Articles

Validation of Scores From the Self-Evaluation Checklist for EFL Teachers in Japan

Akihiro Mikami
Kindai University

The purpose of this study was to validate scores produced by a reflection tool called the Self-Evaluation Checklist for EFL Teachers (SECEFLT) in the Japanese context. A survey was conducted with 760 EFL teachers in 984 junior high and lower secondary schools throughout Japan. The collected data were divided into 2 datasets for exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). The EFA results demonstrate that the original model with 32 items should be modified to a new 4-factor model with 18 items, including Content Knowledge and Skills, Pedagogical Knowledge and Skills, Professional Development Knowledge and Skills, and English Language Use by Teachers and Students. The CFA results show acceptable model fit between the new model and the second dataset. Therefore, both the EFA and CFA results provide positive evidence that the revised SECEFLT is a useful reflection tool for Japanese teachers of English in junior high schools and lower secondary schools in Japan.

　本研究の目的は、英語教師のための自己評価チェックリストにおける構成概念妥当性と信頼性を検証することである。日本全国の中学校・中等教育学校から無作為に抽出された984校に所属する760名の英語教師を対象に調査を実施した。収集されたデータを2つのセットに分け、それぞれのセットを使用して探索的因子分析と検証的因子分析を実施した。探索的因子分析では、自己評価チェックリストの回答データは、初版の32項目より修正版の18項目から成る4因子構造　（教科の知識・技能、教科を教えるための知識・技能、教師の成長に関する知識・技能、教師と生徒による英語使用）　に修正する必要があることが確認された。検証的因子分析では、修

正版のモデルはデータに適合していることが確認された。したがって、自己評価チェックリスト （修正版）は、日本の中学校・中等教育学校の日本人英語教師の専門能力を測定するための論理的な基盤を持っていることが示唆された。

Keywords: factor analysis; professional development; reflection tool; self-evaluation checklist; validation

As reflective practitioners, teachers are encouraged to develop their professional competencies throughout their careers. Many studies show that teachers should reflect on their teaching practices for professional development (e.g., Richards & Farrell, 2005; Richards & Lockhart, 1994; Wallace, 1998). However, few valid and reliable reflection tools are available to language teachers. Mikami (2015) developed a reflection tool called the Self-Evaluation Checklist for EFL Teachers (SECEFLT), but no psychometric analysis of the reliability and validity of scores it produces has been conducted. The purpose of this study was to validate scores produced by the SECEFLT for EFL teachers in the Japanese context.

The Significance of Reflection in Teacher Growth

Despite common agreement that reflection can facilitate teacher growth, it is difficult to determine why or what types of reflection are beneficial. As Farrell (2012) and Rodgers (2002) noted, the vagueness of the definition of reflection makes it difficult to understand it and discuss its effects.

The origin of reflection can be traced to the works of the 20th century American philosopher John Dewey. Dewey (1933) defines reflective thinking as "that operation in which present facts suggest other facts (or truths) in such a way as to induce belief in what is suggested on the ground of real relation in the things themselves" (p. 12). That is, through reflective thinking, people draw meaning from facts. Dewey believed that reflection could lead to learning (Dewey, 1933; Rodgers, 2002).

American philosopher Donald A. Schön (1983), who elaborated on Dewey's findings and discussed the epistemology of practice, pointed out the limits of technical rationality, the view that professional practice is based on the application of scientific theories and techniques. He also identified two types of reflection: *reflection-in-action* and *reflection-on-action*. Reflection-in-action means thinking about one's action while in the middle of it; reflection-on-action means looking back on and analysing one's past action. Schön emphasized the importance of reflection-in-action when professionals take action. He claimed that competent professionals engage in

reflection-in-action to address problems in uncertain situations. This does not mean he downplayed the significance of reflection-on-action. Rather, he showed that professionals use both reflection-in-action and reflection-on-action to improve their performance.

In teacher education, teacher improvement through reflection is called reflective teaching, reflective practice, or action research (Burns, 1999; Farrell, 2015; Richards & Lockhart, 1994). The definitions and usages of these terms are not fully agreed on, but most agree that these terms refer to the improvement of teaching practice through reflection, leading to professional development. Wallace (1998) suggested a model for teacher education that places reflection at the core of the process. According to the model, repeated reflection in everyday situations can promote teachers' professional development.

The Current State and Challenges of EFL Teacher Reflection Tools

Many in-service teacher education programs have introduced reflective practice to teachers, enabling them to reflect on their everyday practices (e.g. Burns, 1999; Farrell, 2015; Mikami, 2011). Through these programs, teachers can reflect on their teaching practices, discover problems in their teaching, and solve them, but it is not clear how they can improve their knowledge and skills for classroom teaching as one of the most important aspects of professional practice through reflection.

