

Test-taker self-assessment accuracy using the TOEIC speaking and writing can-do statements

Jonathan Harrison¹ and Ruth Vanbaelen²

jon.harrison@nihon-u.ac.jp

1. *Nihon University, College of Science and Technology*

2. *University of Tsukuba*

<https://doi.org/10.37546/JALTSIG.TEVAL26.2-2>

Abstract

Can-do statements have been integrated for decades into research and marketing strategies related to standardized tests of English for non-native language learners. For instance, while ETS indicated that TOEIC can-do statements are specifically for “decision making in organizations,” the Council of Europe created user-friendly statements that all stakeholders, including test takers, can use and understand. This study discusses a partial replication of a large-scale validity study with the purpose of developing a scale which can be used by examinees in combination with the TOEIC SW can-do statements for self-assessment. As part of an extracurricular program, 23 graduate-level EAL university students took the TOEIC SW and speaking and writing can-do self-assessment surveys. Results indicated that both self-assessment scales used in this study had moderate to strong correlations with TOEIC SW scores ($r = 0.62-0.73$) and TOEIC proficiency levels ($r = 0.47-0.66$). Self-assessment results using a Level Mean Average (LMA) scale correlated significantly with both speaking and writing test results and produced slightly higher percentages of test takers self-assessing within the true score zones of 1 SEM. For the most part, self-assessment predictive accuracy using these scales and true score zones was 52–57% for speaking and 43% for writing. Other findings were that EAL learners tended to underestimate their scores, and they more accurately self-assessed speaking than writing. If LMA scale accuracy at a particular level can be verified in future studies, claims could be made that TOEIC SW can-do statements have value for examinees.

Keywords: Assessment, can-do statements, English language learning, self-assessment, speaking, TOEIC, predictive accuracy, writing

Student self-assessment is a common educational practice, and can-do statements, which can be a form of self-assessment, were integrated into research and marketing related to international standardized tests of English language skills for English as a second language (ESL) and English as a Foreign Language (EFL), hereafter referred to as English as an Additional Language (EAL). Can-do statements were adopted by Educational Testing Service (ETS) and the Council of Europe (CoE) around 2000. However, ETS states that the can-do statements are for “decision-makers in organizations” (The Chauncey Group International, 2000), and CoE (2001) states that the can-do statements are designed to be user-friendly and educational to test takers as well as being useful for other stakeholders, e.g., potential employers. The TOEIC Speaking and Writing (SW) literature does not directly state that can-do statements can be used for self-assessment.

Self-assessment and can-do statements

There is a plethora of studies on self-assessment, and for those interested in summative literature Brantmeier, Vanderplank, and Strube (2012) and Andrade (2019) are recommended. Student self-assessment is a common educational practice that has been defined as a “descriptive and evaluative act carried out by the student concerning his or her own work and academic abilities” (Brown & Harris, 2013, p. 368) and as a “wide variety of mechanisms and techniques through which students describe (i.e., assess) and possibly assign merit or worth to (i.e., evaluate) the qualities of their own learning processes and products” (Panadero, Alonso-Tapia, & Huertas, 2012, p. 804). Andrade (2010, 2018, 2019) writes that self-assessment is feedback which should be used to inform adjustments to processes and products that deepen learning and enhance performance. Therefore, self-assessment should be formative, meaning for ongoing learning instead of being summative at the end of a project, unit, or course, because if it is not, then for the most part it lacks value.

When self-assessment is used in the classroom, researchers say that can-do statements should describe concrete tasks (Blanche & Merino, 1989) that are familiar to the target group, and self-assessment of language learner achievement needs to be designed according to specific curricular content to achieve “clear potential for predictive accuracy of criterion skills based on self-assessment measures” (Ross, 1998, p. 17). Therefore, self-assessment using can-do statements based on specific curricular content that is familiar to the learners may have better predictive accuracy and may be a formative part of the learning process.

Self-assessment in EAL

The self-assessment literature on EAL learners indicates that speaking is more accurately assessed than writing (Rasch, 1979). Also, there is research that suggests EAL learner proficiency level has an effect on the accuracy of self-assessment, and that advanced learners know which skills they are better and worse at doing (Brantmeier et al., 2012; Heilenman, 1990; Oskarsson, 1984). In regard to correlations between test scores and self-assessments, the literature shows that correlations tend to be moderate or moderately strong (Brantmeier et al., 2012; Edele, Seuring, Kristen, & Stanat, 2015). Finally, research suggests EAL learners tend to underestimate their self-assessments, especially more proficient students underestimate their abilities (Heilenman, 1990; Oskarsson, 1984). Other research suggests self-assessed language skills are systematically biased in certain groups (Edele et al., 2015); specifically, Japanese EAL learners may self-assess more harshly than might be expected because of cultural issues with modesty (Matsuno, 2009). International comparison data support the finding that Japanese students report low self-efficacy (OECD, 2019).

TOEIC can-do statements and proficiency levels

The TOEIC Can-Do Guide (2000), published by The Chauncey Group International in coordination with ETS, states that (underscore added for emphasis):

This *TOEIC Can-Do Guide* allows users of the TOEIC test to link TOEIC scores to the activities that examinees may or may not be able to do in English. The tables in the *Guide* provide examples of the activities that examinees are likely to be able to perform in English given certain Reading Comprehension scores and Listening Comprehension scores (p. 1).

In this statement, a distinction is made between users and examinees. Later, these roles are more concretely defined.

The tables on the following pages allow users of the TOEIC test to determine which English-language activities a test taker can be expected to perform given a certain TOEIC score. This is generally how decision-makers in organizations use TOEIC test scores: Test scores help them make decisions about such things as employee selection, placement, and training (p. 5). Note that “examinees” are referred to as “test takers” and “users” are more thoroughly defined as “decision-makers in organizations” who use the test scores to make human resources decisions.

The Chauncey Group International (2000) offers three caveats regarding “using the score interpretation information”, which are summarized here: Examinees self-rated their abilities to perform tasks, and “it is still unclear whether or not test takers can actually perform the English-language activities which they reported being able to do” (p. 1). The tables are based on average ratings of examinee groups. Therefore, “proficiency scales are useful for describing the abilities of a group of examinees, but can only describe what a given individual is likely to do, not necessarily what that person can actually do” (p. 3, underlined emphasis by The Chauncey Group International). This is soon followed by a similar remark: “There will be people in a given score range who may be able to do more, or fewer, tasks than are indicated. These tables are intended to serve as guidelines only and will not necessarily apply equally to every individual.” The sample is completely Japanese, yet it “may not be representative of the population of test-takers in Japan.” This is followed with “The information presented here may not apply to test takers from other countries, educational backgrounds, or industries. Decision-makers should be cautious when applying these score interpretations to other groups of employees” (p. 1).

Given these statements, (1) it is unclear if examinees can actually perform the activities, (2) these are guidelines and individual task performance will vary, and (3) the usefulness of can-do survey responses for validity purposes may be in question as the population the research was based on may not have been representative. To use an analogy, if a company employee was tasked with buying a fleet of vehicles for company use, and the salesman told him that the cars under consideration may or may not be able to be used for certain tasks, individual car performance would generally be fine but could vary individually, and the research on the cars may not represent these specific cars, it is doubtful that the employee would purchase those vehicles.

To summarize, the above statements from The Chauncey Group International focus on two main issues: the users of the can-do statements and the disclaimers regarding the research. From the above, it can be clearly stated that can-do statements are not for the examinees (test takers) but for decision-makers in organizations that use TOEIC results, and that these decision-makers “should be cautious” as examinees may or may not be able to actually perform the activities, as research methods studied groups and did not account for differences in individual performance, and as the research may or may not be representative of people inside or outside of Japan.

