal assessments and paper and pencil assessments. The instructors in Canada and Hong Kong tended to believe more that formal assessments provide for a better assessment than informal assessments (Panel A) ($m = 3.67$ and 3.46) than the instructors in Beijing ($m = 2.72$). The instructors in Canada and Hong Kong believed in the use of published assessment methods (e.g., assessments found in textbooks, tests bought from publishers or obtained from a government agency (Panel B) ($m = 4.37$ and 4.46); in contrast the instructors in Beijing were uncertain ($m = 2.74$). In addition, while the instructors in Canada and Hong Kong held neutral beliefs about the use of paper-and-pencil assessments as their primary method of assessment (Panel A) ($m = 3.00$ and 2.85), the instructors in Beijing believed less so ($m = 2.33$).

Despite holding overall beliefs that assessment and evaluation enhance instruction and student learning as reported above, the instructors in Hong Kong and Beijing held neutral beliefs about whether assessments took up more time and effort than they were worth ($m = 3.12$ and 3.05), while the instructors in Canada believed less so ($m = 2.57$). While the instructors in all three contexts tended to hold neutral views about the time needed to develop their assessment methods ($w_m = 2.88$; Panel C, Table 4, page 58), the instructors in Canada and Hong Kong tended to feel that, due to other commitments, they did not have sufficient time to properly prepare their assessments ($m = 3.46$ and 3.56); the instructors in Beijing tended to feel otherwise ($m = 2.63$). Taken together, these results suggest that the instructors are not sure about what time is needed for assessment and evaluation.

However, the differences seem to be more between the instructors in Canada/Hong Kong and the instructors in Beijing except for one item.

Taken together, these results, which at times appear to be contradictory, reflect a pragmatic view. As pointed out earlier, the students who attend these courses are preparing to write a standardized examinations which have a heavy pencil and paper component, such as the TOEFL, IELTS, or CAEL, as part of the entrance requirements of an English speaking university in Canada and the College English Test (CET) as part of the degree requirements in Beijing; there is no similar requirement for students in Hong Kong. At the same time, these ESL/EFL instructors also recognized the value of and need to use performance assessments (e.g., speaking and, in the case of conversations, listening, that are not amenable to the use of paper-and-pencil assessments). These results suggest that the instructors are not sure about what time is needed for their assessments and evaluations in relation to their worth.

Beliefs about Standardized Tests

Despite the difference among the instructors in the three contexts, the instructors in all three contexts believed that the use of external standardized tests focused on passing rather than communicating (see Table 5, page 59) ($m = 4.26, 3.88, \text{ and } 4.28$). While majority of the students in the EFL courses in Beijing need to pass the College English Test (CET) to graduate (Cheng & Gao, 2002), the instructors in Beijing did not feel pressured to teach toward the test ($m = 2.55$). In contrast, the instructors in Canada and, surprisingly, Hong Kong felt pressured to teach toward standardized examinations. Like the students in Beijing, the majority of students in the ESL courses in Canada need to pass the TOEFL, IELTS, or CAEL to gain admission to a university program; there were no standardized tests for university students in Hong Kong (Cheng, et al., 2004). Based on these results, there does not seem to be a relationship between the existence of standardized tests and how instructors feel about the pressure to teach toward these tests. In addition, the instructors in all three contexts believed that having students take practice examinations was beneficial ($w_m = 3.76$), yet they were neutral about whether the standardized examinations assessed only what is easy to measure and not what is important ($w_m = 3.33$).

Beliefs about Instructors' Understanding of and Preparation for Assessment and Evaluation

While the instructors in Canada and Hong Kong believed that they had an adequate understanding of how to construct assessments ($m = 3.71 \text{ and } 3.73$), they also believed that they did not have an adequate understanding of how to use the assessment results to evaluate their students ($m = 3.91 \text{ and } 3.85$; see Table 6, page 59). In contrast, the instructors in Beijing were less sure ($m = 3.13; 3.14$). Lastly, the instructors in all three contexts held neutral beliefs about the degree to which the courses they took to prepare them to be an ESL/EFL instructor helped them to construct assessments ($w_m = 3.27$).

As indicated earlier, *assessment* was defined as the process of collecting information about a student to aid in making an evaluation about the progress of the student and *evaluation* was defined as the interpretation of results of an assessment, which describes the worth or merit of a student's performance in relation to a set of learner outcomes or standards of performance. Interpretation and evaluation of student performance are complex activities that, in contrast to developing and choosing assessment methods, are very difficult to describe and teach. More attention is paid to how to develop or chose selection items and performance tasks and how to compute grades than on how to formulate an interpretation and report this interpretation to students and others with a right-to-know. Further, as reported in Table 2 (page 56), approximately 45% of the instructors in Canada, 34% of the instructors in Hong Kong, and 16% in Beijing reported that they had taken a full course in assessment and evaluation. Rogers (1991) pointed out that when assessment and evaluation were taught as part of other courses (e.g., curriculum, learning), the treatment varied and was often shallow. Further, Impara, Plake, and Fager (1993) and Wise, Lukin, and Roos (1991) also found in their studies of teacher preparation that teacher' preparation for classroom assessment and measurement is inadequate. Hence, the mixed beliefs expressed about level of understanding are not surprising.

Relationship between Beliefs and Purposes and Assessments Used

In an earlier paper, Cheng, et al. (2004) reported and discussed the first set of survey results of this three-year study. In that paper they compared the assessment and evaluation practices employed by the ESL/EFL instructors in the three contexts. The beliefs presented and discussed above coincide somewhat with the assessment and evaluation practices the ESL/EFL instructors reported they used.

Instruction and Student Learning

While nearly equal percentages of instructors in the three contexts reported they used their assessments and evaluations for student-centered purposes (e.g., formative purposes like providing feedback, monitoring progress, diagnosing strengths and weaknesses, and motivating students to learn; and summative purposes like determining final grades), a greater percentage of instructors in Canada and Beijing than in Hong Kong used their assessments for instructional purposes (e.g., plan instruction, diagnose strengths and weaknesses in their own teaching and instruction) (Cheng, et al., 2004). These uses reflect the ESL/EFL instructors' beliefs that assessment and evaluation are important for instruction and help improve student learning. The instructors in all three contexts believed that assessment and evaluation are important for instruction and helped to focus what is to be taught. However, while the instructors in Canada and Beijing believed that the assessment results for their students indicated how well they had taught their students, the instructors in Hong Kong were not sure that the assessment results for their students indicated how well they had taught their students. This apparent mixed scenario in Hong Kong may be ascribed to ensuring

that their assessment methods align the curriculum they are teaching rather than taking into account the performance of their students.

Item Types, Classroom Assessment Procedures, and Time to Prepare Assessments

The types of assessment methods used to assess the students' performances in reading, writing, and in speaking and listening in the three ESL/EFL classroom settings clearly differed among the three contexts. For example, the most common reading assessment strategies used in Canada were student summaries of what is read and short answer items; in Beijing, objectively scored assessment methods such as multiple-choice items, true-false items, short answer questions, and standardized reading and writing tests were used most frequently; and in Hong Kong student summaries of what was read and short answer items. Student-conducted assessments, such as student journals and portfolios, were most frequently used in Canada followed by Hong Kong and then Beijing (Cheng, et al., 2004). First, these results reveal a clear relationship between the purposes of assessment and the assessment methods used by the instructors in the three contexts. Second, the assessments used by the ESL/EFL instructors reflect somewhat their beliefs. For example, the instructors in all three contexts believed that they need a variety of assessment methods and that more frequent shorter (objective) assessments are more effective than less frequent longer (subjective) assessments. However, while the Beijing instructors more than the Canadian and Hong Kong instructors believed that informal assessments are better than formal assessments for evaluating ESL/EFL students and less than the Canadian and Hong Kong instructors in the use of paper-and-pencil assessments, in practice they used objectively scored assessments more than the Canadian and Hong Kong instructors. The observation that the assessment methods used by the Beijing instructors were the more structured is likely attributable to their need to prepare their students for the CET and their larger class sizes, which were on average 2 to 2 ½ times greater than the average class sizes in Canada and Hong Kong. The difference between beliefs and actions reflects a desire on one hand and pragmatic need on the other.

Regarding the superiority of one type of assessment method over another, while the instructors in all three contexts indicated that they believed assessment methods similar to real life situations are better than paper-and-pencil procedures, at the same time they held neutral beliefs about the superiority of paper-and-pencil methods for determining what the student have learned. This apparent contradiction may again reflect a pragmatic need: performance assessments require a greater amount of time required to score than objectively scored assessments, such as alternative response and multiple-choice assessments, and to respond to students querying the marks assigned to their performance assessment given the subjective nature of the marking performance assessments. While the instructors in all three contexts used performance or constructed response assessments to assess writing and speaking, they likely wanted to use these assessment methods more in reading and listening but felt they could not because of the greater labor and time needed for marking.

Lastly, the instructors in all three contexts spend approximately 25% of their instructional time on assessment and evaluation (Cheng, et al., 2004). However, the instructors generally were not sure about what time is needed for their assessment and evaluation activities. When coupled with their beliefs about the superiority of the different forms of assessments and what assessment methods they use, the finding of uncertainty in all three contexts may be attributable to the strain between their desire to use more constructed response assessments and their feelings that the time available to develop, administer, and score these assessments is not sufficient.

Standardized Examinations

More instructors in Canada and Beijing than in Hong Kong reported they used the results of their assessments and evaluations to prepare their students for standardized tests such as the TOEFL in Canada, and, in Beijing, the College English Test (CET) (Cheng, et al., 2004). This finding agrees with the focus of instruction in the three contexts: the instructional focus in Canada is, for the most part, on preparing students to enter an English speaking university, while in Hong Kong the focus is on students completing studies in English as the medium of instruction and in Beijing the focus is on preparing students to meet the English requirements for graduation. Again the beliefs expressed by the instructors agree somewhat with what they did. For example, all agreed that their students should take practice examinations. But while the instructors in Canada and, unexpectedly, Hong Kong believed that they were pressured to teach toward standardized examinations (e.g. TOEFL, government examinations, university entrance examinations), the instructors in Beijing did not. However, the instructors in the three contexts strongly agreed that the use of external standardized examinations focuses on passing rather than communication. This again shows a mixed and pragmatic view of the instructors.

Instructors' Understanding of and Preparation for Assessment and Evaluation

The instructors in the three contexts were equally uncertain about their understanding of assessment and evaluation despite the differences in their assessment training (see Table 1, page 56) However, while the instructors believed that they possessed an adequate understanding of how to construct assessments and that they were pleased with the quality of their assessments, they believed that they did not possess an adequate understanding of how to interpret and use the assessment results to evaluate their students. Frequently, interpretations are based on rankings (i.e., norm-referenced) with much less attention paid to the quality of what the students actually did. These findings may be attributable to the observation reported earlier that less than half of the instructors in each context had completed a full course on assessment and evaluation, and that greater attention and time of many of these courses are spent on the need to relate assessment items and tasks to their instructional objectives or expected outcomes, how to construct assessment items, and on the concepts of

reliability and validity rather than on how to interpret and evaluate the scores and information yielded by these assessment methods (Rogers, 1991).

Conclusion

In the introduction to this paper and in agreement with Pajares (1992), we stated that teachers and instructors who possess more favourable or positive beliefs about assessment and evaluation are likely to embrace these activities as part of instruction and regularly use a variety of assessment procedures, while teachers and instructors who possess less favourable or negative beliefs are likely to approach these activities in less positive ways with a more restricted set of activities. This statement is partially supported by the findings of the present study. While the beliefs that assessment and evaluation were important for instruction and help improve student learning and the actual purposes of and uses of assessment and evaluation held by the instructors were positively related, the instructors' beliefs how they conducted their assessments and evaluations, the time required for assessments and evaluations, and their understanding of and preparation for assessment and evaluation were only somewhat related to their actual assessment practices. As reported above, the beliefs expressed by the instructors in all three contexts were somewhat mixed, uncertain, and, at times, contradictory. The contingency table analyses revealed that responses of the instructors in the three contexts were often distributed across all or all but one of the five-points of the response scale, which suggests that the respondents were more mixed in their beliefs, particularly about the use of paper-and-pencil and performance assessments, the time required for assessments and evaluations, and their understanding of and preparation for assessment and evaluation. The first and second of these findings likely reflect the positive and negative characteristics of the two classes of assessment methods (e.g., paper and pencil assessments can be group administered while performance assessments are individually administered, thus taking up more time, vs. performance assessments are more authentic in nature than paper and pencil assessments. The third finding reflects the differences in preparation of the instructors as reflected in the training they have received and, perhaps, differences in their confidence in applying what they have learned about assessment and evaluation.

Further research is needed to clarify why the differences and lack of differences among the three groups of ESL/EFL instructors occurred. In fact, we discovered more similarities among instructors' beliefs (23 items of the 32 items we surveyed) than differences (9 out of the 32 items). It may be that the conduct of assessment and evaluation in the classroom occurs within a policy and procedural framework that is largely determined by outside expectations and by personal experience. Regulations of the university or department to which the instructors belong, students' expectations, and the English proficiency required for admission to English speaking universities and for employment might influence more the assessment and evaluation practices used by the instructors than their own beliefs (Wilson, 1990). Likewise, the assumptions instructors make about

their students and how they learn likely will influence how the instructors assess and evaluate their students and how they report the results to the students. Negative experiences with assessment and evaluation and, perhaps, lower proficiency of the students they teach may also influence instructors' beliefs about assessment and evaluation and the procedures they use.

Following his review of teacher beliefs, Calderhead (1996) questioned whether beliefs precede actions or actions precede and, therefore, influence beliefs. For example, do instructors' beliefs about the value of different assessment procedures precede the use of these procedures or are the beliefs formed by the instructors' prior experience with different assessment procedures? Since it is unlikely that the use of assessment and evaluation of students will diminish, and that ESL/EFL instructors will continue to make decisions that are crucial to the future of their students, it is essential that instructors be knowledgeable about the strengths and weaknesses of the methods and procedures they use. But if student assessments and evaluations are to be sound and valid, ESL/EFL instructors also need to believe in the value of all aspects – developing and choosing a variety of assessments, interpreting and scoring student responses, and communicating and reporting student results of the assessments and evaluations they conduct.

While questionnaire surveys afford the opportunity to survey relatively large numbers of respondents and to identify group differences and trends, they are somewhat limited in the depth of information that can be obtained and that the data provided are self-report data that cannot be probed directly. Further, they are subject to withholding of information and are susceptible to social mores. To clarify the findings of this study, interviews of members of each sample who agreed to be interviewed have been conducted in which explanations were sought to clarify these two issues. At the same time, the interview participants were asked to clarify what influenced them when they assessed and evaluated their students, and to provide suggestions that they thought could be used to foster more positive opinions about assessment and evaluation given the centrality of assessment and evaluation in instruction and learning (Note 3).

Notes

1. Factor analysis was used in an attempt to reduce the number of variables to be used to compare the beliefs of the instructors in the three settings. However, a clean interpretable solution that yielded reliable factor scores was not realized. Therefore, the analyses were conducted at the item level.
2. For example, there was a significant difference for the statement “All things considered, paper-and-pencil assessments are the best methods for determining what each student has learned. (Item 20)”. While the mean belief for instructors in Beijing was significantly lower than the mean for Canada, the differences between Beijing and Hong Kong and between Canada and Hong Kong were not significant. In cases such as these, the significant difference was not claimed. Instead, a weighted mean of all three samples was reported.

3. The interviews were conducted during the second year of the research project. The first set of results derived from the interviews can be found in Cheng, Rogers, and Wang (2007).

References

- Anderson, J. O. (1989). Evaluation of student achievement: Teacher practices and educational measurement. *The Alberta Journal of Educational Research*, 35 (2), 123 – 133.
- Anderson, J. O. (1990). Assessing classroom achievement. *The Alberta Journal of Educational Research*, 36 (1), 1 –3.
- Blok, H., Otter, M. E., & Roeleveld, J. (2002). Coping with conflicting demands: Student assessment in Dutch primary schools. *Studies in Educational Evaluation*, 28, 177-188.
- Breen, M.P., Barratt-Pugh, C., Derewianka, B., House, H., Hudson, C., Lumley, T. and Rohl, M. (1997). *Profiling ESL Children: How teachers interpret and use national and state assessment frameworks*. Department of Employment, Education, Training and Youth Affairs. Canberra City: Australia.
- Brown, G. I., & Forsythe, A. B. (1974). The ANOVA and multiple comparisons for data with heterogeneous variances. *Biometrics*, 30, 719-724.
- Calderhead, J. (1996). Teachers: Beliefs and knowledge. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 709-725). New York: Macmillan.
- Cheng, L., and Gao, L. (2002). Passage dependence in standardized reading comprehension: Exploring the College English Test. *Asian Journal of English Language Testing*, 12, 161-178.
- Cheng, L., Rogers, W. T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: purposes, methods, and procedures. *Language Testing*, 21, 359-389.
- Cheng, L., Rogers, W. T., & Wang, X. (2007). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment and Evaluation in Higher Education*. 32 (6).
- Code of Fair Testing Practices for Education*. (1988). Washington, DC: Joint Committee on Testing Practices. Available online from <http://www.apa.org/science/fairtestcode.html>
- Dubin, F. & Olshtain, E. (1986). *Course design: Developing programs and materials for language learning*. Cambridge: Cambridge University Press.
- Educational Testing Services. (1995). *Capturing the power of classroom assessment*. Princeton, NJ: Author.
- Finn, J. D. (1974). *A general model for multivariate analysis*. New York: Holt, Rinehart and Winston.
- Green, K. E., & Stager, S. F. (1986). Measuring attitudes of teachers toward testing. *Measurement and Evaluation in Counseling and Development*, 19, 141-150.

- Gullickson, A. R. (1984) Teacher perspectives of their instructional use of tests. *Journal of Educational Research*, 77(4), 244-248.
- Hummel, T. J., & Sligo, J. R. (1971). Empirical comparison of univariate and multivariate analyses of variance procedures. *Psychological Bulletin*, 76 (1), 49 – 57.
- Impara, J. C., Plake, B S., & Fager J. J. (1993). Teachers' assessment background and attitudes toward testing. *Theory into practice*, 32, 113-7.
- 1McMillan, J. H. and Workman, D.J. (1998). *Classroom assessment and grading practices: A review of the literature*. Richmond, VA: Metropolitan Educational Research Consortium. (ERIC Document Reproduction Service No. ED453263).
- Pajares, M. F. (1992). Teacher' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62 (3), 307-332.
- Quilter, S. M., & Chester, C. (Oct., 1998). *Inservice teachers' perceptions of educational assessment*. Paper presented at the Annual Meeting of the Mid-Western Educational Research Association, Chicago, Illinois.
- Quilter, S. M., & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator*, 36, 115-131.
- Report of the Task Force on Testing Standards (TFTS)* (1995). International Language Testing Association (ILTA). Available online from <http://www.dundee.ac.uk/languagestudies/lttest/ilta/ilta.html>
- Rogers, W. T. (1991). Educational assessment in Canada: Evolution or extinction? *The Alberta Journal of Educational Research*, 37 (2), 179 –182.
- Rogers, W. T. (1993). *Principles for fair student assessment practices for education in Canada*. Edmonton, AB: University of Alberta, Centre for Research in Applied Measurement and Evaluation. Available online from: <http://www.education.ualberta.ca/educ/psych/crame/>
- SPSS. (2006). Statistical package for the social sciences, 14.0 for Windows. Chicago: Il: Author.
- Standards for Teacher Competence in Educational Assessments of Students*. (1990). Washington, DC: American Federation of Teachers, National Council on Measurement in Education, and National Educational Association. Available online from: <http://www.unl.edu/buros/article3.html>
- Tamhane, A. C. (1979). A comparison of procedures for multiple comparisons of means with unequal variances. *Journal of the American Statistical Association*, 74, 471-480.
- Wilson, R. J. (1990). Classroom processes in evaluating student achievement. *The Alberta Journal of Educational Research*, 36 (1), 4-17.
- Wilson, R. J. (May, 2000). *A model of assessment-in-practice*. Paper presented at the Annual Conference of the Canadian Society for the Study of Education, Edmonton, Alberta.
- Wise, S. L., Lukin, L., & Roos, L. L (1991). Teacher beliefs about training in testing and measurement. *Journal of teacher Education*, 42, 37-42.

TABLE 1
Bio-demographic Description of Sample

Variable	Context		
	Canada (n = 95)	Hong Kong (n = 44)	Beijing (n = 124)
Gender			
Female	81 (85.3)	28 (63.6)	105 (84.7)
Male	14 (14.7)	16 (36.4)	18 (14.5)
Age			
20-25	0 (0.0)	0 (0.0)	8 (6.4)
26-30	3 (3.2)	3 (6.8)	34 (27.4)
31-35	16 (16.8)	4 (9.1)	29 (23.4)
36-40	14 (14.7)	8 (18.2)	33 (26.6)
41-45	18 (18.9)	7 (15.9)	6 (4.8)
46-50	15 (15.8)	9 (20.4)	8 (6.4)
Greater than 50	24 (25.3)	13 (29.5)	6 (4.8)
Educational Qualifications^a			
Certificate in Teaching ESL/EFL	49 (51.6)	12 (27.3)	45 (36.2)
Diploma in Teaching ESL/EFL	12 (12.6)	12 (27.3)	36 (29.0)
B. A.	54 (56.8)	19 (43.2)	79 (63.7)
B. Sc.	8 (8.4)	2 (4.5)	2 (1.6)
B. Ed.	28 (29.5)	7 (15.9)	6 (4.8)
Masters	51 (53.7)	34 (77.3)	45 (36.3)
Doctorate	5 (5.3)	10 (22.7)	2 (1.6)
Measurement and Evaluation Training^a			
Full Course	43 (45.2)	15 (34.1)	20 (16.1)
Partial Course	39 (41.0)	26 (59.1)	33 (26.6)
Workshop	44 (46.3)	12 (27.3)	22 (17.6)
None	7 (7.4)	3 (6.8)	52 (41.9)

Note: n (%); the percentages may not add to 100% due to rounding or missing data.

^a The respondents were asked to indicate all that applied when indicating their educational qualifications and what education they had received in measurement and evaluation..

TABLE 2
Teaching Experience and Assignments

Variable	Context		
	Canada (n = 95)	HongKong (n = 44)	Beijing (n = 124)
ESL/EFL Teaching Experience			
Range in years	1 – 36	3-33	1-30
Mean in years	18.6	17.1	10.3
Teaching Appointment			
Part-time	24 (25.3)	3 (6.8)	3 (2.4)
Full-time	71 (74.7)	41 (93.2)	120 (96.7)
Type of Courses Now Teaching			
Diploma/Certificate courses	43 (45.3)	3 (6.8)	0 (0.0)
Undergraduate ESL/EFL courses	42 (44.2)	29 (65.9)	101 (81.4)
Graduate ESL/EFL courses	0 (0.0)	0 (0.0)	2 (1.6)

Undergraduate & Graduate ESL/EFL courses	8 (8.4)	12 (27.3)	13 (10.5)
Number of Levels Now Teaching			
Level 1	44 (46.3)	12 (27.3)	40 (32.2)
Level 2	43 (45.3)	19 (43.2)	49 (39.5)
Level 3	6 (6.3)	9 (20.4)	12 (9.7)
Level 4	1 (1.1)	1 (2.3)	18 (14.5)
Number of Classes Now Teaching			
Range	1-20	2-20	1-22
Mean	5.3	7.0	7.0
Average Class Size			
Range	10-30	6-50	4-120
Mean	14.9	18.8	46.2
All Classes at Same University?			
Yes	88 (92.6)	41 (93.2)	102 (82.2)
No	7 (7.4)	3 (6.8)	21 (16.9)

Note: *n* (%); the percentages may not add to 100% due to rounding or missing data.

TABLE 3
Beliefs about Enhancing Instruction and Student Learning

Item No.		Mean		
		Canada	HongKong	Beijing
	<i>Panel A Enhancing Instruction</i>			
31	Assessment results tell me how well I have taught my students. ^a	3.44	2.98	3.62
	Items with no differences			
1	Assessment and evaluation results are important for instruction		4.23	
16	Assessment helps me to focus my teaching.		3.78	
25	My instruction is tied closely to student assessment results.		3.31	
	<i>Panel B Enhancing Student Learning</i>			
27	Assessment and evaluation create competition among students. ^b	2.97	2.76	2.13
	Items with no differences.			
2	Assessments provide a valuable learning experience for students.		4.01	
29	Assessment results have an important effect on student self-concept.		3.94	
5	Assessment and evaluation improve ESL/EFL learning.		3.90	
22	Assessments and evaluations make my students work harder.		3.80	
15	Almost all of my students try to achieve their best when they are assessed in my class.		3.70	
8	Students dislike being assessed and evaluated.		3.25	

^a Hong Kong < Canada = Beijing . This indicates that the mean of the Hong Kong instructors is lower than the mean of the Canadian instructors, yet it is the same, statistically, as the mean of Beijing instructors.

^b Beijing < Hong Kong = Canada

TABLE 4
Beliefs about Item Formats and Classroom Assessment Procedures

Item No.		Mean		
		Canada	HongKong	Beijing
<i>Panel A Item Formats</i>				
4	Formal assessments provide for a better evaluation of ESL/EFL students than do informal assessments. ^b	3.67	3.46	2.72
11	Paper and pencil assessments provide the primary basis for the grades I assign to my students. ^b	3.00	2.85	2.33
Items with no differences				
30	Assessment methods that are similar to real life situations are better than paper-and-pencil procedures.		3.91	
20	All things considered, paper-and-pencil assessments are the best methods for determining what each student has learned.		3.46	
9	Speaking and listening assessments are better for assessing ESL/EFL learning than paper-and-pencil assessments.		2.79	
<i>Panel B Assessment Procedures</i>				
26	I use published assessment methods (e. g., in textbooks; tests bought from a test publisher or a government agency) as my primary basis for assigning grades to my students. ^b	4.37	4.46	2.74
Items with no differences.				
21	I need a variety of assessment methods to assess my students.		4.31	
10	More frequent shorter assessments are more effective in encouraging learning than less frequent longer assessments.		3.89	
13	I am happy about the quality of my assessment.		3.44	
6	I assess and evaluate my students more than other ESL/EFL teachers.		2.90	
24	The evaluation of students should be performance-based rather than based on comparison of students.		4.07	
<i>Panel C Adequacy of Time for Assessment and Evaluation</i>				
3	Other commitments do not allow me sufficient time to properly prepare my assessments. ^b	3.46	3.56	2.63
19	In an instructional sense, assessments take up more time and effort than they are worth. ^c	2.57	3.12	3.05
Item with no differences				
12	I have sufficient time to develop the methods I use to assess my students.		2.88	

^b Beijing < Hong Kong = Canada

^c Canada < Beijing = Hong Kong

TABLE 5
Beliefs about Standardized Tests

Item No.		Mean		
		Canada	HongKong-K	Beijing
18	The use of external standardized examinations (e.g., TOEFL, government examinations, university entrance examinations) focuses on passing rather than communicating. ^a	4.26	3.88	4.28
32	I feel pressured to teach toward standardized examinations (e.g., TOEFL, government examinations, university entrance examinations). ^b	3.92	3.76	2.55
Items with no differences				
28	Having students take practice examinations in class helps them perform better on certification examinations (e.g., TOEFL, government examinations, university entrance examinations).		3.76	
14	Standardized assessments only assess what is easy to measure, not what is important.		3.33	

^a Hong Kong < Beijing = Canada

^b Beijing < Hong Kong = Canada

TABLE 6
Beliefs about Instructor Understanding of and Preparation for Assessment and Evaluation

Item No.		Mean		
		Canada	Hong-Kong	Beijing
	<i>Panel A Understanding of Assessment and Evaluation</i>			
17	I have an adequate understanding of how to construct assessments ^b .	3.71	3.73	3.13
23	I do not have an adequate understanding of how to use assessment results to evaluate my students ^b .	3.91	3.85	3.14
	<i>Panel B Preparation for Assessment and Evaluation</i>			
7	The courses I took to prepare me to be an ESL/EFL instructor helped me to construct assessment methods.		3.27	

^b Beijing < Hong Kong = Canada

Appendix

Beliefs about Assessment and Evaluation

Please indicate the degree to which you agree with each of the following statements. Circle the number corresponding to your selection. Please use the following scale:

- 1 = Strongly Disagree (SD)
- 2 = Disagree (D)
- 3 = Neither agree nor disagree (?)
- 4 = Agree (A)
- 5 = Strongly Agree (SA)

		SD	D	?	A	SA
1	Assessment and evaluation results are important for instruction.	1	2	3	4	5
2	Assessments provide a valuable learning experience for students.	1	2	3	4	5
3	Other commitments do not allow me sufficient time to properly prepare my assessments.	1	2	3	4	5
4	Formal assessments provide for a better evaluation of ESL/EFL students than do informal assessments.	1	2	3	4	5
5	Assessment and evaluation improve ESL/EFL learning.	1	2	3	4	5
6	I assess and evaluate my students more than other ESL/EFL teachers.	1	2	3	4	5
7	The courses I took to prepare me to be an ESL/EFL instructor helped me to construct assessment methods.	1	2	3	4	5
8	Students dislike being assessed and evaluated.	1	2	3	4	5
9	Speaking and listening assessments are better for assessing ESL/EFL learning than paper-and-pencil assessments.	1	2	3	4	5
10	More frequent shorter assessments are more effective in encouraging learning than less frequent longer assessments.	1	2	3	4	5
11	Paper and pencil assessments provide the primary basis for the grades I assign to my students.	1	2	3	4	5
12	I have sufficient time to develop the methods I use to assess my students.	1	2	3	4	5
13	I am happy about the quality of my assessment.	1	2	3	4	5
14	Standardized assessments only assess what is easy to measure, not what is important.	1	2	3	4	5
15	Almost all of my students try to achieve their best when they are assessed in my class.	1	2	3	4	5
16	Assessment helps me to focus my teaching.	1	2	3	4	5
17	I have an adequate understanding of how to construct assessments.	1	2	3	4	5
18	The use of external standardized examinations (e.g., TOEFL, government examinations, university entrance examinations) focuses on passing rather than communicating.	1	2	3	4	5
19	In an instructional sense, assessments take up more time and effort than they are worth.	1	2	3	4	5
20	All things considered, paper-and-pencil assessments are the best methods for determining what each student has learned.	1	2	3	4	5

21	I need a variety of assessment methods to assess my students.	1	2	3	4	5
22	Assessments and evaluations make my students work harder.	1	2	3	4	5
23	I do not have an adequate understanding of how to use assessment results to evaluate my students.	1	2	3	4	5
24	The evaluation of students should be performance-based rather than based on comparison of students.	1	2	3	4	5
25	My instruction is tied closely to my student assessment results.	1	2	3	4	5
26	I use published assessment methods (e. g., in textbooks; tests bought from a test publisher or a government agency) as my primary basis for assigning grades to my students	1	2	3	4	5
27	Assessment and evaluation create competition among students.	1	2	3	4	5
28	Having students take practice examinations in class helps them perform better on external examinations (e.g., TOEFL, government examinations, university entrance examinations).	1	2	3	4	5
29	Assessment results have an important effect on student self-concept	1	2	3	4	5
30	Assessment methods that are similar to real life situations are better than paper-and-pencil procedures.	1	2	3	4	5
31	Assessment results tell me how well I have taught my students.	1	2	3	4	5
32	I feel pressured to teach toward standardized examinations (e.g., TOEFL, government examinations, university entrance examinations).	1	2	3	4	5

W. Todd Rogers (PhD) is a Professor, Centre for Research in Applied Measurement and Evaluation, University of Alberta in Canada. He is interested in research in the area of measurement and evaluation of students with a focus on large-scale testing. He is particularly interested in the fair assessment of students. He was the chair of the group that developed the *Principles of Fair Student Assessment Practices for Education in Canada*, and he worked on a project to develop the *Student Evaluation Standards* in the United States.

Liyong Cheng (PhD) is an Associate Professor and a Director of Assessment and Evaluation Group (AEG) at the Faculty of Education, Queen's University in Canada. Before she joined Queen's University, she was a Killam Postdoctoral Fellow at the Center for Research in Applied Measurement and Evaluation, University of Alberta. Her primary research interests are the impact of large-scale testing on instruction, and the relationship between assessment and instruction in classrooms.

Huiqin Hu (PhD) is a psychometrician in Data Recognition Corporation in the United States. She is interested in research on large-scale assessment item development, analysis, and score reporting.