It is therefore necessary to develop a tool that enables teachers to focus on their professional competencies because they are difficult to operationalize, in contrast to easily observable teaching practices. The only widely known reflection tool for language teacher competencies is the European Portfolio for Student Teachers of Languages (EPOSTL; Newby et al., 2007). At the heart of EPOSTL are 193 "can-do" descriptors, grouped into seven categories. In Japan, EPOSTL's can-do descriptors were adapted to the Japanese linguistic, educational, and cultural context under the name Japanese Portfolio for Student Teachers of Language (JPOSTL; JACET SIG on English Language Education, 2014).

However, teachers face challenges in using EPOSTL. For example, students often feel overwhelmed by EPOSTL's numerous can-do descriptors (JACET SIG on English Language Education, 2014). Furthermore, teacher educators or supervisors must be well organised, as EPOSTL requires long-term usage to be effective (JACET SIG on English Language Education, 2014). JPOSTL has received similar criticism. A Japanese teacher of English explained that using JPOSTL in long-term teacher education programs is effective but dif-

ficult for teachers with heavy workloads (Koide, 2016). Therefore, a more practical and user-friendly tool should be developed for EFL teachers to use for daily self-reflection. The process of developing such a tool should begin with a discussion of the professional competencies essential for EFL teachers in Japan.

In sum, a user-friendly instrument for comprehensive self-evaluation of EFL teachers' professional competencies can streamline teachers' self-reflection on their teaching practices, allowing them to focus more on the professional competencies that are essential for their practices.

Development of the Self-Evaluation Checklist for EFL Teachers

In the absence of a practical reflection tool for EFL teachers' professional competencies, Mikami (2015) developed the Self-Evaluation Checklist for EFL Teachers (SECEFLT) by drawing upon two major studies on teacher growth, Roberts (1998) and Hatta (2000). As Table 1 shows, the original SECEFLT was comprised of four components: Content Knowledge and Skills, Pedagogical Knowledge and Skills, Classroom Teaching Skills, and Professional Development Knowledge and Skills. All 32 items in the SECEFLT are shown in the Appendix.

The SECEFLT has three predominant features. First, its development process began with the discussion of professional competencies necessary for EFL teachers; in contrast, the development of JPOSTL began with a review of the can-do descriptors of EPOSTL. Second, 32 items related to core professional competencies were selected, taking into consideration the number of items that teachers can reflect on at one time so they can use the checklist without adding to their heavy workloads. Third, the SECEFLT provides EFL teachers with common criteria to self-evaluate their professional growth throughout their careers.

Table 1. The Tentative Structure of the Original SECEFLT

Professional competency (components)	Definition	Number of items
Content Knowledge and Skills	Knowledge and skills of the target language (English). It includes skills necessary to communicate in English; knowledge of vocabulary, grammar, and language usage; and knowledge of the culture of the English-speaking world.	8
Pedagogical Knowledge and Skills	General pedagogical knowledge and professional English language education knowledge and skills. It includes the abilities to select appropriate teaching materials and to change teaching content and methods flexibly depending on student comprehension.	8
Classroom Teaching Skills	Understanding of the context of English language education and teaching techniques used in classroom settings. It includes the abilities to understand learners, curricula, and schools and to promote learners' English use in classrooms.	8
Professional Development Knowledge and Skills	Knowledge and skills necessary for teacher development. It includes the abilities to reflect on one's own classroom teaching objectively and to improve one's own classroom teaching based on feedback from students and other teachers.	8

Validation of the Self-Evaluation Checklist for EFL Teachers

There are many studies on validity in the psychometric and educational measurement fields, but there is not yet a consensus on its definition (Hubley & Zumbo, 1996; Messick, 1989; Sireci, 1998). Since 1950, multiple types of validity have been posed (Messick, 1989). According to Hubley and Zumbo (1996), validity is traditionally conceptualized in the following ways: content validity, criterion-related validity, and construct validity. Criterion-related validity is subcategorised into concurrent validity and predictive validity. However, a single concept of validity was represented in the Standards

for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1985), indicating that the traditional categories cannot be distinguished rigorously. Messick (1989, 1995, 1996) suggested that validity is a unified concept with six distinguishable aspects of construct validity: the content, substantive, structural, generalizability, external, and consequential aspects.

Despite these different stances towards validity, it is common to put importance on examining how accurately the results obtained by using the instruments reflect what is intended to be measured. Thus, as the first step for the validation of scores from the SECEFLT, construct validity is verified through statistical data analysis. Construct validity is equivalent to one type of validity in the traditional conceptualization and also to the central part of the unified conceptualization, considered to be a structural aspect. If it can be shown that the data collected with the SECEFLT are consistent with the theoretical construct resulting from the investigations of construct validity in this study, this will demonstrate that the SECEFLT meets the basic conditions of an effective educational measurement instrument.

Research Questions

The purpose of this study was to verify the validity and reliability of scores from the SECEFLT through a nationwide survey of EFL teachers in Japan. The following three research questions were posed:

RQ1.	Does an EFA of the scores produced by the SECEFLT produce a factor model corresponding with the four dimensions originally hypothesized for the instrument?

