Education Research and Perspectives
An international journal

Volume 52, 2025, Pages 104-131

Using Free Association as an Implicit Assessment to Measure Intercultural Competence in Virtual Exchange

Curtis Chu[1†]

Mariko Takahashi[2]

1. Setsunan University, Japan

2. Kwansei Gakuen University, Japan

While explicit assessments like surveys and written reflections are commonly used to measure intercultural competence (IC) in virtual exchange (VE), the limitations of the assessments, such as cognitive loading, social desirability bias, and Dunning-Kruger effect, are rarely addressed. Implicit assessments have the potential to reduce the influence of such factors, but no relevant studies were found in the VE field. Hence, the goal of this paper is to pioneer an implicit assessment approach based on free association to explore changes in the internal outcomes component of IC in VE. 28 undergraduate students in Japan participated in a five-week asynchronous VE with Taiwanese partners. Free association was administered pre-VE and before the start of the last session of VE, which involved a near one-month break in between. Five single-word prompts were provided separately, and participants wrote down as many words as they could associate in one minute per prompt. The prompts aligned with the contents of the VE: tourism, hometown, sustainability, Taiwan, and Taiwanese. Including presumptuous words, six themes were identified: concepts, actions, feelings, places, people, and objects. In the post-VE free association, the total number of words for three prompts increased by over 15%, and the types of words for three prompts increased by over 20%. Additionally, fewer words related to content and more words related to culture were associated, thus implying expanded cultural knowledge and awareness. It is argued that intuitive data could be gathered to observe IC in VE using free association, and the limitations of explicit assessments could be mitigated to a certain extent.

†Address for correspondence: Setsunan University, 17-8 Ikeda-Nakamachi, Neyagawa City, Osaka Prefecture, 572-8508. Email: curtis.chu@setsunan.ac.jp

Introduction

Assessments for measuring intercultural competence (IC) of students and their learning outcomes from participating in virtual exchange (VE) are commonly conducted through qualitative methods such as reflective writings, interviews, and observations, and also quantitative methods such as standardized tests, inventories, and scales. While these forms of assessment offer insights into the impact of VE programs in their respective ways, all of them share a common form of harvesting data from students. These forms of explicit metacognitive assessments ask respondents to reflect on their thinking, which could be influenced by various factors such as social expectations, linguistic competence, and how self-reflective one is. As a result of monitoring and controlling one's own thinking processes, or generally referred to as thinking about one's own thinking (Flavell, 1976), explicit assessments extract data from deliberative responses.

In a recent global survey on the practice of assessing IC in various contexts including VE, over 200 respondents reported the usage of various qualitative assessments and 36 different kinds of qualitative assessments, where over 80% of the respondents reported using mixed methods of assessment (Chu et al., 2024a). While explicit assessments are the dominant type of assessment in the field of assessing IC and VE, it is also worth noting that Zotzmann (2015) argued that intercultural learning is fluid and unpredictable, therefore it becomes contradictory when the learning is reduced to fixed sets of competencies for assessment.

Implicit assessments, on the other hand, are very different in nature compared to explicit assessments, and the usage of implicit assessments is rarely discussed in the field of VE to measure IC. It could be a promising way to gain insights into the change in perception to complement explicit assessments. Bazgan and Norel (2013) have argued that implicit assessments could be an effective method for assessing IC, as this type of assessment can bypass the conscious mind of the respondents. They will not be aware of the measured construct or be able

to deliberately control their response. Implicit assessments gather data under specific conditions, including the absence of specific goals, substantial awareness, cognitive resources, and time (De Houwer & Moors, 2007; De Houwer et al., 2009). In other words, if an assessment prompts a respondent to respond as fast and casually as possible, without a thorough understanding of what and how the assessment works, it is believed that an intuitive response from the respondent's automatic processing could be gathered.

One might argue that surveys and questionnaires often ask respondents to intuitively choose the best response, but cognitive loading is much higher when one attempts to intentionally respond to a question. Based on the dual-process model of intention attribution proposed by Rosset (2008), when humans process information, we either use a fast and automatic process that is intuitive or a slower and analytical process that involves higher-level cognitive processing. For example, when a respondent begins reading and attempting to respond to a question, it requires cognitive resources for comprehending and decoding a question; constructing a response that conforms to linguistic skills; and even considering the potential consequences of the response. These factors demand cognitive resources and activate the slower and analytical process of information processing, which takes a respondent farther away from responding intuitively.

