第三部

Assessing Performance Data in IBC EFL Classes, 2021-2022

Theron E. Fairchild

The purpose of this report was to share two years of data that were collected in EFL classes within the International Business Communications (IBC) course at Kanagawa University. Due to the Covid-19 pandemic, the fall term of 2021 was conducted entirely online, whereas the spring term of 2022 was conducted entirely in a live campus classroom. This situation presented a unique opportunity to evaluate and compare student performances alongside self-reports on attitude and affect. In particular, could the data reveal any signals as to which students might perform better than others in IBC English classes? Could an assessment indicate any specific behaviors or evaluation questions that might predict which students were likely to struggle? Could the data inform the instructor on how to design student tasks and evaluate performances?

1. Introduction

The assessment compared class performance and psychological markers. The purpose was to ascertain whether certain student characteristics or self-reports could provide hints as to which students might likely succeed in the first-year EFL portion of the IBC program. More practically, could an English instructor identify early on which students might struggle in the program? Could the program screen for students who lacked the tools to succeed in IBC in the first place?

Within the ordinary routine of designing course tasks and evaluating students, I accumulated much information and insight about student performance and motivation. As an EFL educator for many years, I routinely observed that students who demonstrated more maturity or emotional intelligence tended to perform better and improve their English ability more readily. This observation suggests something about the students but also about how I design, conduct, and evaluate performance in a

participation-oriented EFL classroom. To put some quantitative assessment to my observations, and to compare a live classroom setting to an online setting, I gathered various types of performance and emotion-related data. In this report, I explain my process and share my findings.

2. Literature Review

Recent research suggests a relationship between psychological factors and EFL learning. In Asia, studies ranging from Iran (Nosratinia & Sarabchian, 2013a, 2013b) to China (Yang, Gao, & Han, 2021; Yu, 2022) pointed to quantifiably significant associations between EFL learning performance and student personality, affective characteristics, and emotional intelligence. Specifically, Dastgoshadeh and Javanmardi (2021) showed emotional intelligence (EI) as a predictor of an EFL learner's willingness to communicate.

In an early framework of emotional intelligence (EI), Salovey and Mayer (1990) identified four components of emotional processing within humans: perceiving, utilizing, understanding, and managing. Building from this framework by incorporating the latest research in psychology and cognitive neuroscience, Drigas and Papoutsi (2018) presented a nine-layer (pyramid) model, which addressed emotions from the base levels of neurological stimuli and encoding to the peak levels of social context and ultimately personal transcendence and emotional unity.

Domain-specific forms of intelligence and cognitive processes, such as academic intelligence, remain important when addressing abilities like planning and problem-solving, including learning problems relevant to language use and acquisition (Kanazawa, 2004). However, other critical components, in academic and business settings, include social skills, emotional adaptation, and empathy (Gendron, 2004). In the context of language acquisition, EI relates to aspects of self-regulation and tolerance towards frustrations and making mistakes. Learning from those frustrations and mistakes and capitalizing on the knowledge gained are critical in the process toward EFL fluency.

Related to emotion, a special section of my report addressed student self-reports on a selection of Japanese aesthetic concepts (taken from Yamakuse, 2012), which included 気 (energy), 融通 (flexibility), and 美 (beauty). Energy 気 and flexibility 融通 are important in tasks involving creative thought, problem-solving, and learning success (Gusroy & Kubra, 2018). Beauty 美 was also included because my suspicion, based on classroom observations over the years, is that students who are image conscious tend to perform poorly in EFL classes. Ackerman (2018) and Lupu and Petrescu (2012) discussed several problems associated with unstable self-image and worry about physical appearances, such as feelings of anxiety, rejection, and difficulty focusing on goals or tasks, including learning tasks.

3. Assessing

To test my assumptions about associations between emotion and EFL learning, I focused on the Fall 2021 term (F2021), which was conducted entirely online, and the Spring 2022 term (S2022), which was conducted entirely in person. The students were first-year members of the IBC program. For purposes of English education, IBC students are grouped by English level, with IBCa considered the highest level group. I evaluated the results from three performance markers that I use commonly for EFL classes. In addition, I asked students to volunteer information about themselves through five self-report psychological instruments. These questionnaire instruments were administered in Japanese to help ensure comprehension. Participants completed them online and were awarded bonus grade-points for completing all five questionnaires. There was no requirement to participate. Students provided permission to use their responses for course evaluation and research.

3.1 Performance Markers

I employed three primary performance evaluations in my EFL classes for IBC terms F2021 and F2022: Performance, Participation, and Production. I also employed two secondary markers, Improvement and Potential. These last two were qualitative assessments built from the three primary performance markers, thus they were redundant and excluded from the following list.

3.1.1 Performance

Performance represents the student's overall graded performance and is identical to the end-of-term grade. This quantitative assessment, scored 0-100, accounts for all points awarded on assignments and tasks completed by the student. For instance, a Performance score of 85 would be reported as a term grade of 85. The purpose of the Performance score is to account for all active and passive English study as well as dedication to learning. Approximately 15-20% of the Performance score is comprised of the Participation score, and 15-20% is comprised of tasks related to the Production score. Approximately 60-70% of the remaining Performance score comes from points awarded on more traditional and passive tasks such as reading, listening, and grammar study. Because the assignments and tasks are intended to be level appropriate for the respective classes, the range of Performance scoring should be roughly equivalent for both IBCa and IBCb (the unadjusted IBCa Performance scoring would be marginally higher due to the 15% inclusion of the Production score).

3.1.2 Participation

Participation is a qualitative assessment, scored 0-100, on factors such as engagement in classroom activities, attention to guided instruction, and on-time submission of assignments. The score is meant to assess the student's effort, not English ability, thus the score is intended to be level appropriate for the respective class (i.e., IBCb Participation is not assessed against IBCa Participation). The Participation score is factored into about 15-20% of the overall Performance score, thus there is some overlap in scoring between the two markers.

3.1.3 Production

Production, scored 0-100, assesses the oral and written output of each student. Some EFL learners are better at speaking than writing, and vice versa. The Production score is a composite of both abilities. A quiet or so-called shy student could still receive a high Production score if that student demonstrated proficiency in writing. Production is an absolute score that is *not* adjusted for class level. Therefore, IBCa students tend to score higher than IBCb students due to their more advanced English ability. Because about 15-20% of this score factors into Performance, I generally

adjust the IBCb Performance tasks in an effort to keep point accumulation and course grading fair between IBCa and IBCb.

3.2 Questionnaire Selection

First-year English students in IBC were presented with five psychological instruments, each one available in a validated Japanese language version. These five instruments were administered online during F2021 and S2022.

3.2.1 TIPI-J: Ten Item Personality Inventory (Japanese)

The standard TIPI is a self-report 10-item measure of the Five-Factor Model of personality. The five dimensions of the model are extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience. Each dimension reports on two items, one scored positively and one negatively. The Five-Factor (Big Five) model remains the most widely used personality inventory globally, and the TIPI has been translated into numerous languages. Oshio, Abe, and Cutrone (2012) evaluated the Japanese version of the TIPI (the TIPI-J) and confirmed its validity and reliability.

3.2.2 PANAS: Positive and Negative Affect Schedule

The 20-item version of the PANAS is a self-report questionnaire on emotional characteristics. Comprised of words that describe different emotions and feelings, the 10 positive and 10 negative affect items range from *attentive* and *inspired* to *irritable* and *distressed*. The PANAS has been used in a variety of clinical and non-clinical settings to assess affect and mood. Kawahito, Otsuka, Kaida, and Nakata (2011) evaluated the Japanese version of the scale and confirmed its validity and reliability.

3.2.3 CERTS-J: Mini-Cambridge Exeter Repetitive Thought Scale

The Mini-CERTS is a self-report questionnaire assessing rumination. This 16-item scale evaluates two dimensions of rumination: (CET) a constructive form of rumination and concrete, experiential thinking, and (AAT) an unconstructive form of rumination and abstract, analytical thinking. Constructive rumination is useful for learning and problem-solving. Unconstructive rumination is a factor when evaluating

conditions such as anxiety, depression, and negative self-assessment. Kambara, Kira, and Ogata (2019) established the validity and reliability of the Japanese version of the CERTS (the CERTS-J).

3.2.4 ATAS: Attitudes Towards Ambiguity Scale

ATAS is a self-report 26-item questionnaire designed to assess tolerance toward ambiguous situations. The scale is an original Japanese design, by Nishimura (2007), and has been translated into English. In subsequent research, Enoki et al. (2018) reconfirmed the Japanese version's validity and reliability. ATAS has been used to assess subclinical depression and anxiety. It is also used to determine learning potentials.

3.2.5 DVQ: Dorsal-Ventral Questionnaire

The DVQ is a self-report 20-item questionnaire intended to evaluate information processing pathways. Cognitive neuroscience has identified multiple systems at work in the brain, which include the dorsal and ventral systems. The dorsal system plays a key role in processing information about spatial properties, and the ventral system plays a key role in processing information about properties of objects. Kawahara (2015) developed the Japanese version of ATAS and established its validity and reliability.

3.3 Questionnaire Collection

The five questionnaires (TIPI-J, PANAS, CERTS-J, ATAS, and DVQ) were administered online for the semesters F2021 and S2022. The students had a three-week window, during each semester, in which to complete the questionnaires. The five instruments were administered using Microsoft Forms, available to the students through their university (Jindai) accounts. I regularly delivered most of my online class assignments using MS Forms, so the students were already familiar with the application. Student identity was automatically verified through the system, so that I could match questionnaire responses to performance markers. Student identity was kept entirely confidential. Participation in the questionnaires was voluntary, and students who completed all five received bonus points applied to their final

Performance score for the class. For F2021 and S2022, 126 of 135 students (93%) completed all five questionnaires, and all 126 provided permission to use their responses for assessment purposes.

3.4 Japanese Aesthetic Concepts

For the term F2021 only, the students received a bonus-credit assignment in which they rated the importance of 10 Japanese aesthetic concepts to themselves. The concepts were taken from Yamakuse (2012), which I turned into a bilingual questionnaire. A six-point Likert scale was used, with a score of 1 representing least important and a score of 6 representing most important. Due to time constraints, this assignment was not offered in S2022. However, the F2021 results indicated some significance in overall student assessment, thus the data were addressed in the current report. The aesthetic concepts serve as a point of possible future study in relation to EFL performance.

3.5 Method

For statistical analysis, I used RStudio for Apple Macintosh, a platform I have used multiple times in the past. In the initial analysis, I constructed a Spearman correlation matrix for all variables collected from F2021 and S2022, to spot patterns and associations of interest. Next, I employed *t*-tests and U-tests on numerous two-sample means and medians, to look for differences in variables between various student groupings.

For correlation analysis, I employed the nonparametric Spearman's Rank-Order Correlation (rho) as the most robust test for the data. Although the continuous data variables passed most assumptions needed for parametric testing, including a Kolmogorov-Smirnov test for normal distribution, there were some concerns about how to address all data collectively. Though all the continuous data were monotonic, some of the variables appeared to show a small amount of heteroscedasticity, that is, the line of best fit lacked linearity. There were also a few outliers in the data, particularly in the variables regarding course scoring (performance markers). Yet these outliers were few and the sample size was large enough that a Kendall-tau test was ruled out. A Spearman's test appeared as the best choice, which could also

accommodate ordinal data in a correlation matrix. The five psychological instruments were scored using ordinal Likert-scales, and I was interested in assessing individual items (dimensions) against student performance markers.

4. Results

4.1 Performance Marker Associations

Shown in Table 1, results from the Spearman correlation testing indicated that there was a significant positive association between the three major and two minor performance markers ($p < .001$). Most noticeably, Performance associated more strongly with Participation, $r_s = .88$, $p < .001$, than with Production, $r_s = .75$, $p < .001$. There was also a significant positive association between Production and Participation. The Spearman test also revealed strong associations with Improvement and Potential. The redundancy of these last two associations demonstrated little need to report on these markers further in this report.

Table 1

Associations of Performance Markers to Themselves

	Participation	Production	Improvement	Potential
Performance	**.88**	.75	.82	.73
Participation		.83	.84	.79
Production			.84	.86
Improvement				.89

Note. Spearman coefficient showed all correlations significant, $r_s(124)$, $p < .001$.

4.2 Instrument Associations

Shown in Table 2, results from the Spearman correlation testing indicated that there was a significant positive association between four of the five questionnaires (instruments) ($p < .001$). The strongest association was between TIPI-J and PANAS, $r_s = .68$, $p < .001$. All associations to DVQ were not significant, and *t*-tests confirmed that DVQ scores were not significant between groups, thus the instrument was omitted from further reporting.

Table 2

Associations of Instruments

	PANAS	CERTS-J	ATAS	DVQ
TIPI-J	**.68**	.55	.37	.12[ns]
PANAS		.53	.45	-.05[ns]
CERTS-J			.49	-.14[ns]
ATAS				-.14[ns]

Note. Spearman coefficient showed all correlations significant between TIPI-J, PANAS, CERTS-J, ansd ATAS, r (124), $p < .001$.
[ns]All associations to DVQ were weak and not significant.

4.3 Performance Markers and Instruments

Table 3 shows the associations between four instruments and the three performance markers. The strongest association was significant and positive, between Participation and PANAS, $r_s = .45$, $p < .001$. Two associations to ATAS were not significant, thus ATAS was omitted from reporting in Table 4.

Table 3

Instruments and Performance Markers

	Performance	Participation	Production
TIPI-J	.30	.37	.31
PANAS	.36	**.45**	.33
CERTS-J	.29	.33	.25
ATAS [r_s (p)]	.13 (.15)[ns]	.22 (.01)	.16 (.08)[ns]

Note. Spearman coefficient showed all correlations significant for TIPI-J, PANAS, and CERTS-J, r_s (124), $p < .001$.
[ns]ATAS p-values listed separately [r_s (p)].

Table 4 shows the associations between combinations of two or three instruments against the three performance markers. All associations were positive and significant at $p < .001$. The strongest associations were between Participation and the instruments.

Table 4

Three Instruments and Performance Markers

	Performance	Participation	Production
TIPI-J & PANAS	.36	.44	.34
TIPI-J & CERTS-J	.35	.40	.33
PANAS & CERTS-J	.38	.45	.35
TIPI-J & PANAS & CERTS-J	.38	.46	.36

Note. Spearman coefficient showed all correlations significant, $r_s(124)$, $p < .001$.

4.4 Group Differences

In the current section, performance markers and instruments were compared based upon different groupings. A Kolmogorov-Smirnov Test of Normality for each variable confirmed their normal distributions. A Levene's test for homogeneity of variance confirmed equal variances for most comparisons. As needed, either an equal or unequal variances *t*-test was employed on each comparison. Each test was two-tailed.

4.4.1 Female and Male Compared

Table 5 shows the comparison of mean performance markers between all female students and male students for the terms F2021 and S2022. The descriptive statistics indicated little difference between female and male students, and results from *t*-tests confirmed no significant differences at $p < .05$. The descriptive statistics for the instruments (not shown) also indicated little difference between female and male students, and the results from *t*-tests confirmed no significant differences at $p < .05$.

Table 5

Female and Male Performance Markers Compared

	Performance	Participation	Production
Female (N = 62)	84.0 (15.7)	75.6 (17.9)	73.4 (14.3)
Male (N = 64)	82.9 (15.6)	75.1 (18.8)	71.7 (14.8)
Total (N = 126)	83.4 (15.6)	75.3 (18.3)	72.5 (14.5)

Note. No differences were significant at $p < .05$.

4.4.2 F2021 and S2022 Compared

Table 6 shows performance marker scores between the F2021 and S2022 classes. *T*-tests indicated that only the mean difference for Participation was significant. The 57 students from S2022 (M = 80.3, SD = 15.2) compared to the 69 students from F2021 (M = 71.2, SD = 19.7) demonstrated significantly higher Participation scores, *t* = 2.9, *p* = .005. Although S2022 had higher mean Performance scores and lower mean Production scores than F2021, the differences were not significant at *p* < .05. *T*-tests (not shown) also indicated no significant differences in means between the class semesters and the instrument scores at *p* < .05.

Table 6

Fall 2021 and Spring 2022 Performance Markers Compared

	Performance	Participation*	Production
F2021 (N = 69)	81.2 (16.2)	**71.2 (19.7)**	74.5 (14.1)
S2022 (N = 57)	86.1 (14.6)	**80.3 (15.2)**	71.7 (14.8)
Total (N = 126)	83.4 (15.6)	75.3 (18.3)	72.5 (14.5)

Note. *Participation was significant, t(124) = 2.9, *p* = .005.
Performance and Production were not significant at *p* < .05.

4.4.3 IBCa and IBCb Compared

Table 7 shows performance marker scores between IBCa and IBCb for terms F2021 and S2022. *T*-tests indicated that only the mean difference for Production was significant. The 60 students from IBCa (M = 79.8, SD = 4.8) compared to the 66 students from IBCb (M = 68.5, SD = 13.8) demonstrated significantly higher Production scores, *t* = 6.2, *p* < .001. The differences in means for Performance and Participation were not significant at *p* < .05. *T*-tests (not shown) also indicated no significant differences in means between the IBC groups and the instrument scores at *p* < .05.

Table 7

IBCa and IBCb Performance Markers Compared

	Performance M (SD)	Participation M (SD)	Production* M (SD)
IBCa (N = 60)	82.0 (18.2)	74.8 (20.4)	**79.8 (4.8)**
IBCb (N = 66)	84.7 (12.8)	75.8 (16.2)	**68.5 (13.8)**
Total (N = 126)	83.4 (15.6)	75.3 (18.3)	72.5 (14.5)

Note. *Production was significant, t(124) = 6.2, *p* < .001.
Performance and Production were not significant at *p* < .05.

4.4.4 Top Quartile and Bottom Quartiles Compared

Based on a composite of Performance, Participation, and Production for the F2021 and S2022 terms, the top 32 student performers (15 female, 17 male) were compared to the bottom 32 performers (14 female, 18 male). From the Top Quartile (TQ), 29 completed a TOEFL, and from the Bottom Quartile (BQ), 22 competed a TOEFL. The means were recorded in Table 8.

Table 8

Top and Bottom Quartiles: Performance Markers & TOEFL

	Top Quartile (N = 32) M (SD)	Bottom Quartile (N = 32) M (SD)
Performance	97.2 (3.4)	63.5 (14.9)
Participation	95.4 (3.7)	51.1 (11.5)
Production	88.5 (4.5)	55.8 (10.6)
	(N = 29) M (SD)	(N = 22) M (SD)
TOEFL	487 (21)	460 (26)

Note. 29 of 32 in Top Q and 22 of 32 in Bottom Q received TOEFL score.

Table 9 shows the comparisons between TQ and BQ on the instrument scores. The *t*-tests indicated significant differences on three of the four instruments. Most notably, TQ's reporting on PANAS was more than one standard deviation higher than BQ's reporting, *t* = 4.9, *p* < .001. Across all four instruments, TQ scored consistently higher than BQ on all positive dimensions and lower on all negative dimensions.

Table 9

Top and Bottom Quartiles: Instrument Scores

	Top Quartile (N = 32) *M* (*SD*)	Bottom Quartile (N = 32) *M* (*SD*)	*t*-test (df = 62)
TIPI-J total	46.8 (6.7)	41.1 (6.8)	t = 3.4, p = .001
positive dimensions	24.7 (3.4)	21.7 (4.4)	
negative dimensions	17.8 (4.7)	20.7 (4.0)	
PANAS total	**85.3 (10.2)**	**74.1 (7.7)**	***t* = 4.9, *p* < .001**
positive dimensions	40.6 (6.3)	32.2 (7.3)	
negative dimensions	25.3 (7.2)	28.1 (7.6)	
CERTS-J total	32.8 (4.5)	29.2 (4.3)	t = 3.2, p = .002
positive dimensions	17.3 (3.0)	15.8 (3.3)	
negative dimensions	14.5 (3.3)	16.7 (4.2)	
ATAS total	96.9 (10.8)	92.8 (10.7)	t = 1.5, p = .13[ns]
CET (positive)	57.5 (6.6)	55.3 (11.4)	
AAT (negative)	51.6 (8.6)	53.5 (9.2)	

Note. [ns]ATAS not significant at $p < .05$.

4.4.5 TOEFL and Non-TOEFL Compared

Of the 126 students who completed the five psychological instruments, 102 received a score on at least one TOEFL attempt, and 24 had No TOEFL attempt or score. Table 10 shows the means and results of three *t*-tests. The students with No TOEFL score also recorded lower scores on all three performance markers, although one of those was not significant at $p < .05$. As for the means between TOEFL and the instruments (not shown), the scores differed little and none were significant at $p < .05$.

Table 10

TOEFL and Performance Markers Compared

Total (N = 126)	Performance *M* (*SD*)	Participation *M* (*SD*)	Production *M* (*SD*)
TOEFL (N = 102)	85.5 (14.3)	76.5 (17.9)	74.1 (13.8)
No TOEFL (N = 24)	74.9 (18.4)	70.1 (19.5)	65.9 (15.9)
	t = 3.1, p = .003	t = 1.6, p = .12[ns]	t = 2.6, p = .01

Note. Two-sample *T*-test, using two tails, compared 126 student respondents, of which 102 received at least one TOEFL score and 24 never took the test.
[ns]Participation not significant at $p < .05$.

All data points for the 102 TOEFL takers were also processed in a Spearman correlation matrix. As shown in Table 11, the only notable associations were between student TOEFL scores and the three performance markers. The strongest association was between TOEFL scores and Production, r_s = .43, p < .001.

Table 11

TOEFL and Performance Markers: Associations

	Performance	Participation	Production
TOEFL (N = 102)	.32	.33	**.43**

Note. All Spearman correlations significant, $r_s(100)$, $p \leq .001$.

4.5 Individual Items and Performance Markers

Table 12 lists six individual instrument items (dimensions) that showed the strongest associations to the three primary performance markers. Table 13 shows results from the Spearman rho matrix, grouped between the six listed items and the three performance markers. The strongest associations concerned TIPI-J item 8, negative conscientiousness (disorganized, careless).

Table 12

Six Individual Questionnaire Items

- TIPI-J item 1: 活発で,外向的だと思う (extraverted, enthusiastic) [positive score] [1]
- TIPI-J item 8: だらしなく,うっかりしていると思う (disorganized, careless) [neg. score]
- PANAS item 2: 活気のある (active, lively) [positive score] [2]
- PANAS item 12: やる気がわいた (inspired, motivated) [positive score]
- CERTS-J item 6: CET (constructive rumination) [positive score] [3]
  私の思考は、新しいことや興味のある方へ向きやすい 。
  My thoughts move in new and interesting directions.
- CERTS-J item 9: AAT (unconstructive rumination) [negative score]
  私は、私にとって恐ろしいことがこれ以上起きないようにしなければいけな いというプレッシャー
  を感じる。I feel under pressure to stop my worst fears from happening.

[1] TIPI-J: Oshio, Abe, & Cutrone (2012)
[2] PANAS: Kawahito et al. (2011)
[3] CERTS-J: Kambara, Kira, & Ogata (2018)

Table 13

Four Dimensions and Performance Markers

	Performance	Participation	Production
TIPI-J 1: pos. extraversion	.26	.37	.34
TIPI-J 8: neg. conscientiousness	**-.39**	**-.42**	**-.34**
PANAS 2: pos. active	.34	.38	.27
PANAS 12: pos. motivated	.32	.39	.31
CERTS-J: 6 CET pos.	.31	.32	.25
CERTS-J 9: AAT neg.	-.29	-.34	-.21[1]

Note. Spearman coefficient showed all correlations but one, $r_s(124)$, $p < .01$.
[1]$p = .02$

Table 14 compares the six individual inventory items, between the Top Quartile (TQ) and Bottom Quartile (BQ) students from F2021 and S2022. Using ordinal data from a small sample size necessitated the use of a nonparametric analysis of the medians rather than the means. A two-tailed Mann-Whitney U Test was used to assess whether the medians between TQ and BQ on each of the six inventory items differed. The z-score was reported. The paired tests indicated that TQ scored significantly higher on the positively scored items and significantly lower on the negatively scored items.

Table 14

Top and Bottom Quartiles: Six Inventory Items

	Top Quartile (N = 32) Mdn	Bottom Quartile (N = 32) Mdn	Mann-Whitney Test z (p)
TIPI-J 1	6	4.5	3.4 (< .001)
TIPI-J 8 (neg)	4.5	6	-3.8 (< .001)
PANAS 2	5	3.5	3.9 (< .001)
PANAS 12	5	3.5	3.4 (< .001)
CERTS-J 6: CET	4	3	3.1 (.002)
CERTS-J 9: AAT (neg)	1	2	-2.6 (.008)

Note. TIPI-J ordinal data (medians) from a 7-point Likert scale (7 = agree strongly).
PANAS ordinal data (medians) from a 6-point Likert scale (6 = extremely).
CERTS-J ordinal data (medians) from a 4-point Likert scale (4 = always).

4.6 Japanese Concepts and Performance Markers

Only the F2021 students, 68 of 76, completed this assignment. To statistically assess differences between the top and bottom performing students, the F2021 class data were divided into equal halves, of the top 34 performers versus the bottom 34 performers. This division was redrawn for a total of five passes, once for each of five performance markers: Performance, Participation, Production, Composite PPP, and TOEFL.

A nonparametric Mann-Whitney U Test was used to assess whether the medians between the top and bottom performers on each of the five divisions for performance markers differed. The z-score was reported. Though the median scores in the descriptive statistics noticeably differed, the small sample size allowed for only three significant differences. The two biggest differences occurred when students were divided along the marker of Performance. As shown in Table 15, a Mann-Whitney test indicated that the top 34 students (Mdn = 5) scored significantly higher than the bottom 34 students (Mdn = 4) on 気 energy, z = 2.3, *p* = .02. The second most significant difference was that the bottom 34 students (Mdn = 5) scored higher than the top 34 students (Mdn = 4) on 美 beauty, z = 2.1, *p* = .03. The same positive-negative pattern followed for the remaining variables, including TOEFL, but the sample size was too small to gain significance.

Table 15

F2021 Top and Bottom Performers: Concepts and Performances

Performance score	Top Perf (N = 34) Mdn	Bottom Perf (N = 34) Mdn	Mann-Whitney Test z (*p*)
気 (ki) energy	5	4	2.3 (.02)
美 (bi) beauty	4	5	-2.1 (.03)
融通 (yuzu) flexibility	4.5	4	1.1 (.28)[ns]

Note. Ordinal data (medians) from a 6-point Likert scale (6 = most important).

The 10 Japanese aesthetic concepts and five performance markers were next entered into a Spearman correlation matrix to assess potential associations. The result indicated several significant weak associations. Table 16 shows the most significant associations, between 気 energy and three of the performance markers.

Table 16

Energy (気) and Performance Marker Associations

	Performance r_s (p)	Participation r_s (p)	Production r_s (p)	Composite r_s (p)
気 (ki) energy	.31 (.01)	.28 (.02)	.23 (.06)[ns]	.28 (.02)

Note. N = 68. All Spearman correlations, r_s (66).
[ns]Production not significant sat $p < .05$.
TOEFL (not shown) not significant at $p < .05$.

5. Discussion

The purpose behind an assessment of class performance and psychological markers was to ascertain whether certain characteristics could provide hints as to which students might likely succeed in the first-year EFL portion of the IBC program. Analysis of the data indicated that a student's positive and negative affective characteristics, or emotional intelligence (EI), played a role in the student's success, at least in the participatory type of EFL classroom under discussion. Since a student's Participation score was largely a qualitative assessment by the instructor, the score was also partly an assessment of the student's EI in relation to English learning. However, given the quantitative nature of the Performance score and the mixed nature of the Production score, positive and negative affect also appeared relevant to the student's overall focus on learning and expressing in English.

The most compelling evidence for considering the role of affect in EFL success came from comparing the Top Quartile (TQ) of student performers, across the semesters F2021 and S2022, to the Bottom Quartile (BQ). In every applicable category, TQ tended toward more positive emotional attributes and fewer negative attributes. Students who self-reported more favorably on attributes related to motivation, enthusiasm, and self-discipline performed significantly better on English expressive production and overall grading performance. This result could partly be the result of the instructor's class design and evaluation methods, but overall performance and especially higher expressive production also associated significantly with higher TOEFL scores. Participation, which associated significantly with positive affect, reflected student engagement with other EFL learners and with the instructor. The purpose of such participation was to help improve a student's expressive production and overall ability.

Other comparisons in the data were also telling. There were no significant differences between female and male students on any of the performance markers or psychological measures. The only significant difference between IBCa and IBCb concerned the Production score. This was expected since the score is measured on an absolute scale, thus IBCa students would naturally have higher scores than IBCb students. In short, the scoring and evaluation of student work appeared consistent across sex and English level.

However, in comparing the F2021 group to the F2022 group, the lack of significant differences was unexpected. The 2021 data came from classes conducted entirely online, during the fall semester. The 2022 data came from classes conducted entirely in a live classroom, during the spring semester. Although the psychological instruments were delivered in an identical way during both semesters, the class assignments and their delivery varied between terms. In addition, the mood, motivation, and mindset of first-year students typically go through adjustments between the initial spring semester and the subsequent fall semester. Yet on the performance markers and psychological measurements alike, the only significant difference was that F2021 scored lower than S2022 on Participation. As demonstrated in online KU classes during the 2020 pandemic, apprehension by some students to participate online, combined with my own inexperience in evaluating online participation, likely contributed to F2021 scoring significantly lower than F2022 on the Participation marker. However, all other performance markers and psychological measures were statistically identical between F2021 and S2022.

The lack of significant differences between F2021 and S2022 points to several possible conclusions. One is that the instructor's class designs and evaluation methods were mostly consistent for both online and live classrooms, which is important since some online education could become standard practice at Kanagawa University. A second conclusion is that the psychological instruments, particularly TIPI-J, PANAS, and CERTS-J, were valid measurements of stable affective characteristics, which is important if planning to use these instruments in future assessments. A third conclusion suggests that the IBC students themselves, particularly the most successful students, were focused, determined, and psychologically stable enough to stay the course in IBC English regardless of the changes around them, which is an indication

of emotional intelligence.

As a final observation, the positive relationships of 気 energy and 融通 flexibility to performance, and the negative relationship of 美 beauty to performance, suggested a possible future topic of research. The concepts of 気 energy and 融通 flexibility have obvious connections to current research on motivation and flexibility in learning. Yet there is also much research regarding overemphasis on physical appearances and its association with conditions such as poor self-image, eating disorders, and depression (Ackerman, 2018; Lupu & Petrescu, 2012). For a future review of literature, self-image in relation to language learning would be a useful point of inquiry. Young people in general are often worried, sometimes excessively, about their appearances and often attempt to build their self-images by emulating established or packaged images of attractiveness and success. In my years of observational experience within classrooms, I have noticed that students who seem preoccupied with appearances and image building tend to struggle in EFL learning.

Finally, one question this report asked was whether any particular markers, instruments, or individual questions might help predict student success in IBC English or IBC overall. Likewise, how would an instructor employ such information to improve the quality of English education within IBC? A program-wide discussion would likely address these questions better. As for delivery of my own EFL classes, the data and analyses presented in this report provided some insight on how to improve my course designs and evaluations, for both live and online class environments. The report also suggested the need to better incorporate a standard measure like TOEFL into the design. Ultimately, there remain several potential research avenues to explore.

References

Ackerman, C. E. (18 November 2018). What is self-image in psychology? How do we improve it? Retrieved from https://positivepsychology.com

Dastgoshadeh, A., & Javanmardi, P. (2021). Emotional intelligence as a predictor of EFL learners' willingness to communicate. Mextesol Journal, 45(3).

Retrieved from https://www.mextesol.net

Drigas, A.S., & Papoutsi, C. (2018). A new layered model on emotional intelligence. Behavioral Science, 8(5). https://doi.org/10.3390/bs8050045

Enoki, H., Koda, M., Saito, S., Nishimura, S., & Kondo, T. (2018). Attitudes towards ambiguity in Japanese healthy volunteers. Current Psychology 37(4): 913–923. https://doi.org/10.1007/s12144-017-9569-9

Gendron, B. (2004). Why emotional capital matters in education and in labour? Toward an optimal exploitation of human capital and knowledge management. Pantheon-Sorbonne University: Paris, France. https://hal.archives-ouvertes.fr/hal-00201223

Gusroy, E, & Kubra, H. (2018). Is it possible to enhance the creative thinking skills of EFL learners through training? Advances in Language and Literary Studies, 9(6), 172–182. http://dx.doi.org/10.7575/aiac.alls.v.9n.6p.172

Kambara, K., Kira, Y., & Ogata, A. (2019). Development of a Japanese Version of the Mini-Cambridge Exeter Repetitive Thought Scale. SAGE Open, 9(2), 1–8. https://doi.org/10.1177/2158244019856722

Kanazawa, S. (2004). General intelligence as a domain-specific adaptation. Psychological Review, 111, 512–523. https://doi.org/10.1037/0033-295X.111.2.512

Kawahara, M. (2015). Development of the Japanese Dorsal-Ventral Questionnaire. Japanese Journal of Mental Imagery, 13, 41–53.

Kawahito, J., Otsuka, Y., Kaida, K., & Nakata, A. (2011). Reliability and validity of the Japanese Version of 20-item Positive and Negative Affect Schedule (PANAS). Hiroshima University Journal of Psychological Research, 11, 225–240.

Lupu, E., & Petrescu, A. L. (2012). A study regarding teenagers' self-image and the importance of physical activities in its formation. Procedia—Social and Behavioral Sciences, 33, 870–874. https://doi.org/10.1016/j.sbspro.2012.01.246

Nishimura, S. (2007). Development of the Japanese Dorsal-Ventral Questionnaire. Japanese Journal of Personality, 15(2), 183–194.

Nosratinia, M., & Sarabchian, E. (2013a). Making EFL students aware of the underlying psychological traits: Bounds between personality factors and emotional intelligence. Global Journal of Foreign Language Teaching, 1, 46–54.

Nosratinia, M., & Sarabchian, E. (2013b). Predicting EFL learners' emotional intelligence and critical thinking ability through Big-Five Personality Traits: A study on psychological characteristics of EFL learners. International Journal of Scientific & Engineering Research, 1(9), 501–515.

Oshio, A., Abe, S., & Cutrone, P. (2012). Development, reliability, and validity of the Japanese Version of Ten Item Personality Inventory (TIPI-J). Japanese Journal of Personality, 21, 40–52. http://dx.doi.org/10.4992/jjpsy.83.91

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. Imagination, Cognition, and Personality, 9(3), 185–211. https://doi.org/10.2190/DUGG-P24E-52WK-6CDG

Yamakuse, Y. (2012). Soul of Japan: The visible essence. IBC Publishing, Tokyo, Japan.

Yang, Y., Gao, Z., & Han, Y. (2021). Exploring Chinese EFL learners' achievement emotions and their antecedents in an online English learning environment. Frontiers in Psychology. https://doi.org/10.3389/fpsyg.2021.722622

Yu, Q. (2022). Review of foreign language learners' emotions. Frontiers in Psychology. https://doi.org/10.3389/fpsyg.2021.827104