RQ2.	To what extent do the responses to the SECEFLT fit the original or revised model for the sample of EFL teachers in Japan?

RQ3.	What are the general tendencies of the professional competencies of EFL teachers in Japan?

Methods

Participants

Junior high schools and lower secondary schools[1] in Japan were randomly selected using the list by Zenkoku Gakkou Data Kenkyuujo (2013). One out of every 11 junior high schools was chosen from 10,547; one out of every two lower secondary schools was chosen from 50. Therefore, the total number of schools in the study was 959 junior high schools and 25 lower secondary

schools, or 984 schools in total. The participants were all Japanese teachers of English working in these 984 schools throughout Japan.

Instrument

Participants were requested to self-evaluate their own professional competencies on a 6-point Likert scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*), making use of the question items in the SECEFLT.

Procedure

The SECEFLT was sent to the randomly selected schools from September to December 2014. The responses of 760 teachers from 369 schools (a response rate of 37.5%) were collected. Forty-four teacher responses were excluded from the data analyses because of missing items or more than one response to the same item. Thus, 716 surveys were used in the analysis. All of the collected data were randomly divided into two independent datasets, based on the participants' school category and years of teaching experience. Two sets of responses from 358 teachers were prepared: Dataset A and Dataset B. Table 2 shows the distribution of school categories in both datasets, with the largest being public junior high schools. Table 3 presents the distribution of the participants' years of teaching experience, showing a wide variety of teaching experience. The distributions in the two datasets were very similar.

Table 2. Categories of Participants' Schools in Both Datasets

	Dataset A		Dataset B	
Categories of schools	*n*	%	*n*	%
National junior high	16	4.5	17	4.7
Public junior high	307	85.8	307	85.8
Private junior high	15	4.2	14	3.9
National lower secondary	5	1.4	5	1.4
Public lower secondary	9	2.5	9	2.5
Private lower secondary	6	1.7	6	1.7
Total	358	100	358	100

Table 3. Participants' Years of Teaching Experience in Both Datasets

Years of experience	Dataset A		Dataset B	
	n	%	*n*	%
1 - 5	77	21.5	78	21.8
5 - 10	65	18.2	65	18.2
10 - 15	51	14.2	51	14.2
15 - 20	40	11.2	39	10.9
20 - 25	52	14.5	52	14.5
25 - 30	42	11.7	42	11.7
30 - 35	25	7.0	25	7.0
35 - 40	6	1.7	6	1.7
Total	358	100	358	100

According to Field (2013), factor analysis can be used to identify the structure of a set of variables and develop an instrument to measure an underlying variable. There are two types of factor analysis: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA; Kline, 2016). EFA is conducted to determine how the observed variables are linked to their underlying latent variables when the links between them are unknown. CFA is conducted to evaluate to what extent the hypothetical structure between the observed and latent variables is appropriate when the researcher has some knowledge of the links between them. By conducting CFA after the underlying structure is established using EFA, it is possible to evaluate how well the EFA models fit the data.

In this study, factor analysis was used to answer the first and second research questions. Specifically, EFA was used to examine whether the four factors in the original model of the SECEFLT were extracted. Following this examination, the theoretical framework was reconsidered carefully, the model was revised, and CFA was used to evaluate this finalized model.

However, as Kline (2016) pointed out, the same data should not be used in both factor analyses because the same chance variation may influence the results. Therefore, in this study, the collected data were divided into two datasets. The data were collected randomly, and each split dataset can be considered to be independent. Dataset A was used to establish the underlying structure of responses to the SECEFLT through EFA, and Dataset B was

used for the CFA to evaluate how well the finalized model fits the data. The reliability of the SECEFLT was also examined using both datasets. Finally, for the third research question, data from both datasets were used separately to uncover the general tendency in the responses of EFL teachers regarding their professional competencies. PASW Statistics 18.0 and Amos 18.0 were used for the data analyses.

Results

Following Field's (2013) recommendations, the normality of the collected responses to each item in both datasets was checked. As a result, five items (8, 12, 19, 24, and 31) were excluded from the analysis because their absolute skewness z scores or their absolute kurtosis z scores were above 3.29 and the shapes of the distributions were also not visually normal. The absolute skewness z scores or the absolute kurtosis z scores of four items (25, 26, 28, and 30) were slightly above 3.29 only in Dataset B, but their distributions were not visually different from a normal distribution, so these four items were not excluded. The values of the final available items for the analysis varied as follows: skewness ranged from -0.44 to 0.07, kurtosis ranged from -0.19 to 0.95, skewness z scores ranged from -3.42 to 0.54, and kurtosis z scores ranged from -0.75 to 3.68, raising no questions about the normal distribution of the data. Moreover, it was found that two respondents wrote on the margins of their questionnaire sheets "I have very few opportunities to observe other teachers' classroom teaching" and "I am the only English language teacher at my school." Thus, Item 32 was also removed from the analysis because its content was not appropriate. Neither ceiling effects nor floor effects were detected. Finally, a total of 26 items were available for analysis.