Implicit assessments could gather intuitive responses, but this has been a neglected area to date to assess the impact of VE in general and IC in particular. Hence, this study aims to examine whether using free association can indicate changes in perceptions related to the internal outcomes component of IC from Deardorff's Process Model of Intercultural Competence (2006) after experiencing VE. It also aims to establish a protocol for using free association to gather and analyze meaningful data.

Literature Review

Limitations of explicit assessments in measuring IC

With a huge variety of explicit qualitative and quantitative assessment methods available for measuring IC, implicit tools that probe into the unconscious are rare. A weakness of explicit assessments is that they might not be capturing participants' unconscious thoughts, as it is a complicated process for many reasons: the process involves assessing different types of knowledge skills; the thinking process could not be directly observed; and existing assessment tools tend to focus on particular aspects that might not align with the context of an educational intervention (Lai, 2011). Another common factor affecting explicit assessments is social desirability bias, where one might consciously respond to questions by conforming to social norms and/or reporting desirable attributes instead of honest thoughts (Nederhof, 1985). Additionally, the Dunning-Kruger effect could be another factor, which is defined as one might overestimate one's knowledge or abilities (Kruger & Dunning, 1999), subsequently affecting their response to explicit assessments. As mentioned in a research report by Griffith and colleagues (2016), the reliance on self-report among explicit assessments is problematic when assessing IC.

Five potential limitations can be identified in explicit assessments. The first limitation is that explicit assessments of IC do not only measure IC. They also measure other types of knowledge skills, such as linguistic competence, reading comprehension, social interaction skills, self-reflection, and so on (Sercu, 2010). In the case of tests and scales, it is required to be able to comprehend and decode the items to select an appropriate response. In the case of open-ended questions or reflective essays, one must have the necessary language skills to write meaningful responses that conform to linguistic rules. In both cases, one must be able to reflect on one's experience of interacting with an international partner during a VE program and elicit scenarios to serve as evidence to support their response to assessment questions. As Fantini (2007) explained, IC is "a complex of abilities needed to perform effectively and appropriately

when interacting with others who are linguistically and culturally different from oneself" (p.12).

Measuring such complex abilities would require situations where the demonstration of the abilities could be directly observed, which is the second limitation of explicit assessments. An interesting point raised by Sercu (2004) was that when a learner cannot solve an intercultural problem, it is not necessarily because of the learner's lack of sufficient IC, but rather not all learners can explicitly express their ideas in a similar way. This means that there is a possibility for a learner to be wrongfully accused of having low IC from the results of an explicit IC assessment because of how the score could be affected by how well one could express oneself.

As the third limitation of explicit assessments, it is also important to acknowledge how established explicit assessments might not align with the context of a VE project or the target components of IC. Established quantitative assessments such as inventories or scales are fixed sets of items that cannot be customized or edited for the sake of maintaining validity and reliability (Juniper, 2009). This means that not all items in an assessment would fit the contents of a VE project, nor would they be suitable to measure targeted changes in IC. Additionally, established assessments were developed from data gathered from mainstream populations (Steward et al., 2012), which is unlikely for them to take into consideration the context of VE.

While qualitative explicit assessments could adapt to various contexts without jeopardizing the validity, they still mainly rely on self-reported responses, and this is the fourth potential limitation of explicit assessments. Factors such as social desirability bias and the Dunning-Kruger effect remain significant challenges in obtaining honest and/or unconscious thoughts of participants' perception. Social desirability bias, according to Nederhof (1985), is a result of self-deception and other-deception. Self-deception refers to when one believes something to be true about oneself even if it is inaccurate, whereas others-deception refers to when one deliberately presents inaccurate information in consideration of how it would affect being evaluated. Common

behaviors include shaping responses that might favor the expectations of an instructor, and responses that would make oneself look good to peers. There lies the possibility of intentionally lying or simply believing that it is true. Thus, social desirability bias could likely affect how students would deliberately respond to explicit assessments.

Finally, as the fifth limitation of explicit assessments, students might underestimate their IC before engaging in a VE program, or overestimate their IC afterwards. This is known as the Dunning-Kruger effect (Kruger & Dunning, 1999). Take for instance students who have never or rarely engaged in intercultural communications before, it might be expected that they would underestimate themselves or feel a lack of confidence before a VE program begins. During the program, instructors ask their students to be patient and have empathy, which is an important aspect of IC, so that students can collaboratively complete tasks with their international partners. After the end of the program, students celebrate a successful VE experience that might have boosted their confidence and respond to explicit assessments while overestimating themselves.