In 2008, a large-scale validity study was undertaken in coordination with ETS in Korea and Japan (ETS, 2018, 2019, 2021; Powers, Kim, Yu, Weng, & VanWinkle, 2009, 2010). The study used self-assessment questionnaires, a speaking test, and a writing test, and the results reportedly showed “TOEIC scores related relatively strongly to test-taker self-reports for both speaking and writing tasks” (Powers et al., 2009, p. i). This was later reworded as “moderately strong relationships of TOEIC scores with test takers’ self-reports... for both the speaking and writing language domains” (ETS, 2019, p. 20). TOEIC speaking scores and self-assessments of the ability to perform speaking tasks are correlated at ($r = 0.54$), and writing scores and the ability to perform everyday writing tasks correlate at ($r = 0.52$) and noted as correlating “reasonably strongly” (ETS, 2019, p. 20). To affirm these correlations, the ETS User Guide cites Cohen (1988) to support the belief that correlations of 0.5 can be seen as large, and then it cites Kuncel and Hezlett’s (2007) use of a 0.45 correlation between tests and first-year GPA.

One point of confusion with the original studies (Powers et al., 2009, 2010) is that both end with the following sentences. It is stated that “In the present study, we are not able to evaluate the soundness of test-taker self-reports as a validity criterion. However, in comparable studies that we have conducted recently in similar contexts, can-do self-reports have exhibited several characteristics that suggest reasonably trustworthy validity criteria” (2009, p. 14). However, the summation in the next paragraph seems to contradict what was previously stated with “In conclusion, the current study provides evidence of the validity of the TOEIC Speaking and Writing tests’ scores by linking them to test takers’ assessments of their ability to perform a variety of everyday (often job-related) English-language activities” (2009, p. 15).

CoE can-do statements and the common reference levels

The CoE (2001), which created the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR), clearly states numerous times that the can-do items are for all stakeholders, including the test takers, and they were purposely designed to be easily understandable by the test takers. They write that “the categorisation and description needs to be theoretically grounded... the description must also remain user-friendly – accessible to practitioners. It should encourage them to think further about what competence means in their context” (p. 21). The document also states that the can-do descriptors are provided for the three metacategories of reception, interaction and production (p. 25). Readers should refer to the Common Reference Levels self-assessment grid (p. 26) for examples. In Appendix D, the CoE states again that the can-do items are “user-oriented scales” that “assist communication between stakeholders in the testing process” and that they are a tool which can be used as a checklist of what language users can do and thus define the stage they are at” (p. 244).

Non-ETS related TOEIC can-do research

ETS has many publications referring to can-do statements and their tests. However, non-ETS related publications on the use of TOEIC can-do statements predictive accuracy of TOEIC scores or other can-do statements and their predictive accuracy on other standardized tests are somewhat limited.

In regard to reading and listening, Yoshizawa (2009) redesigned TOEIC can-do statements to fit the curriculum and researched test score correlations for first-year EAL university students ($n = 151$), which were 0.39 and 0.31, respectively. These were lower than the 0.46 and 0.53 in Powers, Kim, and Weng (2008), but this could be due to a neutral response being added to the TOEIC can-do survey tool. Runnels (2016) noted that some universities have shifted away from TOEIC use, for reasons including that interpretability of its scores is problematic and limited in its usefulness for test takers (p. 2), toward the CEFR “to create a synergy between the three areas of learning, teaching and assessment, resulting in a positive learning and assessment culture” (p. 3). Results from this correlation study on university students’ ($n = 57$) TOEIC scores and CEFR can-do self-assessment scores indicated that a moderate correlation was found for listening, but not for reading.

A few studies have called into question the validity of the TOEIC SW. Irwin and Nagy (2011) question the validity of the TOEIC SW based on can-do questionnaires saying, “Relying on can-do self-evaluations as a measure of validity is untenable. Future TOEIC SW Test research should explore the extent to which this test actually assesses the skills and constructs that ETS claims it measures” (p. 14). Uenishi et al. (2017) question the validity of the tests due to the fact that they are taken using a computer, which removes factors of human interaction (e.g., communication strategies and negotiation of meaning).

The present study

To sum up the introduction, the TOEIC SW literature does not directly state that can-do statements can be used by examinees for self-assessment. Therefore, because no studies have sought to replicate the Powers et al. (2009) validation study, the authors decided to partially replicate it on a small scale using the original survey questions for speaking and writing and newly collected test score results. The following research questions were thus formulated:

1. When replicated, what will the correlations be between test-taker TOEIC SW results and scaled self-assessment can-do statement survey results?
2. Using scaled self-assessment can-do surveys, to what degree can test takers accurately self-assess their speaking and writing skills to predict their TOEIC SW scores?
3. Will any other patterns emerge from the data?

Methods

Participants

The participants in this study took part in a voluntary extra-curricular graduate school English program (Harrison et al., 2022) in which the TOEIC Speaking and Writing test was offered. All 32 test takers were contacted by email and asked to take the skills self-assessment surveys (described below) via Google Forms. In total, 23 participants (22 male, 1 female) who were in their first or second year of a master's degree program in Japan completed the surveys and then the test.

The self-assessment survey tool, response analysis and response scaling

The Powers et al. (2009) survey had test takers of the TOEIC Speaking and Writing tests perform a bilingual self-evaluation, of “their ability to perform a variety of everyday (often job-related) English-language activities” (Powers et al., 2009, p. 15). As the original survey could not be procured from ETS, the self-assessment tool in Appendices A and B was recreated from ETS (English) and IIBC (Japanese) publications (Powers et al., 2009; IIBC, 2018). The tool was updated due to issues with jargon and clarity as well as discrepancies such as missing items, altered items, and translation discrepancies in items and choices. These are given in Appendix C. The tool had two parts: a 39-item speaking skills self-assessment survey and a 29-item writing skills self-assessment survey.

The authors followed the design of the original study and participants responded on a 0–5 scale:

- 0 = 答えられない (私に関係ない・判断できない) cannot answer: does not apply to me/unable to judge
- 1 = 全くできない not at all
- 2 = 非常に難しい with great difficulty
- 3 = ある程度難しい with some difficulty
- 4 =それほど難しくなく with little difficulty
- 5 = 簡単 easily

The original study matched activities that test takers in certain score proficiency levels claimed to be “able to perform, are likely to be able to perform with difficulty, and are unlikely to be able to perform” (p. 12). This was done by coding (0–5) the responses of participants with certain test scores into three categories: PND (Probably cannot do, responses 0–2), PCDwdiff (Probably can do with difficulty, responses 3–4), and PCD (Probably can do, response 5). The original study tabulated responses of test takers in each proficiency level resulting in a total percent of test takers who probably can do, probably can do with difficulty, and probably cannot do particular activities.

The results of the original research and subsequent publications (Powers et al., 2009; IIBC, 2018) placed certain skills within certain proficiency levels. Table 1 indicates the score ranges for each proficiency level. Using those results, the authors reverse engineered scales which could potentially be used to predict scores based on self-analyses using the can-do statements. The self-assessment categories were scaled to find the method which correlated best with the test-taker TOEIC SW scores and TOEIC proficiency level results. The reason that correlations were analyzed for both scores and proficiency levels was to see if differences existed. ETS reports both as results to test takers. Therefore, both seem to have value in the eyes of ETS, possibly test takers, and other stakeholders.

Self-assessment survey scaling methods

The original study (Powers et al., 2009, pp. 7-13) indicated which can-do statements were associated with each proficiency level as well as the rationale and verification process for the placements. Using those can-do item placements and the proficiency levels given by ETS and IIBC, the following scales were reverse engineered. The participant self-analysis results and the test results were used to verify which scale is potentially more accurate. Descriptions of the two scaling methods are given below.

Level midpoint average scale

The first of the two scales in this study is the Level Midpoint Average (LMA) scale. The rationale behind this scale is that the values of the midpoints of the given ETS/IIBC proficiency levels for each self-assessment item can be averaged to indicate a value that is potentially near the actual test result (score or proficiency level).

To calculate the LMA scaled scores, the midpoints of each level, Levels 1–8 for speaking and Levels 1–9 for writing, on the 200-point scale were used as a two-tier scale with responses for “Probably can do” and “Probably can do with difficulty” being awarded points, as shown in Table 1. The total points for all levels were then averaged. For example, if survey takers marked that they “Probably can do” a speaking item under Level 4 (scaled score 80–100), then they received 90 points for that item as that is the midpoint of that level. The same item is listed under “Probably can do with difficulty” in Level 2. Therefore, survey takers who marked that choice would receive 45 points for that survey item as that is the midpoint score of Level 2 (scaled score 40–50). Points for all items were then averaged to calculate the LMA self-assessment scores.

The same method was used for the scaling of proficiency levels. The actual proficiency level numbers were used for PCD and PCDwdiff calculations. These were then averaged to get the LMA self-assessment levels. For example, if survey takers marked that they “Probably can do” a speaking item under Level 4, then they received 4 points for that item. The same item is listed under “Probably can do with difficulty” in Level 2. Therefore, that choice would receive 2 points. Points for all items were then averaged to calculate the LMA self-assessment proficiency levels.

Table 1

Level midpoint average scale (number of survey items at that level)

Speaking (39 items)			Writing (29 items)		
Prof. level (Range)	PCD	PCDwdiff	Prof. level (Range)	PCD	PCDwdiff
8 (190–200)	195 (5)		9 (200)	200 (10)	
7 (160–189)	175 (15)		8 (170–199)	185 (10)	
6 (130–159)	145 (5)	145 (1)	7 (140–169)	155 (9)	155 (6)
5 (110–129)	120 (12)	120 (10)	6 (110–139)		125 (11)
4 (80–109)	95 (2)	95 (8)	5 (90–109)		100 (12)
3 (60–79)		70 (15)	4 (70–89)		
2 (40–59)		50 (5)	3 (50–69)		
1 (0–39)			2 (40–49)		
			1 (0–39)		

Note. See IIBC (2018) and Powers et al. (2009) for which items are situated at each level.

Equal weight sum scale

The second scale used in this research assigned equal weight to each item and then took the sum; hence it was termed the Equal Weight Sum (EWS) scale. The rationale behind the EWS scale is that, given a number of skills and a group of participants, a bell curve will form in regard to those who probably can do and probably cannot perform each skill. If the self-assessment items are equally weighted and then summed, the scale could potentially predict the number of skills that respondents indicated they can possibly perform. Those with fewer skills would have lower scores and those with more skills would have higher scores. A sum of the indicated skills could produce a value that is expected to correlate with actual test results (score or proficiency level).

There were 39 items on the 200-point speaking test and 29 items on the 200-point writing test, so to calculate the EWS scale each PCD item on the speaking test was 5.128 points and each of the 29 writing items was 6.896 points. PCDwdiff

scores were assigned half weight, 2.564 points and 3.448 points, respectively. Survey takers who marked that they probably could do items or probably could do items with difficulty were given their respective points, and then the scores were summed. For example, if survey takers marked that they probably could do 15 speaking items and probably could do 10 items with difficulty, then their EWS speaking score would be 103 ($[15 \times 5.128] + [10 \times 2.564]$).

For proficiency level analysis, the value of PCD for each of the 39 speaking items was 0.231 points, and for each of the 29 writing items it was 0.31 points. PCDwdiff level scores were assigned half weight, 0.116 points and 0.155 points, respectively. Survey response totals were summed as above.

Methods of obtaining TOEIC SW results

ETS and their affiliates score the speaking and writing tests separately. ETS provides test takers with both a speaking score and a writing score, and they also provide test takers with both a speaking proficiency level and a writing proficiency level. In this way, test takers can use their scores to understand where they stand within a proficiency level. Test scores and proficiency levels were used to scale the results as well as for correlations and other analyses.

Results

Statistical descriptions and distributions

Test and self-assessment results for both scores and proficiency levels for this research sample ($n = 23$) are shown in Table 2. For speaking scores and proficiency levels, the EWS scale results were more consistent with the test scores in regard to ranges and means. The LMA scale produced lower maximums and lower means, but the SDs and SEMs were more consistent with test scores. For writing scores and proficiency levels, the LMA scale produced ranges that were more consistent with the test ranges. The mean score of the EWS scale was slightly closer to the mean of the test, but for proficiency levels the LMA scale mean was closer to the test results. The SDs and SEMs with these scales were similar.

Table 2

Descriptive statistics for TOEIC SW results and both self-assessment scales

	Scores			Proficiency levels		
	TOEIC	LMA	EWS	TOEIC	LMA	EWS
	Speaking					
Range	30–150	2–128	3–156	1–6	0–5	0–7
Mean	69	49	64	3.17	2.09	2.91
SD	28	33	42	1.17	1.35	1.88
SEM	6	7	9	0.24	0.28	0.39
	Writing					
Range	30–140	3–163	3–172	3–7	0–7	0–8
Mean	88	72	68	4.35	3.38	3.07
SD	33	52	52	0.87	2.37	2.36
SEM	7	11	11	0.18	0.49	0.49

Normal distributions have 68% of the scores within 1 SD, 95% within 2 SDs, and 99% within 3 SDs of the mean. The test speaking score distribution was near-normal with 61% of scores within 1 SD, but with writing scores 57% were within 1 SD, and 100% were within 2 SDs. For speaking self-assessments, both self-assessment scales revealed score results with near-normal distributions having 74% of scores within 1 SD. For writing scores, the LMA scale results had the same distribution as the test. The EWS scale writing score results were closer to a normal distribution with 61% within 1 SD, 96% within 2 SDs and 100% within 3 SDs.

Regarding the proficiency level distributions for speaking, the test results formed a near-normal distribution, but the data were closer to the mean with 83% within 1 SD. Both self-assessment proficiency level results were distributed the same with near-normal distributions with 74% within 1 SD. For the writing test results, the proficiency level distributions within 1, 2 and 3 SDs were 30, 70, and 96%, respectively, which means the data points were further from the mean than in the case of a normal distribution. Both of the proficiency level self-assessment result distributions showed 74% within 1 SD

and 100% within 2 SDs, which means overall the data were closer to the mean.

Correlations

The correlation (r) between TOEIC speaking and writing scale scores was 0.66 (ETS, 2021, p. 22), and the correlation of the participants' ($n = 23$) speaking and writing scores in this report was 0.78. This was perhaps due to the limited sample size and a narrower range of test-taker proficiencies.

ETS reported correlations between test results and self-assessments as 0.54 for speaking and 0.52 for writing (ETS, 2019). To answer RQ1 (When replicated, what will the correlations be between test-taker TOEIC SW results and scaled self-assessment can-do statement survey results?), correlations between LMA and EWS self-assessments and TOEIC SW scores and the proficiency levels for the current sample are shown in Table 3. Both the LMA and EWS scaled scores had higher correlations with the test scores than the self-assessment proficiency levels had with the test proficiency levels. The EWS scale scores and proficiency levels correlated slightly higher with test results than those of the LMA scale. The LMA speaking, LMA writing, and EWS writing self-assessments correlated significantly ($p < .05$), indicated with an asterisk in Table 3. Only the EWS speaking score and proficiency level self-assessments did not correlate significantly with the test results. Therefore, both LMA and EWS scales correlated highly with test results, and the LMA scale seems to be the better of the two scales as it correlated significantly with both speaking and writing scores and proficiency levels.

Table 3

Correlations of test and self-assessment results

	Scores				Proficiency levels			
	r	t	df	p	r	t	df	p
LMA speaking	0.73	4.18	22	0.0003*	0.65	4.78	22	0.00009*
EWS speaking	0.73	0.77	22	0.45	0.66	0.87	22	0.39
LMA writing	0.62	2.74	22	0.006*	0.47	2.16	22	0.02*
EWS writing	0.63	2.26	22	0.017*	0.54	2.97	22	0.004*

Note. Significant at $p < .05$

“True score” zone, proficiency level, and self-assessment zone results

ETS reports TOEIC SW score results on scales of 0–200 and by proficiency levels, and they also provide information on the standard error of measurement (SEM) in the *Score User Guide TOEIC Speaking & Writing Tests*. For the TOEIC SW, the SEM for speaking is 13 scale points, and for writing it is 17 points (ETS, 2021, p. 21). According to the guide, this “can be used to estimate the expected variation of each test taker’s observed scores around their ‘true score.’...the true score can be thought of as the average score a test taker would earn on a very large number of forms...and across a number of testing dates.” The guide further explains that scores “would be expected” to be within the SEM 68% of the time (p. 22).

To answer RQ2 (Using scaled self-assessment can-do surveys, to what degree can test takers accurately self-assess their speaking and writing skills to predict their TOEIC SW scores?), the percentage of survey takers who self-assessed within the true score zone (1 SEM) of their test scores was 26–39%, and 9–35% self-assessed within the proficiency levels reached through the tests, shown in Table 4.

ETS used “true score” zones of 1 SEM. If the SEMs of the self-assessment scales, given in Table 2, are considered as “true score” zones, then 52–57% of test takers accurately self-assessed speaking and 43% accurately self-assessed writing.

Table 4

Percentage of participants within the true score zone or proficiency level

	Speaking			Writing		
	True zone (SEM 13)	Proficiency level	True + SA zone	True zone (SEM 17)	Proficiency level	True + SA zone
LMA	39	35	57	30	9	43
EWS	35	35	52	26	17	43

Note. Self-assessment is abbreviated as SA.

Although not shown in the table, if self-assessments for both speaking and writing are considered together, both the LMA and EWS score results indicated that 17% of survey takers self-assessed both tests accurately, and those that assessed accurately were of high, middle, and low proficiencies. Regarding proficiency level assessment with the LMA scale, no one self-assessed accurately for both speaking and writing. With the EWS scale, 9% of self-assessments were accurate for both speaking and writing, and they were in the middle- and high-proficiency ranges.

When TOEIC true score zones and speaking and writing self-assessment scale true score zones of 1 SEM were considered together, 26% of participants on the LMA scale and 22% of participants on the EWS scale self-assessed both speaking and writing accurately. These participants were in the top and bottom, not middle, proficiency groupings. This suggests that in total 22–26% of participants self-assessed accurately on both speaking and writing, and an additional 17–31% of participants assessed either speaking or writing accurately.

Speaking result visualizations

Answers to RQ3 (Will any other patterns emerge from the data?), are discussed forthwith. Figure 1 shows test-taker TOEIC scores, the “true score” zones, and the LMA and EWS self-assessment scores, and Figure 2 shows the proficiency levels. Participant order is from high to low based on speaking score test results. As can be seen, self-assessment scores with both methods fluctuated above and below the true score zone.

Figure 1

Speaking test scores and true score zone with LMA and EWS score self-assessments

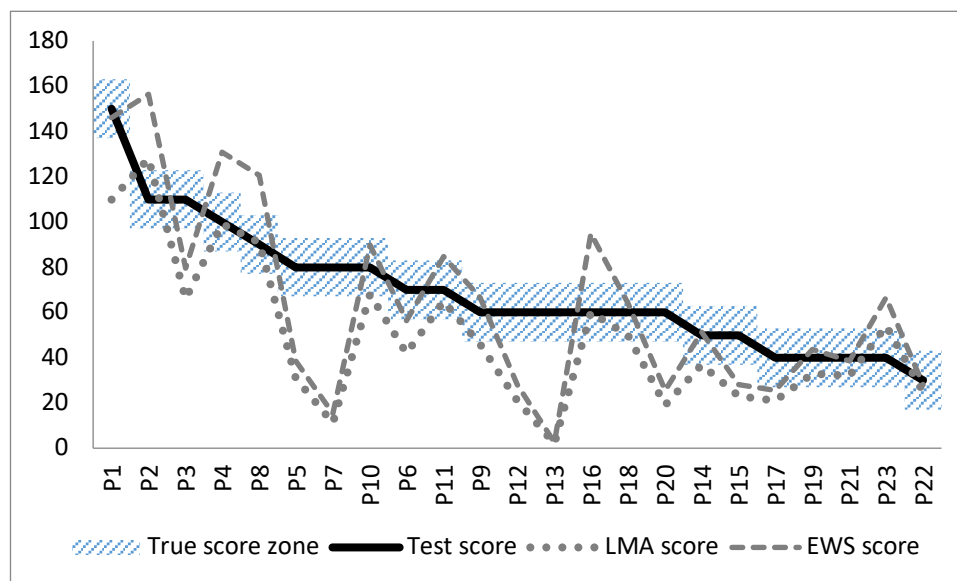
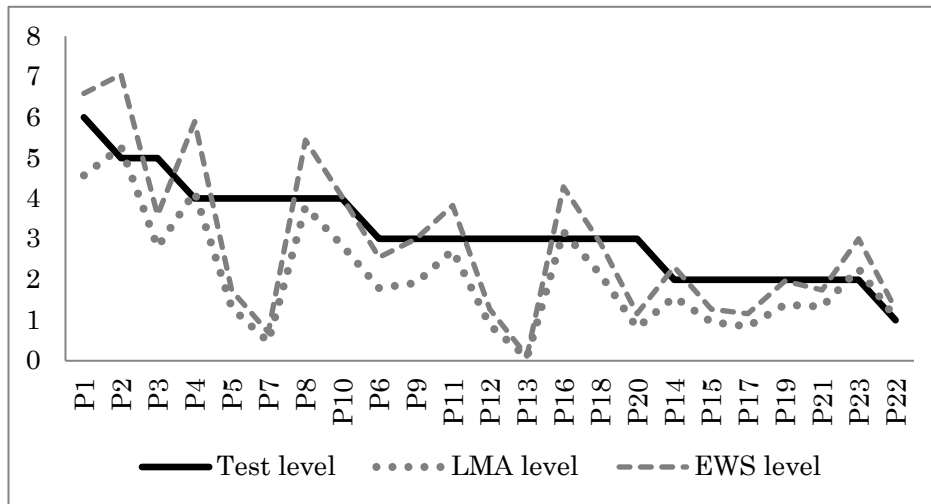


Figure 2

Speaking test, LMA, and EWS proficiency levels



Writing result visualizations

Results for writing are shown in Figures 3 and 4. As the speaking and writing tests were separate, the order of participants is from high to low based on writing test results, but participant numbers (e.g., P1, P2, are in order by writing score). Writing self-assessment scores with both methods fluctuated above and below the true score zone.

Figure 3

Writing test scores and true score zone with LMA and EWS score self-assessments

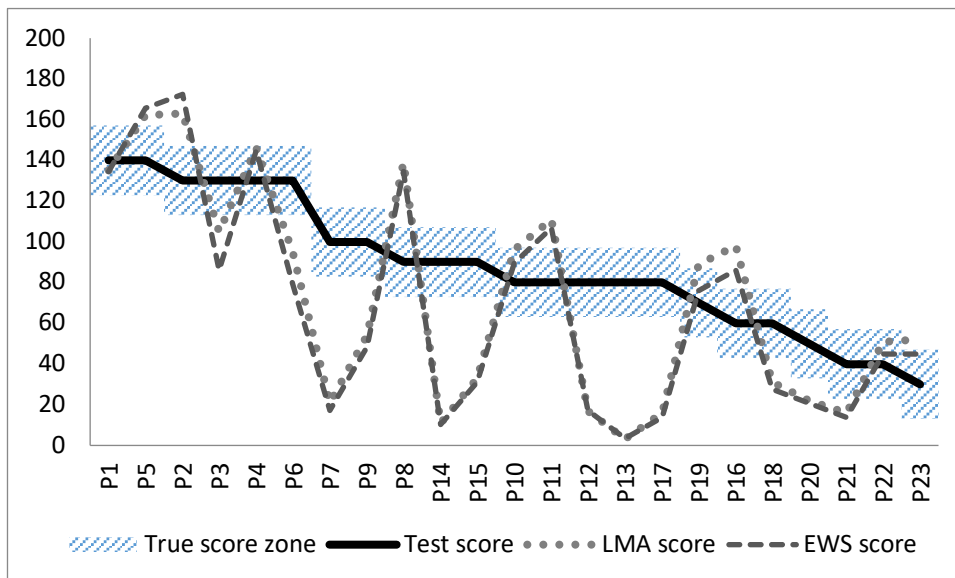


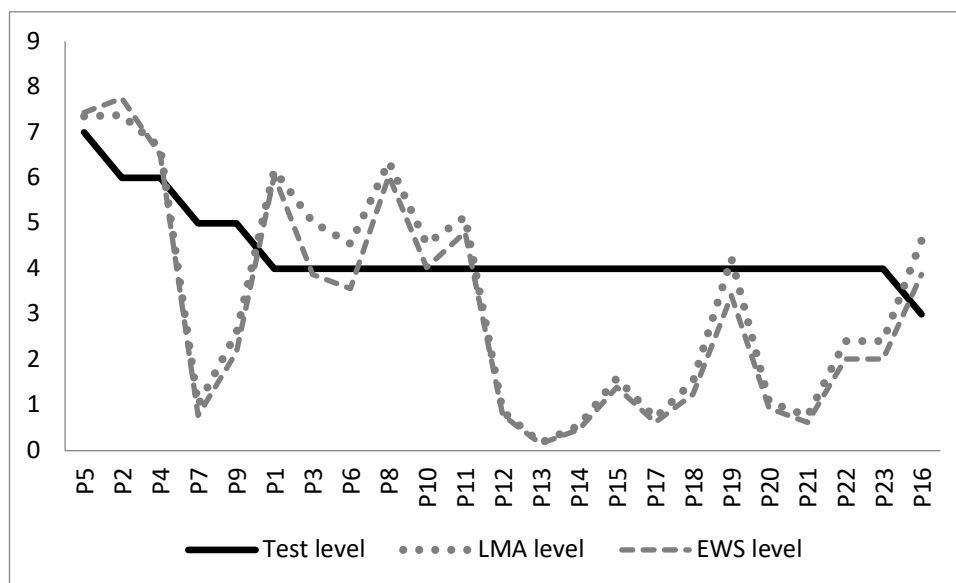
Figure 4*Writing test, LMA, and EWS proficiency levels*

Figure 4 shows that the majority of participants in the middle and bottom proficiency groupings self-assessed very low, but the reasons were unclear. The majority of items for writing are set at the higher proficiencies (Levels 7–9), and it could be that survey takers became tired with the length of the survey or discouraged with the difficulty of the items and filled in answers leading to scores of zero under “Probably cannot do.”

Differences between self-assessment and test results

The differences between self-assessments and test results per participant were calculated to understand if participants were underestimating or overestimating their results. The results in Table 5 indicated that overall, on both scales participants underestimated their results. The minimums of the ranges, which were negative, were further from zero than the maximum values, and the mean values for scores and proficiency levels were all negative.

Table 5*Differences between self-assessments and test results*

	Scores		Proficiency levels	
	LMA	EWS	LMA	EWS
Speaking				
Range	-69–18	-65–46	-3.54–0.26	-3.31–2.05
Mean	-20	-5	-1.08	-0.26
SD	22	29	1.06	1.42
Writing				
Range	-81–38	-83–44	-4.00–2.36	-4.23–2.05
Mean	-22	-20	-0.97	-1.28
SD	38	41	2.1	2.02

The results in Table 5 indicated that the self-assessment results generated with these scales tended to be underestimations, and for both scales raw data tallies indicated more students underestimated in all cases. With the LMA scale, 87% ($n = 20$) for speaking and 57% ($n = 13$) for writing self-assessed lower than their scores. With the EWS scale, respectively, 52% ($n = 12$) and 57% ($n = 13$) self-assessed low. Regarding proficiency level assessment for speaking with the LMA scale 65% ($n = 15$) underestimated, and 35% ($n = 8$) were perfect. For speaking levels with the EWS scale, 35% ($n = 8$) were below, 35% ($n = 8$) were in alignment with the true score zone, and 30% ($n = 7$) were high. For proficiency levels of writing with the LMA scale, 52% ($n = 12$) were below, 9% ($n = 2$) were perfect, and 39% ($n = 9$) were high. For writing levels with the

EWS scale, 57% ($n = 13$) were below, 17% ($n = 4$) were perfect, and 26% ($n = 6$) were high.

To analyze the data by proficiency groupings, the participants were divided into roughly equal groupings based on their test results. These groupings of top, middle, and bottom are relative to each other in this analysis and not to proficiency levels in general or as defined by any organization. Table 6 shows results with groupings of 7, 9, and 7 participants; however, groupings of 8, 7, and 8 participants yielded similar results. Results in Table 6 indicated that when analyzed by proficiency groupings, all proficiency groups underestimated their test results for both scores and proficiency levels. Regarding scores, for both speaking and writing the middle-proficiency groups underestimated their scores the most, followed by the top and then the bottom groups. Regarding proficiency levels, the EWS speaking scale followed this trend, but the LMA speaking scale showed the amount of underestimation in descending order from top to bottom groups. For writing proficiency levels, the bottom and middle group participants underestimated more than the top proficiency groups, and this was believed to be because the majority of these participants tested into the same proficiency level (Level 4).

Table 6

Differences between self-assessment and test results by proficiency grouping

	Speaking				Writing			
	LMA scores	EWS scores	LMA proficiency level	EWS proficiency level	LMA scores	EWS scores	LMA proficiency level	EWS proficiency level
Top ($n=7$)	-22	-3	-1.38	-0.14	-20	-14	-0.12	-0.35
Middle ($n=9$)	-23	-8	-1.29	-0.54	-37	-35	-1.28*	-1.57*
Bottom ($n=7$)	-9	-2	-0.51	-0.04	-4	-5	-1.43*	-1.84*

Note. Proficiency level is abbreviated as *prof. lvl.* *Test results place the majority of these participants at proficiency level 4.

Results for absolute values of the differences between self-assessments and tests

The absolute values of the differences between self-assessments and test results were calculated to measure the distance by which participants were either underestimating or overestimating. Regarding scores, for speaking, self-assessments were on average 23 points off and for writing 36–38 points off the mark. Regarding proficiency levels, for speaking they were on average 1.11–1.16 levels off and for writing 2.02 levels different from the test results. This demonstrated that participants more accurately self-assessed their speaking than their writing in regard to both scores and proficiency levels.

Table 7

Differences between self-assessment and test result absolute values

	Scores		Proficiency levels	
	LMA	EWS	LMA	EWS
	Speaking			
Range	0–69	1–65	0.05–3.54	0.01–3.31
Mean	23	23	1.16	1.11
SD	19	18	0.98	0.92
	Writing			
Range	2–81	5–83	0.22–4.00	0.03–4.23
Mean	36	38	2.02	2.02
SD	25	24	1.14	1.29

When results were calculated by proficiency levels with nearly equal numbers of participants in each group, the results, in Table 8, indicated that for speaking scores and proficiency levels, the bottom group self-assessed most accurately, then the middle group and then the top group. Regarding self-assessment scores for writing, the results showed that the bottom group self-assessed most accurately, then the top group, and then the middle group. Yet, for proficiency levels the top group was the most accurate, followed by the bottom group, and then the middle group. However, the bottom and the middle group were mostly the same level according to the test results. For this reason and due to the limited sample size, these results were not considered conclusive for writing. Groupings of 8, 7, and 8 participants yielded similar results.

Table 8*Differences between self-assessment and test result absolute values by proficiency grouping*

	Speaking				Writing			
	LMA scores	EWS scores	LMA proficiency level	EWS proficiency level	LMA scores	EWS scores	LMA proficiency level	EWS proficiency level
Top (n=7)	29	32	1.50	1.85	33	40	1.71	1.70
Middle (n=9)	22	24	1.33	1.02	46	51	2.30*	2.21*
Bottom (n=7)	17	13	0.60	0.49	27	26	1.95*	2.09*

Note. *Test results placed the majority of these participants at proficiency level 4.

Discussion

Based on the results presented in this study, strengths and weaknesses of the self-assessments can be discussed. First, to review the answer to RQ1, self-assessments using both the LMA and the EWS scales correlated highly with TOEIC SW scores and proficiency levels, and the correlations were significant for both speaking and writing using the LMA scale. Correlations between test scores and self-assessments in this study were moderate to strong, which is in line with previous studies.

Second, as a review of the answer to RQ2, we will focus on the LMA scale because it outperformed the EWS scale. The percentage of survey takers who self-assessed within the true score zone (1 SEM) of their test scores was 30-39% and within their proficiency level it was 9-35%. When “true score” zones of 1 SEM were considered for both the tests and the self-assessments, then 57% of test takers accurately self-assessed speaking and 43% accurately self-assessed writing. These results are consistent with the literature (Rasch, 1979), as EAL learners more accurately assessed speaking than writing.

Finally, regarding RQ3, there were two other patterns that emerged from the data. The first was that with writing, the majority of participants in the middle and bottom proficiency groupings self-assessed very low, but the reasons were unclear. The self-assessment results generated with both scales tended to be underestimations. With the LMA scale, 87% ($n = 20$) for speaking and 57% ($n = 13$) for writing self-assessed lower than their scores. This underestimation by EAL learners is consistent with other studies (Edele et al., 2015; Heilenman, 1990; Matsuno, 2009; OECD, 2019; Oskarsson, 1984).

The second was a pattern of irregularity regarding participant results when grouped by proficiency (top, middle, and bottom). For writing proficiency levels in this study, the top group self-assessed most accurately, which follows other studies (Brantmeier et al., 2012; Heilenman, 1990; Oskarsson, 1984) that suggest that EAL learner proficiency level has an effect on accuracy of self-assessment and that advanced learners know which skills they are better and worse at doing. However, contradicting previous studies, writing score results and speaking score results in this study demonstrate that the bottom group self-assessed most accurately. This could be due to the small sample size and that in this study proficiency groupings were used and levels of proficiency were not defined by any outside criteria.

Limitations

This study has multiple limitations. First, this is only a partial replication study, as the original study was used to set the can-do items within the proficiency levels to indicate to stakeholders what skills test takers may or may not actually possess at those levels. This study used the can-do items and participant test results to reverse engineer scales which could potentially be used by learners and educators in combination with the can-do items to predict test results.

Second, results may differ because the original survey in Powers et al. (2009) had to be recreated and differences in wording and translations between the original English and IIBC publications in Japanese had to be reconciled. Items also had to be adjusted due to changes in society and technology. This could have affected the results.

Third, in this study the sample size was limited, and the testing situation was extracurricular. The self-assessment accuracy would probably improve with a larger sample size of learners who are familiar with the concrete tasks listed in the can-do items, via curricular content and experience. If the self-assessment could be designed to be a formative part of the learning process, it would benefit the learners more.

Conclusion

In summary, the background literature indicated that TOEIC can-do statements are for decision-makers in organizations and not specifically for test takers. This study indicated that TOEIC SW can-do statement survey responses scaled using the LMA scale could be used for predictive purposes because they correlated strongly with TOEIC SW scores and moderately with proficiency levels. These correlations were significant at $p < .05$. Regarding predictive accuracy of self-assessments using the LMA scale and true score zones of 1 SEM, participants accurately predicted their TOEIC SW scores 57% of the time for speaking and 43% for writing. Further research is needed with a larger sample of participants to confirm the accuracy and reliability of the LMA scale developed in this study. In the future, if such research supports the use of the LMA scale, then claims could be made by educators and by ETS that for test takers the can-do statements can be useful at a particular degree of predictive accuracy.

References

- Andrade, H. (2010). Students as the definitive source of formative assessment: academic self-assessment and the self-regulation of learning. In H. Andrade and G. Cizek (Eds.) *Handbook of Formative Assessment* (pp. 90-105). Routledge.
- Andrade, H. L. (2018). Feedback in the context of self-assessment. In A. Lipnevich and J. Smith (Eds.) *Cambridge Handbook of Instructional Feedback* (pp. 376-408). Cambridge University Press.
- Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4, 87, 1-13. <https://doi.org/10.3389/educ.2019.00087>
- Blanche, P., & Merino, B. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning*, 39 (3), 313-338. <https://doi.org/10.1111/j.1467-1770.1989.tb00595.x>
- Brantmeier, C., Vanderplank, R., & Strube, M. (2012). What about me?: Individual self-assessment by skill and level of language instruction. *System*, 40 (1), 144-160.
- Brown, G. T., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.) *Sage Handbook of Research on Classroom Assessment* (pp. 367-393). Sage. <https://dx.doi.org/10.4135/9781452218649.n21>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Erlbaum.
- Council of Europe. (2001). *The Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.
- Edele, A., Seuring, J., Kristen, C. & Stanat, P. (2015). Why bother with testing? The validity of immigrants' self-assessed language proficiency. *Social Science Research*, 52, 99-123.
- ETS. (2018). *TOEIC User Guide Speaking & Writing Tests*. Educational Testing Services.
- ETS. (2019). *Score User Guide TOEIC Speaking & Writing Tests*. Educational Testing Services.
- ETS. (2021). *Score User Guide TOEIC Speaking & Writing Tests*. Educational Testing Services.
- Harrison, J., Suzuki, T., & Falout, J. (2022). 2018-2019 Graduate School English Speaking and Writing Proficiency Program. *日本大学理工学部一般教育教室彙報 (Nihondaigaku Rikougakubu Ippan Kyouiku Ihou, Bulletin of the Department of General Education of the College of Science and Technology at Nihon University)*, 112, 1-8.
- Heilenman, L. K. (1990). Self-assessment of second language ability: the role of response effects. *Language Testing*, 7(2), 174-201. <https://doi.org/10.1177/026553229000700204>
- IIBC. (2018). *TOEIC Speaking & Writing Tests のご案内* (No goannai, A guide to). The Institute for International Business Communication.
- Irwin, B., & Nagy, P. (2011). The 2010 revision of the TOEIC speaking test. *SHIKEN: JALT Testing & Evaluation SIG Newsletter*, 15(2), 10-19.
- Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, 315, 1080-1081.
- Matsuno, S. (2009). Self-, peer- and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26 (1), 75-100.
- OECD. (2019). *PISA 2018 Results (Volume III): What School Life Means for Students' Lives*. PISA, OECD Publishing. <https://doi.org/10.1787/acd78851-en>
- Oskarsson, M. (1984). *Self-Assessment of Foreign Language Skills: A Survey of Research and Development Work*. Council for Cultural Cooperation.

- Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, 22, 806–813. <https://doi.org/10.1016/j.lindif.2012.04.007>
- Powers, D. E., Kim, H.-J. & Weng, V. Z., (2008). *TOEIC Compendium Study: The Redesigned TOEIC (Listening and Reading) Test: Relations to Test-Taker Perceptions of Proficiency in English*. (ETS Research Rep. RR-08-56). Educational Testing Services. <https://www.ets.org/Media/Research/pdf/TC-10-06.pdf>
- Powers, D. E., Kim, H.-J., Yu, F., Weng, V. Z., & VanWinkle, W. (2009). *The TOEIC Speaking and Writing Tests: Relations to Test-Taker Perceptions of Proficiency in English*. (ETS Research Rep. RR-09-18). Educational Testing Services. <https://doi.org/10.1002/j.2333-8504.2009.tb02175.x>
- Powers, D. E., Kim, H.-J., Yu, F., Weng, V. Z., & VanWinkle, W.H. (2010). *The TOEIC Speaking and Writing Tests: Relations to Test-Taker Perceptions of Proficiency in English*. Compendium Study. Educational Testing Services.
- Rasch, G. (1979). *Probabilistic Models for Some Intelligence and Achievement Tests*. Danish Institute for Educational Research (Expanded edition, 1980. University of Chicago Press, Chicago).
- Runnels, J. (2016). Self-assessment accuracy: Correlations between Japanese English learners' self-assessment on the CEFR-Japan's can-do statements and scores on the TOEIC. *Taiwan Journal of TESOL*, 13(1), 105-137.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experimental factors. *Language Testing*, 15, 1-20.
- The Chauncey Group International. (2000). *The TOEIC Can-Do Guide: Linking TOEIC Scores to Activities Performed Using English*. https://www.ets.org/Media/Research/pdf/TOEIC_CAN_DO.pdf
- Uenishi, K., Sakaue, T., Lauer, J., Davies, W., Fraiser, S., Howell, P., Selwood, J., Song, K., Morita, M., & Kida, S. (2017). Integrating the TOEIC speaking test with small group classes. *広島外国語教育研究 (Hiroshima Gaikokugo Kyouiku Kenkyu, Hiroshima Studies in Language and Language Education)*, 20, 1-15. <https://doi.org/10.15027/42610>
- Yoshizawa, K. (2009). To what extent can self-assessment of language skills predict language proficiency of EFL learners in school context in Japan? *関西大学外国語教育研究 (Kansai Daigaku Gaikokugo Kyouiku Kenkyu, Kansai University Research in Foreign Language Education)*, 17, 65-82.

Appendices

Appendix A: Self-Report on English Language Speaking Skills for Business

ビジネス場面における英語のスピーキングスキルに関する自己報告

Self-Report on English Language **Speaking** Skills for Business

- 0 = 答えられない (私に関係ない・判断できない) cannot answer: does not apply to me/unable to judge
 1 = 全くできない not at all
 2 = 非常に難しい with great difficulty
 3 = ある程度難しい with some difficulty
 4 =それほど難しくなく with little difficulty
 5 = 簡単 easily

英語のスピーキングなら、私は: Regarding my **English speaking**, I can:

1. メニューを見ながら、喫茶店やレストランで料理を注文することができる
use a menu, order food at a café or restaurant
2. 人と会う約束を取る・変更する・キャンセルすることができる
make/change/cancel an appointment to see a person
3. 非公式 (インフォーマル) の場面で、自分の趣味や故郷、興味のあることについて話すことができる
discuss my hobby, hometown, or interests when introducing myself in informal situations
4. トイレの場所の説明等、来訪者の案内をすることができる
show a guest around (e.g., where the restroom is)
5. 留守番電話に、折り返し電話をもらえるよう伝言を残すことができる
leave a message on an answering machine/voicemail to ask a person to call me back
6. 電話で自分の名前・用件を告げ、担当者への取り次ぎを依頼することができる
give my name and the reason I called and ask to connect me to a person in charge
7. レストランの営業時間を問い合わせたり、夕食の予約をしたりすることができる
request information about business hours of a restaurant and make dinner reservations
8. 最寄り駅から、特定の目的地 (オフィス等) までの道順を説明することができる
give directions on how to get to some destinations (e.g. the office) from the nearest station
9. 航空会社に電話して、飛行機の予約を変更することができる
telephone the airline to change my flight plans
10. 電話で、商品の注文や注文状況の問い合わせをすることができる
telephone a company to place (or follow-up) an order for an item
11. 社交の場面で、適切な挨拶・自己紹介をすることができる
use appropriate greeting expressions and introduce myself in social situations
12. 来訪者と商談前に、天気等の一般的な話題について、世間話をするすることができる
have “small talk” with a guest about topics of general interest (e.g. the weather) before discussing business
13. 海外からの来訪者に、ちょっとした観光案内をすることができる
give a brief sightseeing tour to a guest from abroad
14. 外国人の同僚や新人に、日常業務の手順を口頭で教えることができる
tell a foreign colleague or newly employed person how to perform a routine task
15. 自分の会社 (または大学) について (例: 主要事業は何か、どのような履修コースがあるか)、資料を見ながら手短かに説明することができる
briefly explain about my company (or university) while looking at some data (e.g. what the main business is or what courses it offers)
16. 将来のキャリア上の目標について話すことができる (例: 来年自分が何をしようとしているか話すことができる)
talk about my future professional goals (e.g., what I plan to be doing next year)
17. 今起きているトラブル (例: フライト・宿に関するトラブル) を係員に説明し、問題を解決するよう要求

- することができる
- explain ongoing troubles (e.g., about flight or hotel accommodations) and make a request to settle a problem
18. 自分の学歴や現在の職務を、フォーマルな場面で説明することができる
describe my academic training or my present job responsibilities in formal settings
19. 自分の意見を、異なる意見を持つ相手に説明することができる
express my opinion to others who may have a different view
20. 同僚や仲間に、使用方法を知っている機械や機器（例：コピー機、PC、オーディオプレーヤー）の操作の仕方を説明することができる
explain (to a co-worker or colleague) how to operate a machine or device (e.g., photocopier, PC, audio player) that I am familiar with
21. 非公式（インフォーマル）の場面で、会話等の通訳をすることができる
translate (e.g., conversations) in an informal setting
22. 外国人の従業員に、会社の歴史や規則について説明することができる
explain company history and rules (in English) to foreign employees
23. 自分がよく知っている仕事関連のプロセスについて、説明することができる
describe a familiar work-related process
24. その場に適した暗記したフレーズを使って、質問したり会話を行ったりすることができる
ask a question and talk by using memorized phrases and expressions correctly in appropriate situations
25. 上司や同僚と、顧客サービスや製品の質の向上について話し合うことができる
discuss the improvement of customer service/product quality with my boss/co-workers
26. 事業計画や方針について、新しい仲間や同僚に説明することができる
explain a business plan or policy to a new colleague or co-worker
27. 議論や会議の場で、自分の意見を述べたり力説したりすることができる
state and emphasize my opinion during a discussion or meeting
28. 議論の場で、相手の意見に対して自分の意見を述べたり反論したりすることができる comment on or react to someone's opinion during a discussion
29. 事前に準備をした上で、新規プロジェクトや新製品に関する 20～30 分のプレゼンテーションや、公式（フォーマルな）スピーチを行うことができる
give a prepared 20-30 minute presentation or formal talk (e.g., about a new project or product)
30. プレゼンテーションやスピーチで議論されている内容について、質問をしたり質問に答えたりすることができる
ask or answer questions about an issue being discussed in a presentation or speech
31. 講義やプレゼンテーションの際、不明瞭な問題点についてその場で質問することができる
ask questions to understand unclear or problematic points made in a lecture or presentation
32. 教師や同僚、友人、仲間等、聞き手の経歴や英語力に応じて話し方を変えることができる
adjust my speaking for different learners (e.g., instructors, co-workers, friends, colleagues, according to their backgrounds and English proficiency)
33. 事前に用意した資料を用いて、他社の製品と比較しながら自社の製品やサービスの品質、効率性、価格などを説明することができる
using prepared materials, discuss my company's products/services, comparing them with other companies' products in terms of quality, efficiency, and prices
34. 来訪者と世界の出来事について話すことができる
discuss (in English) world events with a guest
35. 少人数の歓送迎会で、事前準備なしに 2～3 分程度の短い挨拶やスピーチをすることができる
without any preparation, give a short (a few minutes) greeting or speech at a small welcome/farewell party
36. 講演やプレゼンテーション等、フォーマルな場面で通訳することができる
translate (e.g., lectures or presentations) in a formal setting
37. 自分が行った研究や調査に関して、結果を報告することができる
report on the outcomes of my research or investigation
38. 自分の専門分野に関係する話題について、分かりやすく伝えることができる communicate in an understandable way about topics related to my field or specialty
39. 不良品や不十分なサービスに対する苦情を処理することができる
deal with complaints about defective product or inadequate service

Appendix B: Self-Report on English Language Writing Skills for Business

ビジネス場面における英語のライティングスキルに関する自己報告

Self-Report on English Language **Writing** Skills for Business

- 0 = 答えられない (私に関係ない・判断できない) cannot answer: does not apply to me/unable to judge
 1 = 全くできない not at all
 2 = 非常に難しい with great difficulty
 3 = ある程度難しい with some difficulty
 4 =それほど難しくない with little difficulty
 5 = 簡単 easily

英語のライティングなら、私は: Regarding my English writing, I can:

1. ホテルの宿泊設備に関する情報を求める E メールを書くことができる
write an email requesting information about hotel accommodations
2. 仕事上の知り合い宛てに、プレゼントや食事に対する短い礼状を書くことができる
write a short thank-you note to a business acquaintance for a gift or dinner
3. 同僚または教師宛てに、会議または授業に出席できなかった理由を説明した簡潔なメモを書くことができる
write a brief note to a co-worker (or instructor) explaining why I was not able to attend a meeting (or class)
4. 歓送迎会のご案内の E メールを書くことができる
convey information in an email about a welcome or farewell party
5. 業務上または授業に関する確認のためのメモを書いたり、E メールを送ったりすることができる
write a memorandum or send an email confirming some information related to my job or class
6. 上司または教師宛てに、この一週間で達成したことを説明する E メールを書くことができる
write an email to my supervisor (or instructor) explaining what I have accomplished during the past week
7. 自分のオフィスへの道順を分かりやすく書くことができる
write clear directions on how to get to my office
8. 商品やサービスに対する、苦情の E メール・文書を企業に対して書くことができる
write a letter or send an email to a company to complain about a product or service
9. ある特定の商品やサービスに関する情報を得るために、企業に対し E メール・文書で問い合わせることができる
write a letter or send an email to a company to request information about a certain product or service
10. 公的機関に対し、必要な情報を得るために E メール・文書で問い合わせることができる send an email or write a letter to a public organization requesting necessary information
11. ビジネスレターの返事を書くことができる
write a reply to a business letter
12. 求人に応募するにあたり、自己紹介し、その仕事に自分が適任であることを説明する手紙を書くことができる
write a letter introducing myself and describing my qualifications for a job
13. 自分が作成した報告書内の図や表で示された情報について、説明文を書くことができる
describe (in writing) information presented in a table/chart of a report that I have written
14. 自分の専門分野に関連したグループ会議で話し合われた要点を、簡単なレポートやメモにまとめることができる
write a brief report (or memo) summarizing the main points discussed in a group meeting related to my field or specialty
15. ビジネスレターやマニュアル等の書類を、英語に翻訳することができる
translate documents (e.g., business letters, manuals) into English
16. 同僚に、使用方法を知っているコピー機や PC 等、オフィス機器の操作の仕方を説明する短いメモを書くことができる
write a short note (to a co-worker) describing how to operate an office machine or device (e.g., photocopier, PC) that I

am familiar with

17. 会議または授業中に議論された内容をメモし、要点をまとめることができる
write discussion notes during a meeting or class and summarize them
18. 上司または教師宛てに、プロジェクトまたは課題の進行状況を説明するメモを書くことができる
write a memorandum to my supervisor or instructor describing progress on a current project or task
19. 潜在顧客宛てに、自社の製品・サービスを説明する手紙を書くことができる
write a letter to a potential client describing the services or products of my company
20. 自分がよく知っている専門書に関して、1~2 ページの要約を書くことができる
write a 1-2 page summary of a technical book that I am familiar with
21. 顧客に対する、フォーマルな礼状を書くことができる
write a formal letter of thanks to a client
22. 自分の専門分野内のよく知っている話題に関して、専門的な報告書を書くことができる
write a technical report on a familiar topic within my area of expertise
23. 自分の仕事に関連する、計画書や提案書を書くことができる
write a plan or proposal related to my work
24. 専門家による会議（または学会で）のプレゼンテーションで使用する文章やスライドを作成することができる
prepare text and slides (in English) for a presentation at a professional conference
25. 上司または教師から割り当てられた、調査報告書を書くことができる
write a research report assigned by a supervisor or instructor
26. 商品やサービスのセールスマニュアルを作成することができる
write a manual (for sales purposes) for a product or service
27. 新しい業務プロセスや新商品・新サービスを提案する企画書を書くことができる
write a proposal to suggest a new business process, a new product, or a new service
28. プロジェクトの進行状況を説明する、数ページにわたる正式な報告書を書くことができる
write a brief, several-page (formal) report explaining the progress being made on a current project
29. 新商品や新サービスを紹介するプレスリリースを書くことができる
create a press release to introduce a new product or service

Appendix C: Can-Do Item Edits in Self-Reports due to Variations in Sources

Regarding recreation of the survey, the Japanese Can-Do List (IIBC) was aligned with the English Can-Do List (Powers et al., 2009). The Can-Do Writing Lists by Powers and IIBC (2018) do contain the same 29 items, but the English and Japanese Can-Do Lists vary in the following ways (A-E):

- A. As the authors did not have the original format of the surveys, the answer choices were taken from Powers et al. (2010, p. 11.4) and translated by the authors, as shown below. The “0” choice was created to account for respondents being allowed to “omit a task statement if they thought it did not apply to them or they were unable to make a judgment” (Powers et al., 2010, p. 11.4). Item numbers represent the final randomized version order, and Japanese precedes English in these bilingual surveys as it is the first language of the majority of survey takers.

0=答えられない（私に関係な・判断できない） cannot answer: does not apply to me/unable to judge

1=全くできない not at all

2=非常に難しい with great difficulty

3=ある程度難しい with some difficulty

4=それほど難しくなく with little difficulty

5=簡単 easily

- B. Item 31 was in IIBC (2018), not in Powers et al. (2009), but something similar in Powers et al. (2010, Appendix B). 講義やプレゼンテーションの際、不明瞭な問題点についてその場で説明することができる

- C. Item 37 was in IIBC (2018), not in Powers et al. (2009), but in Powers et al. (2010, Appendix B). 自分が行った研究や調査に関し、結果を報告することができる report on the outcomes of my research or investigation

- D. Item 38 was in IIBC (2018), not in Powers et al. (2009), but in Powers et al. (2010) Appendix B. 自分の専門分野に関係する話題について、分かりやすく伝えることができる communicate in an understandable way about topics related to my field or specialty

- E. Item 39 was in IIBC (2018), not in Powers et al. (2009), but in Powers et al. (2010) Appendix B. 不良品や不十分なサービスに対する苦情を処理することができる deal with complaints about defective product or inadequate service

- F. Item 40 was in Powers et al. (2009, 2010), but not in IIBC (2018). The English was translated into Japanese, but then not included in the survey as not “everyday” item. ビジネス交渉や表敬訪問など、さまざまな場面で経営陣の通訳を務めることができる serve as an interpreter for top management on various occasions such as business negotiations and courtesy calls

A-F are the differences between the lists. However, upon inspection of the ETS and IIBC lists, the authors made the following adjustments to what they believe is more accurate translation, easier to understand, or for continuity of style. The translation changes were verified as “correct” by 2 other native Japanese professionals who said these changes “should be made.”

On the ビジネス場面における英語のスピーキングスキルに関する自己報告 / Self-Report on English Language Speaking Skills for Business, the following changes were made.

- A. In Item 1 the English was changed from “using” to “use.”
メニューを見ながら、喫茶店やレストランで料理を注文することができる using a menu, order food at a café or restaurant
- B. In Item 3 the Japanese was changed to “非公式（インフォーマル）...” as the IIBC (2018) translation uses フォー

マル instead of 公式, see Item 18.

非公式の場面で、自分の趣味や故郷、興味のあることについて話すことができる discuss my hobby, hometown, or interests when introducing myself in informal situations

- C. In Item 21 the Japanese was changed to “非公式（インフォーマル）...”, for the same reason as B above.
非公式の場面で、会話等の通訳をすることができる translate (e.g., conversations) in an informal setting
- D. In Item 23 the Japanese was changed to “自分が...”.
自分やよく知っている仕事関連のプロセスについて、説明することができる describe a familiar work-related process
- E. In Item 27 the Japanese was changed to “述べたり力説したりする”.
議論や会議の場で、自分の意見を述べたり力説することができる state and emphasize my opinion during a discussion or meeting
- F. In Item 29, the Japanese was changed to “公式（フォーマルな）”.
事前に準備をした上で、新規プロジェクトや新製品に関する 20～30 分のプレゼンテーションや、公式スピーチを行うことができる give a prepared 20-30 minute presentation or formal talk (e.g., about a new project or product)
- G. Item 30 was very similar to Item 31, but different. For our study, we used the Japanese from Item 31 and made a proper English translation.
プレゼンテーションやスピーチで議論されている内容について、質問をしたり質問に答えたりすることができる ask or answer questions about an issue being discussed in a presentation or speech
- H. In Item 31 the Japanese was changed to “質問”.
講義やプレゼンテーションの際、不明瞭な問題点についてその場で説明することができる (IIBC, 2018) ask questions to understand unclear or problematic points made in a lecture or presentation (Powers et al., 2010)
- I. In Item 37, the Japanese was changed to “関して”.
自分が行った研究や調査に関し、結果を報告することができる report on the outcomes of my research or investigation

On the “ビジネス場面における英語のライティングスキルに関する自己報告/ Self-Report on English Language Writing Skills for Business, the following changes were made.

- J. In Item 15, the English was English changed to “translate”.
ビジネスレターやマニュアル等の書類を、英語に翻訳することができる transcribe documents (e.g., business letters, manuals) into English
- K. In Item 21, the Japanese was changed to “フォーマル”.
顧客に対する、正式な礼状を書くことができる write a formal letter of thanks to a client
- L. In Item 24, “（または学会で）” was added to the Japanese both for accuracy of translation and as it suited the group of engineers being surveyed.
専門家による会議のプレゼンテーションで使用する文章やスライドを作成することができる prepare text and slides (in English) for a presentation at a professional conference