Exploratory Factor Analysis (EFA)

To determine empirical support for the hypothesised four-factor structure model based on the 26 items in Dataset A, EFA was conducted using the maximum likelihood method with promax rotation. Following Zwick and Velicer (1986) and Hori (2005), both the parallel analysis and MAP methods were conducted on Dataset A using the SPSS script developed by Hori (2001) to determine the number of components. The results of both methods suggested retaining four components. Therefore, it was determined that the final number of components was four. The minimum item-loading threshold was set at .50. This stringent criterion was used to select items that accounted for more variance, which suggested their importance. In the

EFA process, Items 15, 16, 17, 29, and 30 were not found to load on any factor at greater than .50, so they were removed from the subsequent analyses. Items 3 and 20 cross-loaded on two factors and were thus discarded. Item 18 loaded on a different factor than the hypothesised one at greater than .50; therefore, this item was deleted. In all, eight items were removed during the EFA, yielding four factors with 18 items. Table 4 shows the descriptive statistics of the 18 individual items retained after the EFA.

Table 4. Descriptive Statistics of 18 Individual Items (Dataset A)

Question item	*M*	*SD*	Skewness	Kurtosis
1	3.68	0.98	-0.11	0.31
2	3.85	1.04	-0.30	-0.00
4	3.86	0.99	-0.25	0.50
5	3.72	0.96	-0.24	0.45
6	4.09	0.95	-0.34	0.56
7	3.81	0.98	-0.05	0.01
9	4.12	0.90	-0.20	0.28
10	4.20	0.85	-0.25	0.42
11	4.14	0.89	-0.36	0.55
13	4.26	0.96	-0.25	-0.01
14	4.15	0.89	-0.27	0.17
21	4.02	1.04	-0.05	-0.19
22	4.01	0.96	-0.23	0.29
23	3.82	0.90	-0.06	0.19
25	4.16	0.88	-0.07	0.15
26	4.05	0.88	-0.00	0.26
27	4.13	0.91	-0.21	0.29
28	4.34	0.89	-0.38	0.73

As shown in Table 5, four factors were extracted. According to the content of the loading items for each factor, Factor 1 was named Content Knowledge and Skills, Factor 2 was named Pedagogical Knowledge and Skills, Factor

3 was named Professional Development Knowledge and Skills, and Factor 4 was named English Language Use by Teachers and Students. All factors except for Factor 4 matched the hypothesized categories. Table 6 shows the correlation between these factors.

Table 5. Pattern Matrix of EFA Results

Question item	Factor 1 (α = .93)	Factor 2 (α = .93)	Factor 3 (α = .91)	Factor 4 (α = .88)	Communality
4	**.95**	-.07	.06	-.02	.87
5	**.94**	-.00	-.02	-.03	.84
1	**.92**	-.09	.01	.07	.84
2	**.88**	-.03	-.05	.03	.73
6	**.66**	.23	.03	-.07	.62
7	**.54**	.21	-.06	.00	.44
11	-.01	**.95**	.04	-.07	.84
10	.00	**.90**	-.04	.02	.80
9	.01	**.76**	.06	.07	.73
13	.11	**.69**	.03	.08	.71
14	.01	**.67**	.15	-.03	.61
27	.00	-.03	**.95**	-.04	.82
25	.09	-.04	**.83**	.05	.78
26	-.02	.16	**.76**	.02	.78
28	-.08	.07	**.74**	-.03	.54
22	-.03	-.01	-.01	**.96**	.86
21	.08	.04	.03	**.72**	.66
23	-.02	.19	.19	**.54**	.69

Note. The numbers in bold indicate factor loadings of .50 or above. Factor 1 = Content Knowledge and Skills; Factor 2 = Pedagogical Knowledge and Skills; Factor 3 = Professional Development Knowledge and Skills; Factor 4 = English Language Use by Teachers and Students.

Table 6. Factor Correlation Matrix

Factor	1	2	3	4
Content Knowledge and Skills	1.00			
Pedagogical Knowledge and Skills	.60	1.00		
Professional Development Knowledge and Skills	.49	.73	1.00	
English Language Use by Teachers and Students	.56	.73	.71	1.00

According to Kline's (2016) criteria for describing internal consistency, coefficients around .90 are excellent, values around .80 are very good, and values about .70 are adequate. Reliability coefficients for each factor (Cronbach's alpha) varied from .88 to .93 and are sufficient to confirm internal consistency.

Based on the EFA results, the original structure model for the SECEFLT was revised and updated. Figure 1 summarizes the changes from the original version to the revised one.