Given the limitations above, perhaps it is worth reflecting on how much could be learned from studies that did not address the limitations of explicit assessments. Positive findings among VE studies bloomed soon after Byram published his Model of Intercultural Communicative Competence in 1997, such as studies on asynchronous VE (Chu et al., 2024b; O'Dowd, 2003; Vinagre, 2010), synchronous VE (Kilimci, 2010; Siergiejczyk, 2020), and large-scale reports (European Commission, 2024; Stevens Initiative, 2023), and they all drew conclusions from explicit assessments. To measure internal outcomes, an important aspect of IC based on Deardorff's Process Model of Intercultural Competence (2006), which refers to the cognitive and affective changes within an individual during the development of IC, explicit assessments that solely rely on self-reported responses might not be sufficient.

Hence, to refrain from possibilities of under- or overestimation of oneself; effects of social desirability bias; using out-of-context assessments; or demanding high cognitive resources from participants, implicit assessments that probe into the unconscious could serve as a

means to gather intuitive data that bypasses the conscious mind. Gathering intuitive raw data that is generated through minimal metacognitive processes and customized to fit the context of VE would be a different approach to measure change in IC and examine the impact of VE interventions.

Implicit assessments

There are well-established implicit assessments with decades of research, for example, the Go/No-go Association Task (GNAT), the Affect Misattribution Procedure (AMP), and the Implicit Association Test (IAT). Free association, on the other hand, has been used in various fields such as memory recall (Galton, 1880), psychoanalysis (Kent & Rosanoff, 1910), and qualitative research (Hollway, 2000). It is also a method that probes into the unconscious, and it gained awareness when Sigmund Freud (1901) further developed it into a method for psychotherapy. Free association is considered as an implicit assessment that does not trigger deliberate recall of memory or other explicit processes (Schnabel & Asendorpf, 2013). To the best of the authors' knowledge, implicit assessments have not been used in the context of VE.

GNAT is a method for measuring associations between target categories and attribute categories, and the associations are measured by accuracy and response time (Nosek & Banaji, 2001). It is generally conducted on a computer where respondents are asked to press the spacebar if words or images that appear in the center of the screen match the specified category /attribution and do nothing if they do not match the category/ attribution. For example, in an experiment testing race and gender attitudes, black, white, and Asian faces of smiling basketball players were shown, and respondents were asked to categorize them into white, black, male, female, good, and/or bad (Nosek & Banaji, 2001). Results of the experiment found that female respondents responded faster and more accurately when categorizing female faces (compared to male faces), hence interpreted as females having strong positive associations with the category of female compared to male.

Another implicit assessment, the AMP, also taps into respondents' attitudes and biases while they are not consciously aware of it. Developed by Payne and colleagues in 2005, the AMP measures how respondents misattribute their feelings. Respondents are first flashed (shown for a brief moment) with a real-life image (prime stimulus), then flashed a neutral pictograph of a Chinese character, (target stimulus) a neutral image only to those without Chinese language knowledge), followed by a flash of a white masking image, and finally asked to judge whether the Chinese pictograph felt pleasant or unpleasant without thinking about the real-life image. The idea is that people will judge the target stimulus consistently with how they feel with the prime stimulus, which could be interesting if applied to pre- and post-VE to measure perception change.

As for the IAT, it was developed earlier than both the AMP and the GNAT. In 1998, Greenwald and colleagues published the IAT. It is a method that measures the strength of association between concepts and attributes. Like the AMP and GNAT, IAT also operates on a computer using the keyboard, and the idea is that people make responses easier when items that are closely related share the same response key (Project Implicit, 2011). Take for example the Skin-tone IAT task on the Project Implicit website, respondents are shown monochrome photos of individuals and also words that belong to either good or bad categories. Respondents are asked to classify them by pressing the keys that match their category (light-skinned, dark-skinned, good, or bad). The IAT could measure underlying preferences from intuitive responses, thus tapping into the unconscious mind of respondents to uncover potential biases.

Regarding free association with words, great thinkers from Plato to Freud recognized its potential, and it is assumed not only would it allow the reconstruction of subjective meaning, but also reveal perception and attitudes (Szalay & Deese, 1978). Free association is applied in various contexts, such as in psychopathology where a study significantly contributed to the understanding of how individuals with and without mental illness associate between ideas (Kent & Rosanoff, 1910); in intercultural understanding where it was used to study cultural meanings (Szalay & Brent, 1967); and in linguistics where 72,000 word pairs of free association data were gathered from over one million entries (Nelson

et al., 2004). Additionally, free association was used in marketing research to explore median brand equity (McDowell, 2004); in psychoanalysis and psychotherapy (Archard & O'Reilly, 2023); and also as a qualitative research methodology (Hollway & