Current trends in language testing education in Japan

Adam Murray
Tokai University

Yasuko Ito
Kanda University of International Studies

Kazumi Kimura
Tokyo University of Foreign Studies

Kahoko Matsumoto
Tokai University

Yuji Nakamura
Keio University

Atsuko Okada
Toho University

This paper reports on the perspectives of textbook writers, teacher trainees and teacher trainers on the treatment of testing concepts in teacher education programs in Japan. Ten domestically published textbooks were examined to see what testing-related concepts are dealt with in comparison to their foreign counterparts. Most major concepts such as validity and reliability are covered. However, many textbooks fail to provide adequate explanations and hands-on practice of basic statistics. In addition to the textbook analysis, two surveys were administered. One was administered to teacher trainees enrolled in an English teacher-training workshop in order to learn more about their perceptions on testing. The second survey asked teacher trainers which testing concepts are covered in their courses and how much classroom time is allocated to them. Based on the findings from the textbook analysis and the survey results, the authors make recommendations about how testing concepts should be covered in teacher-training programs and textbooks.

　言語テストの作成、実施、分析は言語技能の指導と同じように語学教師の重要かつ避けられない職務であるにもかかわらず、テスティングに関する理論を踏まえた実践技術に関しては十分な教育あるいは補完指導がなされているとは言い難い。日本ではテスティングの指導は一般的には英語科教育法の講義において、主に教員志望者に対して教科書や補助的資料を使用して授業を展開する。また、現職教員研修においては資料を用意してワークショップ形式で行うこともある。本稿では次の3点を研究テーマとした調査分析結果を報告し、最後に今後の研究への示唆を述べる。1) 国内で出版された英語科教育法の教

科書のテスティングに関する記述の質的、量的分析（英語で出版されたものと比較、2）テスティング・ワークショップ参加者（主に教員志望者）に対するテスティング関連項目のアンケート調査結果分析、3）英語科教育法または教員養成講座担当者へのテスティング関連項目のアンケート調査結果分析。結論として、教科書分析においては、教科書全体に占めるテスティング関連の記述の割合が低く、妥当性、信頼性などの代表的な概念の説明はあるものの、多くの場合記述統計の解釈方法など実践的訓練につながる内容が少ないことが分かった。また、ワークショップ参加者と教員へのアンケート結果を総合すると、実践形式のトレーニングを行うことが、理論と実践、さらに指導と評価の関連性を認識かつ経験させる有効な方法の一つであることがわかった。

IN ADDITION to being effective instructors of subject content, teachers are faced with a variety of responsibilities. One of these responsibilities is the accurate assessment of their teacher trainees. Cumming (2009) pointed out three common problems in classroom-based assessment, namely teachers' training in assessment, assessment of professional or curriculum standards in their teaching, and assessment of their teacher trainees' learning. Stiggins (1998) also summarizes the importance of assessment and the need for teacher training, "Classroom assessment requires a great deal of time and effort; teachers may spend as much as 40% of their time directly involved in assessment-related activities. Yet teachers are neither trained nor prepared to face this demanding task" (p. 363). Although not specifically referring to the Japanese teaching context, this statement could also apply to Japan. In order to get a better understanding of the current situation in Japan, we decided to look at language testing education from the perspectives of textbook writers, teacher trainees (pre-service and in-service teachers) and teacher trainers.

From the textbook writers' perspective

An analysis was conducted on textbooks published in Japan for pre-service and in-service language teachers. The majority of the 10 textbooks selected (Appendix 1) were published after 2000. These books were analyzed from quantitative and qualitative perspectives. The quantitative analysis consisted of counting the number of pages allocated to testing-related chapters as well as the number of textbooks mentioning specific testing-related concepts and terms. These concepts and terms were considered important based on their treatment in typical testing textbooks and on our cumulative experience in both testing and teacher education.

The analysis revealed that the length of the testing-related chapters in these textbooks ranged from 9 to 37 pages, with the average being 19 pages. In addition to the length of the testing-related chapters, we also looked at testing-related topics. In total, 26 testing-related concepts (Appendix 2) and 14 types of tests (Appendix 3) were observed (see Douglas, 2010 for description of testing terminology). Table 1 shows which testing-related concepts the majority of the textbooks covered. Only three (reliability, validity, and practicality) of the 26 concepts were dealt with in more than six textbooks. Nine textbooks dealt with reliability, eight with validity, and six with practicality. Since the total number of the textbooks is 10, this indicates that textbook writers tend to consider two concepts (reliability and validity) to be rather essential for all teachers.

Table 1. Number of textbooks published in Japan covering testing-related concepts ($n=10$)

Testing-related concept	Number of books
Reliability	9
Validity	8
Practicality	6

Next, we looked at the types of tests covered by the textbooks. Table 2 shows the types of tests described by the majority of the textbooks. The most important test types are diagnostic, formative

and norm-referenced tests. These test types were covered by seven of the textbooks. In addition, six of the textbooks dealt with criterion-referenced, objective, subjective and summative tests. When teachers make or conduct tests, they should have testing purposes which lead to the testing types. Table 2 suggests that teachers should have a better understanding of a variety of test types to make effective tests so that they can fulfill their testing purposes.

Table 2. Number of textbooks published in Japan covering types of tests ($n=10$)

Type of test	Number of books
Diagnostic tests	7
Formative tests	7
Norm-referenced tests	7
Criterion-referenced tests	6
Objective tests	6
Subjective tests	6
Summative tests	6

In the qualitative analysis, we examined the explanations and examples offered in the testing-related chapters of the textbooks. The main strength of these textbooks is that they offer an informative overview and provide pre-service and in-service teachers with a good foundation for language testing. Their weakness, on the other hand, is that no concrete examples of actual tests are given with stated purposes. This can make it rather difficult for novice teachers to prepare tests for their own classes simply by reading English teacher education textbooks. It would be ideal to offer test-making practice and to explain basic statistics used for test results analysis in teacher training programs so that the teachers can properly measure the abilities and progress of their teacher trainees.

To put our textbook analysis into a wider perspective, an additional analysis was done on eight typical or widely used English teacher education textbooks published abroad. Special attention was paid to how these foreign counterparts cover the assessment-related contents that we found lacking in the textbooks published in Japan. All eight textbooks we analyzed devote more than one chapter (average 24 pages) to testing and evaluation and the most important testing-related concepts are covered in all of them. Specialized terms for statistical procedures are not included.

In sum, the eight widely used textbooks published abroad (Appendix 4) tend to give more detailed, full-fledged descriptions or explanations of important testing concepts, often in a "step-by-step" process-oriented manner, compared to the Japanese books. Naturally, each book reflects the author's experience and interests in terms of emphases given to specific areas, but they all provide a general philosophical background of assessment with the reasons why tests are given and the meaning of testing itself. In addition, the necessity of establishing validity by preparing a good test specification is given along with ample examples of communicative or performance testing. As Davies (2008) suggests, the tendency to incorporate "principles" with the knowledge and skills of testing is clearly seen. Notably, some of them introduce various alternative, longitudinal assessment methods such as portfolio and classroom-based formative assessment, as well as relatively new notions such as washback effect and test-making ethics.

Although the above-mentioned textbooks published abroad would also be beneficial for Japanese teachers of English, most of them are probably too technical and voluminous to use in ordinary English teacher education courses in Japan. In addition, the level of English used might also be too difficult. Therefore, rather than directly inserting the aspects and contents of these textbooks into present Japanese teacher education courses, care-

ful modifications and adjustments should be made to meet the needs and limitations of our EFL context, which will, hopefully lead to the establishment of our own "assessment literacy" (Inbar-Lourie, 2008; Malone, 2008).

From the teacher trainees' perspective

Taking our findings from the textbook analysis into consideration, we held one-day workshops for teacher trainees who are interested in testing in 2009 and 2010. The purpose of our workshops was to link teaching with testing by providing model lessons and assessment examples along with statistical information. There were 10 participants in 2009 and seven in 2010. They were teacher trainees who were enrolled in English teacher training courses at a variety of universities, and most of them intended to be teachers. Upon completion of the workshops, the participants provided written feedback about both the workshops and the textbooks used in their teacher training courses.

The perceived benefits of the workshops were as follows:
- Learning statistics
- Learning the connection between practice and theory
- Learning why some test items are inappropriate
- Hands-on experiences in test making and in assessing their peers
- Receiving feedback for test items they made

They felt that their textbooks and current teacher training courses were lacking in the following:
- Testing and assessment contents
- More specific examples of test items
- Hands-on experiences in testing

From their feedback, it can be considered that teacher trainees need more hands-on experiences and practical exercises in

testing to prepare themselves to be qualified teachers. Therefore, workshops could potentially fill the gap between their needs and the contents of current teacher training courses and textbooks.

From the teacher trainers' perspective

In 2007, a survey was conducted of language testing courses around the world (Brown & Bailey, 2008). Unfortunately, only three of the 97 respondents were Japanese, so no conclusions could be made about the current situation in Japan. As a result, we decided to conduct a similar survey of teacher trainers in Japan. Instead of focusing on language testing courses, we wanted to gain insight into the treatment of testing and testing-related concepts in teacher education courses. In particular, we wanted to answer the following research questions about teacher education courses in Japan:

- To what degree are general topics on testing covered?
- To what degree are hands-on experiences on testing covered?
- Which textbooks are being used?
- How do teacher trainees (pre-service and in-service teachers) feel about testing before and after they take the course?

In an attempt to answer these research questions, we conducted a small-scale pilot survey during the summer of 2010. The survey consisted of 56 statements, four open-ended questions and one multiple-choice question. The statements used a five-point Likert scale ranging from "none" (0) to "extensive" (4) to indicate how much classroom time was allocated to the item. The statements and their mean scores are presented in Appendices 5-8.

Tables 3 to 6 organize the statements into four general categories: "Time spent on testing concepts", "Time spent on test types", "Time spent on test administration" and "Hands-on

time spent on test administration". Table 3 shows how much classroom time was spent covering general testing concepts. Only 10 of the 26 testing concepts (Appendix 5) are given more than a little classroom time. This can be expected because many of these concepts (i.e., Rasch model, item response theory, logit model) would only be covered in testing courses not in general teacher education courses. The most classroom time is spent on validity (Table 3, Item 1; M=2.50) and reliability (Table 3, Item 2; M=2.21). Both are required components of accurate assessment. Because of their importance, an emphasis is placed on these fundamental concepts in all teacher education courses. A moderate amount of time is given to washback effect (Table 3, Item 3; M=2.07). Given the importance of entrance exams in Japan, teachers need to be aware of both the positive and negative aspects of washback effect.

Table 3. Time spent on testing concepts (n=14)

Concept	Mean
1. Validity	2.50
2. Reliability	2.21
3. Washback Effect	2.07

Note: 0 = none, 4 = extensive

After examining the treatment of testing concepts, we looked at the coverage of the various types of tests. Table 4 shows the amount of time allocated to test types. In total, 14 forms of tests (Appendix 6) were examined. With the exception of aptitude tests (M=0.86), teachers spend some time on all forms of tests (M>1.28). The most time is spent on attainment tests (Table 4, Item 1; M=1.86). Attainment tests are a priority because teachers are expected to be able to create classroom tests such as quizzes and end-of-term tests. Objective (Table 4, Item 2; M=1.64) and

subjective (Table 4, Item 3; M=1.64) tests are essential because teachers need to understand when to use both of these test types effectively. Due to the wide use of proficiency tests (Table 4, Item 4; M=1.64) and criterion-referenced tests such as *The EIKEN Test in Practical English*, criterion-referenced (Table 4, Item 5; M=1.57) tests are an important topic. Finally, a moderate amount of time is dedicated to norm-referenced tests (Table 4, Item 6; M=1.50). An understanding of these tests is necessary because of the importance of tests such as *The Test of English for International Communication* (TOEIC) and *The Test of English as a Foreign Language* (TOEFL).

Table 4. Time spent on test types (n=14)

Type	Mean
1. Attainment	1.86
2. Objective	1.64
3. Subjective	1.64
4. Proficiency	1.64
5. Criterion-Referenced	1.57
6. Norm-Referenced	1.50

Note: 0 = none, 4 = extensive

We then looked at eight aspects (Appendix 7) of test administration. The most classroom time is spent on scoring (Table 5, Item 1, M=1.57), interpretation (Table 5, Item 2, M=1.21) and analysis (Table 5, Item 3, M=1.07). Teachers are required to know how to assess their teacher trainees and give them letter and numerical grades. To do this, they need to understand how to make measurement tools such as grading scales and rubrics. After grades have been given, teachers need to understand the implications for the teacher trainees. In order to improve how their teacher trainees are evaluated, critiquing (Table 5, Item 4,

M=1) and revision (Table 5, Item 5, M=1) are essential. Teachers need to be able to identify both effective and ineffective test items and make appropriate revisions.

Table 5. Time spent on test administration (n=14)

Concept	Mean
1. Test Scoring	1.57
2. Test Score Interpretation	1.21
3. Test Analysis	1.07
4. Test Critiquing	1.00
5. Test Revision	1.00

Note: 0 = none, 4 = extensive

When looking at the amount of hands-on time allocated to the eight aspects of test administration (Appendix 8), the most time is allocated to item writing (Table 6, Item 1; M=1.43). Since teachers will be required to write (and administer) quizzes and tests in their classes, it is essential for them to have an understanding of test items (e.g., multiple choice, gap-fill). After administering a test, they need to be able to interpret the scores (Table 6, Item 2, M=1.07) in order to evaluate the performance of their students. To a lesser extent, scoring (Table 6, Item 3, M=1) and analysis (Table 6, Item 4, M=1) also require hands-on practice.

The responses were varied to the research question, "Which textbooks are being used?" Three of the respondents do not use textbooks and rely on self-made handouts. Of the remaining 11 trainers, nine different textbooks are used. Mochizuki (2002) and Negishi (2007) are the only two textbooks that are used by more than one person. As a result, it is impossible to make a conclusion about which textbooks are most commonly used.

Table 6. Hands-on time spent on test administration (n=14 respondents)

Concept	Mean
Item Writing	1.43
Test Score Interpretation	1.07
Test Scoring	1
Test Analysis	1

Note: 0 = none, 4 = extensive

To answer the first part of the fourth research question, we asked the teacher trainers, "How do teacher trainees feel about testing *before* they take the course?" The teacher trainers commented that their teacher trainees do not understand the role of testing before their courses. Many of the teacher trainees only have "vague" perceptions based on their own experiences as teacher trainees. One trainer commented, "They seem to have a very narrow understanding of testing before the course". Another said that some teacher trainees had very negative conceptions of testing, "Testing for screening, punishment, grading are the only notions they have about testing". The teacher trainees also had not thought about tests from the teacher's perspective, "they had never thought about making tests; for them, tests were things to take" and "they had never thought about analyzing tests". Finally, teacher trainees had the impression that "language testing is difficult" and they were "afraid of the word *statistics*".

In answering the second part of the fourth research question, we asked, "How do teacher trainees feel about testing *after* they take the course?" The majority of the trainers commented that the perceptions of their teacher trainees change drastically during the courses. The teacher trainees understand the importance of tests and that they "are not just for teachers to give grades".

As they learn about testing concepts, they become "interested in the concepts of validity, test analysis, and test development". They also become more critical of tests in general and by using newly learned concepts such as validity and relevance, "they can recognize the good points along with the weak points of the tests they have taken in the past". They also understand that it is difficult to construct a well-designed test. As a result, they feel that "more time should be spent on making tests". Finally, they also realize the importance of tests. One trainer commented, "Teacher trainees understand the significant influences that tests can have on their students' lives".

Discussion and conclusion

A good teacher-training program (in-service or pre-service), like a good test, must have three qualities: validity, reliability and practicality (Douglas, 2010). The validity of a teacher-training program must also be checked to confirm if the contents meet the goals of the program. The purpose of the testing (evaluation) section of a teacher training program should be to not only give testing knowledge, but also to provide hands-on opportunities to make tests and analyze test results.

A practical suggestion is to conduct workshops in addition to regular classroom lectures. A typical workshop could consist of several components or stages as shown in Figure 1. In the first stage, a lecture is delivered on testing theory and relevant testing terminology such as validity, reliability, practicality, and test specifications. In the second stage, teaching, a lesson plan is made that deals with one of the four skills. In the third stage, test making, the participants gain essential hands-on experience constructing tests for the lesson made in the previous stage. At this point, the strengths and weaknesses of various test types and item types can be discussed. In addition, topics such as rating scales, rating criteria, and grading rubrics should also be covered. The fourth stage, test data analysis, is also hands-on.

Here statistical concepts such as means, standard deviations, histograms, standard scores and correlations can be covered in a meaningful context. Also, more practical aspects of item analysis such as item difficulty and item discrimination could be dealt with. After completing the training, the pre-service and in-service teachers would then apply these concepts and use the associated techniques with their classes.

If held on a regular basis, these workshops could become cyclical in nature by going through teaching, learning, testing and analyzing cycles (cf. McNamara, 2000). They would also conform to the concept behind the Common European Framework of Reference (CEFR) for languages: teaching, learning and assessment. The CEFR was proposed by the Council of Europe (2001) to provide a common basis for the elaboration of language syllabi, curriculum guidelines, examinations, textbooks, etc. across Europe. Its content and purpose bring language learning, teaching and assessment into a much closer relation to each other than before (cf. Little, 2006). This idea has some impact on the cyclical nature of the workshops.
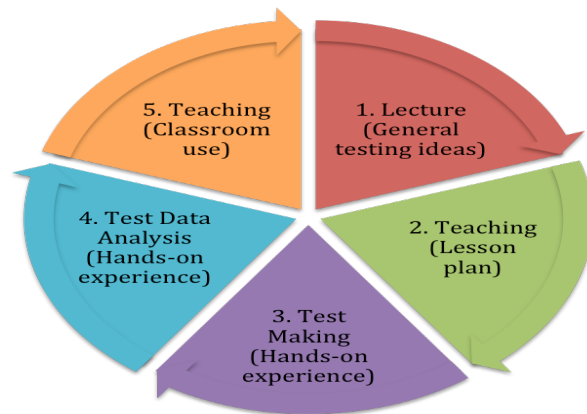

Figure 1. Workshop cycle with an emphasis on testing

This workshop approach is a response to the teacher trainees who participated in the summer workshops of 2009 and 2010 and who commented on the usefulness of the hands-on test-making component of the workshops. The survey of teacher educators also confirms the importance of hands-on experience in a good teacher-training program.

In the current textbooks published in Japan, few pages are spent on statistics and testing explanations. More pages need to be dedicated to testing concepts. A good teacher-training textbook, at a minimum, must include the following testing content: mean, standard score, standard deviation, histogram, correlation, t-test, and item analysis (item difficulty, item discrimination). Rating scales should be covered in the lectures and hands-on sections.

Ideally, testing should be taught as a separate course like it is taught in many foreign teacher education programs. However, this is not feasible in most teacher education programs in Japan. As the next best option, testing content should be integrated into present Japanese teacher education courses in a balanced way.

Bio data

Adam Murray is a junior Associate Professor at Tokai University, Shimizu campus. His current research interests are listening comprehension, writing fluency and technology-enhanced language learning. <murray@scc.u-tokai.ac.jp>

Yasuko Ito is an Associate Professor at Kanda University of International Studies. She has a PhD from the University of Hawaii at Manoa, and her research interests include second language acquisition, second language phonology, and language testing. <yasukoi@kanda.kuis.ac.jp>

Kazumi Kimura is a Lecturer at Tokyo University of Foreign Studies. She has a MA from the University of California, Los Angeles. Her research interests are second language acquisition, second language writing, and language testing. <kazumi.kimura@tufs.ac.jp>

Yuji Nakamura is a Professor of English at the Department of English in the Faculty of Letters at Keio University. His current interests include language assessment (specifically L2 reading assessment), and second and foreign language teaching and learning. His main focus has been on developing in-house placement tests and doing longitudinal analysis of students' English proficiency. <nkyj@flet.keio.ac.jp>

Kahoko Matsumoto is a Professor at Tokai University. She teaches English and applied linguistics subjects. Her main research areas are teaching and evaluation of writing, testing, and educational engineering. She has been involved in teacher training and faculty development. <mkahoko@tsc.u-tokai.ac.jp>

Atsuko Okada is a Professor at Toho University in Tokyo. She teaches English to nursing students. She is particularly interested in testing, and curriculum and material development for nursing and medical students. <okatsuko@med.toho-u.ac.jp>

References

Brown, J. D., & Bailey, K. M. (2008). Language testing courses: What are they in 2007? *Language Testing*, *25*(3*)*, 349-383.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching assessment.* Cambridge: Cambridge University Press.

Cumming, A. (2009). What needs to be developed to facilitate classroom-based assessment? *TESOL Quarterly, 43*(3)*, 515-519.

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

Douglas, D. (2010). *Understanding language testing*. London: Fodder Education.

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

Little, D. (2006). The Common European Framework of Reference for Languages: Content, purpose, origin, reception and impact. *Language Teaching, 39*(3), 167-190.

Malone, M. (2008). Training in language assessment. In E. Shohamy & N.H. Hornberger (Eds.), *Encyclopedia of language and education, 2nd Edition, Volume 7: Language testing and assessment,* (pp. 225-239). New York: Springer Science+Business Media LLC.

McNamara, T. F. (2000). *Language testing.* Oxford: Oxford University Press.

Mochizuki, A. (Ed.). (2002). *Shin Gakusyu shido yoryo ni motozuku eigoka kyoiku ho* [Methods of English teaching based on new guidelines for the course of study]. Tokyo: Taishukan.

Negishi, M., & Tokyo Society for the Study of English Education at Junior High School. (2007). *Communicative testing eno chosen* [Challenges to communicative testing]. Tokyo: Sanseido.

Stiggins, R.J. (1998). Revitalizing classroom assessment: The highest instructional priority. *Phi Delta Kappan, 69*(5), 363-368.

Appendix 1

Textbooks published in Japan analyzed

Asaba, R., Sato, T., Chiba, K., Toyoda, K., Nakamura, N., & Yamazaki, A. (2009). *Wakariyasui eigo kyoiku ho: Shochuko deno jissenteki shido* [Comprehensible methods of English teaching: Practical instruction at elementary, junior high, and senior high schools]. Tokyo: Sanshusha.

Higuchi, A., & Shimatani, H. (Eds.). (2007). *Nijuisseiki no eigo ka kyoiku* [English education in the 21st century]. Tokyo: Kairyudo.

Iino, S., & Shimizu, T. (1989). *Eigo no kyoiku: Hensen to jissen* [English education: Changes and practices] (Revised ed.). Tokyo: Taishukan.

JACET English Education SIG. (Ed.). (2005). *Shin eigo ka kyoiku no kiso to jissen: Jugyoryoku no saranaru kojo wo mezashite* [New basics and practices of English education: Aiming at more advancements in teaching]. Tokyo: Sanshusha.

Kotera, S., & Yoshida, H. (Eds.). (2005). *Eigo kyoiku no kiso chishiki* [Basic knowledge of English education]. Tokyo: Taishukan.

Matsumura, M. (2009). *Eigo kyoiku wo shiru gojuhachi no kagi* [58 key points to learn English education]. Tokyo: Taishukan.

Takanashi, T., & Takahashi, M. (2007). *Shin eigo kyoiku gaku gairon* [New introduction to the study on English education]. Tokyo: Kinseido.

Tsuchiya, S., & Hirono, T. (2000). *Shin eigo ka kyoiku ho nyumon* [New introduction to methods of English teaching]. Tokyo: Kenkyusha.

Yamauchi, S. (Ed.). (2003). *Gengo kyoiku gaku nyumon* [Introduction to the study on language teaching]. Tokyo: Taishukan.

Yoneyama, A. (2002). *Eigo kyoiku: Jissen kara riron e* [English education: From practice to theory]. Tokyo: Shohakusha.

Appendix 2

Testing-related terms covered by textbooks analyzed

Testing-related term	Japanese translation	Number of books
Absolute Evaluation	絶対評価	4
Classical Testing Theory	古典的テスト理論	1
Continuous Assessment	継続的評価	1
Correlation Coefficient	相関係数	3
Item Characteristic Curve	項目特性曲線	1
Item Difficulty	項目困難度	3
Item Difficulty Parameter	項目困難度パラメータ	1
Item Discrimination	項目弁別力	3
Item Information Function	項目情報関数	1
Item Response Theory	項目応答理論	2

Testing-related term	Japanese translation	Number of books
Logit Model	ロジット得点	1
Mean	平均値	3
Median	中央値	1
Mode	最頻値	1
Practicality	実用性	6
Range	範囲	2
Rasch Model	1パラメーター・ラッシュ・モデル	1
Reliability	信頼性	9
Reliability Coefficient	信頼性係数	1
Significant Difference	有意差	1
Standard Deviation	標準偏差	4
Statistical Tests by Mean Scores	平均値差の統計処理	5
Test-taker Ability Parameter	受験者能力パラメータ	1
Validity	妥当性	8
Variance	分散	2
Washback effect	波及効果	1

Appendix 3

Test types covered by textbooks analyzed

Test type	Japanese translation	Number of books
Aptitude Tests	適性テスト	3
Attainment Tests	到達テスト	5
Criterion-Referenced Tests	目標基準準拠テスト	6
Diagnostic Tests	診断テスト	7
Direct Tests	直接テスト	4
Discrete Item Tests	個別項目テスト	3
Formative Tests	形成的テスト	7

Test type	Japanese translation	Number of books
Indirect Tests	間接テスト	4
Integrative Tests	総合テスト	3
Norm-Referenced Tests	集団基準準拠テスト	7
Objective Tests	客観テスト	6
Proficiency Tests	熟達度テスト	5
Subjective Tests	主観テスト	6
Summative Tests	総括的テスト	6

Appendix 4

Textbooks published abroad analyzed

Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy*. (Third edition). New York: Pearson Education.

Celce-Murcia, M. (2001). *Teaching English as a second or foreign language*. (Third edition). Boston: Heinle & Heinle Publishers.

Cohen, L., Manion, L., & Morrison, K. (2008). *A guide to teaching practice*. (Fifth Edition). New York: Routledge-Falmer.

Harmer, J. (2007). *The practice of English Language Teaching*. (Fourth Edition). Edinburgh: Pearson Education.

Johnson, K. (2008). *An Introduction to foreign language learning and teaching*. (Second Edition). New York: Pearson.

Nunan, D. (Ed.) (2003). *Practical English language teaching*. (International Edition). New York: McGraw-Hill.

Nation, I.S.P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking*. New York: Routledge.

Savignon, S. J. (1997). *Communicative competence: Theory and classroom practice: Texts and contexts in second language learning*. (Second Edition). New York: McGraw-Hill.

Appendix 5

Classroom time spent on testing concepts (mean scores)

- Absolute Evaluation (1.86)
- Classical Testing Theory (0.71)
- Continuous Assessment (1.36)
- Correlation Coefficient (0.93)
- Item Characteristic Curve (0.21)
- Item Difficulty (0.71)
- Item Difficulty Parameter (0.28)
- Item Discrimination (0.50)
- Item Information Function (0.21)
- Item Response Theory (0.36)
- Logit Model (0.28)
- Mean (1.43)
- Median (0.93)
- Mode (0.93)
- Practicality (1.93)
- Range (1.07)
- Rasch Model (0.36)
- Reliability (2.21)
- Reliability Coefficient (0.86)
- Significant Difference (0.57)
- Standard Deviation (1.43)
- Statistical Tests by Mean Scores (0.64)
- Test-Taker Ability Parameter (0.21)
- Validity (2.50)
- Variance (1.07)
- Washback Effect (2.07)

Appendix 6

Classroom time spent on test types (mean scores)

- Aptitude Tests (0.86)
- Attainment Tests (1.86)
- Criterion-Referenced Tests (1.57)
- Diagnostic Tests (1.36)
- Direct Tests (1.43)
- Discrete Item Tests (1.43)
- Formative Tests (1.28)
- Indirect Tests (1.43)
- Integrative Tests (1.28)
- Norm-Referenced Tests (1.50)
- Objective Tests (1.64)
- Proficiency Tests (1.64)
- Subjective Tests (1.64)
- Summative Tests (1.28)

Appendix 7

Classroom time spent on test administration (mean scores)

- Item Writing (0.93)
- Test Administration (0.78)
- Test Analysis (1.07)
- Test Critiquing (1.00)

- Test Revision (1.00)
- Test Score Interpretation (1.21)
- Test Scoring (1.57)
- Test Taking (0.64)

Appendix 8

Hands-on time spent on test administration (mean scores)

- Item Writing (1.43)
- Test Administration (0.36)
- Test Analysis (1.00)
- Test Critiquing (0.93)
- Test Revision (0.57)
- Test Score Interpretation (1.07)
- Test Scoring (1.00)
- Test Taking (0.57)