The original version (32 items in total)	**The revised version (18 items in total)**
1. Content Knowledge and Skills (8 items)	1. Content Knowledge and Skills (6 items)
2. Pedagogical Knowledge and Skills (8 items)	2. Pedagogical Knowledge and Skills (5 items)
3. Classroom Teaching Skills (8 items) → (not extracted)	**3. English Language Use by Teachers and Students (3 items)**
4. Professional Development Knowledge and Skills (8 items)	4. Professional Development Knowledge and Skills (4 items)

Components of professional competencies narrowed down from EFA results		
Professional competencies (components)	Definitions	Number of items
English Language Use by Teachers and Students	In classroom settings, teachers can use English, encourage students to use English, and evaluate students' English use appropriately.	3

Figure 1. Changes to the SECEFLT: Different components and item numbers between the original and revised versions.

Confirmatory Factor Analysis (CFA)

To test and evaluate the revised four-factor model with the 18 items supported by the EFA results, CFA was conducted using Dataset B. Table 7 shows the descriptive statistics of these 18 individual items using Dataset B.

Table 7. Descriptive Statistics of 18 Individual Items (Dataset B)

Question item	*M*	*SD*	Skewness	Kurtosis
1	3.75	0.99	-0.01	0.10
2	3.89	0.99	-0.20	0.04
4	4.00	1.01	-0.27	0.16
5	3.82	1.00	-0.31	0.27
6	4.19	0.96	-0.35	0.37
7	3.82	0.98	-0.21	0.01
9	4.13	0.83	-0.00	0.09
10	4.22	0.80	-0.25	0.28
11	4.16	0.81	-0.16	0.25
13	4.28	0.90	-0.17	0.23
14	4.21	0.88	0.01	0.01
21	4.01	1.00	-0.40	0.28
22	4.01	0.91	-0.21	0.42
23	3.87	0.83	0.07	0.35
25	4.19	0.85	-0.24	0.89
26	4.10	0.84	-0.39	0.95
27	4.15	0.89	-0.13	0.28
28	4.36	0.90	-0.44	0.56

Multivariate distribution was checked using Mardia's normalized estimate of multivariate kurtosis. The *z* statistic of 37.29 is suggestive of nonnormality in the sample. The maximum likelihood method was used, so the degree to which the estimates are statistically significant may be overestimated (see Byrne, 2010, for further details).
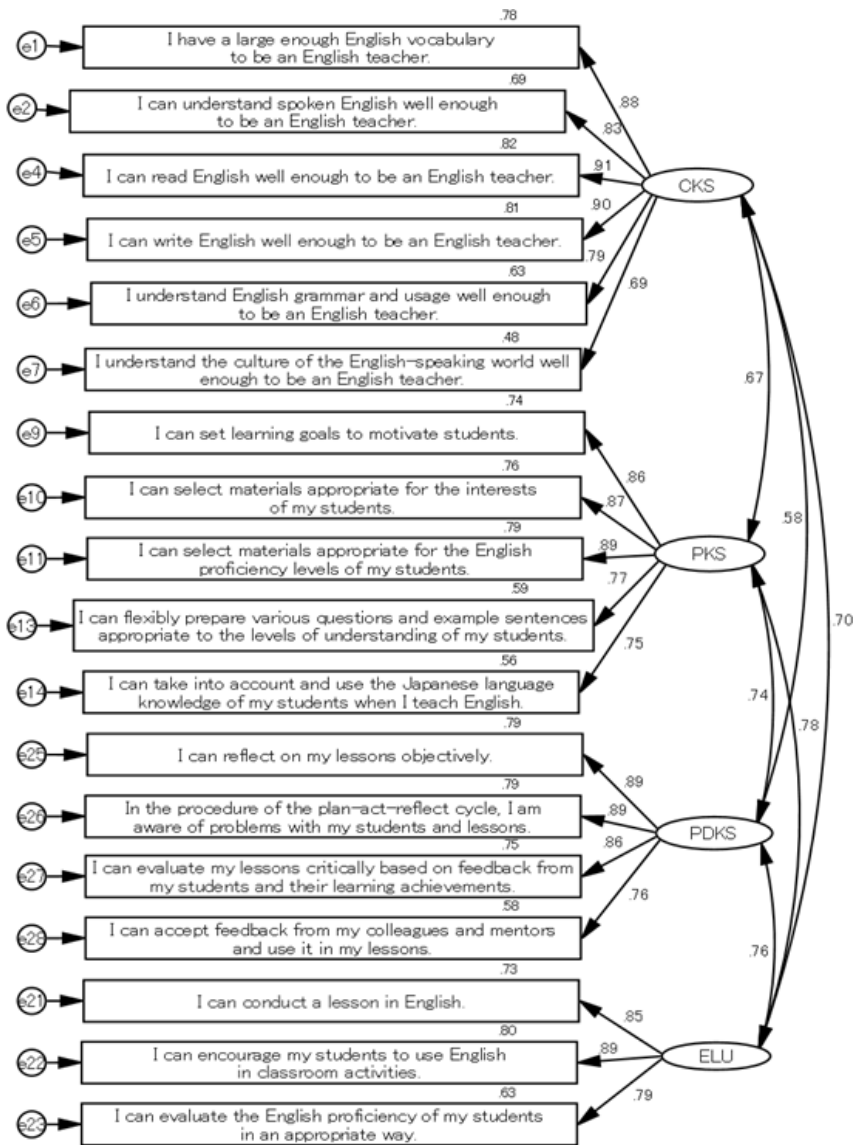
Figure 2. CFA model with standardized estimates for the revised SECEFLT. Ellipses represent latent variables and squares represent observed variables. CKS = Content Knowledge and Skills; PKS = Pedagogical Knowledge and Skills; PDKS = Professional Development Knowledge and Skills; ELU = English Language Use by Teachers and Students; e = measurement error.

The CFA model with standardized estimates for the revised SECEFLT is presented in Figure 2. All the loadings between the indicators and the latent variables as well as the covariances among the factors were statistically significant ($p < .001$).

To evaluate the fit between the CFA model and the observed data, many goodness-of-fit indices are available. Brown (2015) recommended considering at least one fit index from each out of the three following categories: absolute fit (e.g., standardized root mean square residual [SRMR]), parsimony correction (e.g., root mean square error of approximation [RMSEA]), and comparative fit (e.g., comparative fit index [CFI] and Tucker-Lewis index [TLI]). Table 8 shows the fit indices' values calculated from Dataset B. Another goodness-of-fit index, chi-square, is rarely used as a sole model fit index because a large sample size inflates it (Brown, 2015). Schumacker and Lomax (2010) pointed out that it is notoriously difficult to meet the criteria for chi-square, especially for sample sizes over 200. The chi-square value was 447.49, and chi-square/*df* was 3.47 ($p < .001$). Although this result is unacceptable, it is likely influenced by the relatively large sample size.

The cutoff criteria for goodness-of-fit indices are hotly debated, and it is difficult to specify clear criteria for model fit because they depend on model conditions such as sample size, model complexity, and estimation method (Brown, 2015). This study used the cutoff criteria suggested by Hu and Bentler (1999), which proposed the recommended value for a relatively good fit as .08 or below for SRMR, .06 or below for RMSEA, and .95 or above for CFI and TLI. As Table 8 shows, the SRMR value showed good model fit. The RMSEA value exceeded the cutoff, but it was less than 0.10, so it was not rejected (Brown, 2015). Both the CFI and TLI values were slightly below the cutoff, but Bentler (1992) originally considered a well-fitting model to have a CFI of greater than .90, and so, these values were considered an acceptable degree of fit.

Goodness-of-fit indices are interpreted on a continuum according to cutoff criteria and not as absolutes. Therefore, these statistics showed acceptable model fit between the CFA model and the data of Dataset B. That is, it was demonstrated that the construct validity of the revised four-factor model in Dataset B was appropriate. As Table 9 shows, reliability coefficients for each factor (Cronbach's alpha) varied from .88 to .93, sufficient to confirm internal consistency.

Table 8. Goodness-of-Fit Indices for the CFA Model

Categories	Absolute fit	Parsimony correction		Comparative fit	
Index	SRMR	RMSEA	90% CI	CFI	TLI
Values	.045	.08	.075, .092	.94	.93

Note. CI = confidence interval; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index.

Table 9. Reliability Coefficients for Each Factor (Dataset B)

Factors	α
Content Knowledge and Skills	.93
Pedagogical Knowledge and Skills	.91
Professional Development Knowledge and Skills	.91
English Language Use by Teachers and Students	.88

Subscale Values in Participants' Self-Evaluation of Professional Competencies

Regarding the four factors extracted based on the results of both the EFA and CFA, the mean scores of all the items included in each were computed as subscale values. Table 10 shows the descriptive statistics of the subscale values in both Datasets A and B.

Table 10. Descriptive Statistics of the Subscale Values in Datasets A and B

Subscales	Dataset A		Dataset B	
	M	*SD*	*M*	*SD*
Content Knowledge and Skills	3.83	0.85	3.91	0.85
Pedagogical Knowledge and Skills	4.17	0.79	4.20	0.73
Professional Development Knowledge and Skills	4.17	0.79	4.20	0.77
English Language Use by Teachers and Students	3.95	0.87	3.97	0.82

Discussion

Regarding Research Question 1, the results of the EFA and CFA indicated that EFL teachers' responses to the SECEFLT can be classified into four components of professional competencies, as hypothesised. However, there are some differences from the theoretical framework for the original SECEFLT. The most remarkable difference is that only three items out of eight for the hypothesised factor Classroom Teaching Skills were retained through the EFA, and a new factor named English Language Use by Teachers and Students emerged. When the SECEFLT was created, it was believed that two components included in the framework, namely Pedagogical Knowledge and Skills and Classroom Teaching Skills, were clearly distinguishable for respondents, because the latter can be interpreted as the techniques used by teachers in actual classroom settings. However, through the EFA process, it was found that the items hypothesised for Classroom Teaching Skills had relatively high loadings on a different factor (Pedagogical Knowledge and Skills), or cross-loadings on two factors (Pedagogical Knowledge and Skills and Classroom Teaching Skills). Thus, some of the items designed for these two factors were not clearly distinguishable for the respondents. That may be why Classroom Teaching Skills was extracted.

On the other hand, it is noteworthy that English Language Use by Teachers and Students was extracted. A reason why this factor was extracted may be the influence of national policy on English language education in Japan. In 2003, Japan's Ministry of Education, Culture, Sports, Science and Technology (MEXT) formulated *An Action Plan to Cultivate "Japanese With English Abilities."* The Commission on the Development of Foreign Language Proficiency (MEXT, 2011) also presented *Five Proposals and Specific Measures for Developing Proficiencies in English for International Communication*. Behind these concrete plans and proposals by MEXT lay the rapid advance of globalisation in fields such as politics, economics, and industrialisation. As such, reinforcing English language skills and the teaching abilities of EFL teachers is seen as critical. In light of this situation, participants in this study may have had a strong awareness of the emphasis of increasing the English language use of teachers and students in the classroom, which may have caused the new factor of English language use by teachers and students to be extracted.

As for the second research question, the results of the CFA showed that the SRMR value indicated good model fit, the CFI and TLI values indicated fit very close to satisfactory, and the RMSEA had mediocre fit. Although it is difficult to judge the CFA results, it is clear that EFL teachers' responses in Dataset B adequately fit the four-factor structure model for the revised SE-

CEFLT. Moreover, Cronbach's alpha coefficients in this study show that participants' responses to the items for each factor were internally consistent. Overall, therefore, the results provide positive evidence for the four-factor revised model of the SECEFLT.

As for the third research question, as shown in Table 10, the results of both data sets showed that the means of two subscales (Pedagogical Knowledge and Skills and Professional Development Knowledge and Skills) were higher than those of the other subscales, but the mean of Content Knowledge and Skills was the lowest. This implies that EFL teachers in Japan are more confident in their professional competencies related to teaching and professional development than in their content knowledge and skills, at least among those who responded to the survey. The fact that the participants were all Japanese teachers of English may have contributed to this result.

This study has several limitations. First, only the construct validity of the scores from the revised SECEFLT, specifically, the structural aspect of the unified concept, was examined. Future research needs to be conducted to accumulate various types of evidence for the validity of the revised SECEFLT. For example, content validity should be examined by asking professionals to evaluate to what extent the content of each item in the revised SECEFLT is related to what it is supposed to measure. Second, the participants in this study were only Japanese teachers of English in junior high schools and lower secondary schools in Japan. Further research should be done to confirm whether the SECEFLT can be used for different populations, such as high school teachers or English-speaking assistant language teachers in Japan. If it is confirmed that the structural properties of scores from the revised SECEFLT are stable among different populations, it can be used to examine the different trends in each population's evaluation of their own professional competencies.

Notes

1. Lower secondary schools in Japan are schools that have educational continuity through 6-year secondary levels, whereas 3-year junior high school and 3-year senior high school levels are integrated.

Acknowledgments

Akihiro Mikami is a professor at Kindai University, Japan. His research interests include TESOL teacher education, action research for classroom teaching improvement, and extensive reading.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Bentler, P. M. (1992). On the fit of models to covariances and methodology to the *Bulletin. Psychological Bulletin, 112,* 400-404. https://doi.org/10.1037/0033-2909.112.3.400

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: The Guilford Press.

Burns, A. (1999). *Collaborative action research for English language teachers.* Cambridge, England: Cambridge University Press.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York, NY: Routledge.

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process.* Boston, MA: D. C. Heath.

Farrell, T. S. C. (2012). Reflecting on reflective practice: (Re)Visiting Dewey and Schön. *TESOL Journal, 3,* 7-16. https://doi.org/10.1002/tesj.10

Farrell, T. S. C. (2015). *Promoting teacher reflection in second language education: A framework for TESOL professionals.* New York, NY: Routledge.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London, England: Sage.

Hatta, G. (2000). *Hanseiteki juugyou jissen: Reflective approach ni yoru eigo kyoushi no yousei* [Reflective classroom teaching practice: A reflective approach to English language teacher education]. Tokyo: Kinseido.

Hori, K. (2001). Inshi bunseki no inshisuu ketteihou SPSS script [How to determine the number of factors in factor analysis using SPSS; script file]. Retrieved from http://www.ec.kagawa-u.ac.jp/~hori/spss/spss.html#nfactors

Hori, K. (2005). Inshi bunseki ni okeru inshisuu ketteihou: Heikou bunseki o chushin ni shite [How to determine the number of factors in factor analysis: On parallel analysis]. *Kagawa University Keizaironso, 77*(4), 35-70.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55. https://doi.org/10.1080/10705519909540118

Hubley, A. M., & Zumbo, B. D. (1996). A dialectic on validity: Where we have been and where we are going. *The Journal of General Psychology, 123,* 207-215. https://doi.org/10.1080/00221309.1996.9921273

JACET SIG on English Language Education. (2014). *Seichou no tame no shousatsu tool gengokyoushi no portfolio eigo kyoushi kyouiku zenpen* [Japanese portfolio for student teachers of languages full version for English teacher education]. Tokyo: Tobunsha.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: The Guilford Press.

Koide, F. (2016). Kiroku hanashiai juugyou koukai wo toshita juugyoukaizen: J-POSTL no katsuyou jirei [Classroom teaching improvement through records, discussions, and demonstrations: Using J-POSTL]. *The English Teachers' Magazine, 64*(11), 34-36.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist, 50,* 741-749. https://doi.org/10.1037/0003-066X.50.9.741

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13,* 241-256. https://doi.org/10.1177/026553229601300302

MEXT. (2003). Eigo ga tsukaeru nihonjin no ikusei no tame no koudou keikaku [Action plan to cultivate Japanese with English abilities]. Retrieved from http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/004/siryo/04031601/005.pdf

MEXT. (2011). *Five proposals and specific measures for developing proficiency in English for international communication (Provisional translation)*. Retrieved from http://www.mext.go.jp/component/english/__icsFiles/afieldfile/2012/07/09/1319707_1.pdf

Mikami, A. (2011). Changes experienced by teachers through action research projects in an in-service EFL teacher education program. *JACET Journal, 53,* 57-74.

Mikami, A. (2015). Eigokyoushi no seichou o unagasu jikohyouka checklist no teian [Suggestions on self-evaluation checklists for EFL teachers in Japan]. *The Language Teacher, 39*(5), 19-22.

Newby, D., Allan, R., Fenner, A., Jones, B., Komorowska, H., & Soghikyan, K. (2007). *European portfolio for student teachers of languages: A reflection tool for language teacher education.* Graz, Austria: ECML, Council of Europe.

Richards, J. C., & Farrell, T. S. C. (2005). *Professional development for language teachers: Strategies for teacher learning.* New York, NY: Cambridge University Press.

Richards, J. C., & Lockhart, C. (1994). *Reflective teaching in second language classrooms*. Cambridge, England: Cambridge University Press.

Roberts, J. (1998). *Language teacher education*. London, England: Arnold.

Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record, 104,* 842-866.

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

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). New York, NY: Routledge.

Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research, 45,* 83-117. https://doi.org/10.1023/A:1006985528729

Wallace, M. J. (1998). *Action research for language teachers*. Cambridge, England: Cambridge University Press.

Zenkoku Gakkou Data Kenkyuujo. (2013). *Zenkoku gakkou soran 2014 nenban* [A comprehensive list of schools in Japan in 2014]. Tokyo: Harashobo.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432-442. https://doi.org/10.1037/0033-2909.99.3.432

Appendix

Items of the Self-Evaluation Checklist for EFL Teachers in Japan (Originally in Japanese)

1. I have a large enough English vocabulary to be an English teacher.
2. I can understand spoken English well enough to be an English teacher.
3. I can speak English well enough to be an English teacher.
4. I can read English well enough to be an English teacher.
5. I can write English well enough to be an English teacher.
6. I understand English grammar and usage well enough to be an English teacher.
7. I understand the culture of the English-speaking world well enough to be an English teacher.
8. I can explain the meaning and benefits of English language learning.
9. I can set learning goals to motivate students.
10. I can select materials appropriate for the interests of my students.
11. I can select materials appropriate for the English proficiency levels of my students.
12. I can predict the learning difficulties of my students.
13. I can flexibly prepare various questions and example sentences appropriate to the levels of understanding of my students.
14. I can take into account and use the Japanese language knowledge of my students when I teach English.
15. I am knowledgeable of the differences between English and Japanese language acquisition.
16. I am knowledgeable of teaching methods and teaching theories.
17. I can plan and conduct a lesson based on the Course of Study.
18. I can plan and conduct a lesson based on the needs of my students.
19. I can plan and conduct a lesson based on the actual status of my classes.
20. I can create an effective classroom atmosphere for English language learning.
21. I can conduct a lesson in English.
22. I can encourage my students to use English in classroom activities.
23. I can evaluate the English proficiency of my students in an appropriate way.
24. I can use whole class, small group, and pair activities effectively.
25. I can reflect on my lessons objectively.

26. In the procedure of the plan-act-reflect cycle, I am aware of problems with my students and lessons.
27. I can evaluate my lessons critically based on feedback from my students and their learning achievements.
28. I can accept feedback from my colleagues and mentors and use it in my lessons.
29. I can use related theories and research findings to improve my lessons.
30. I can evaluate the learning growth of my students in an appropriate way.
31. I can plan my lessons with other teachers.
32. I can give constructive feedback by observing the lessons of my colleagues.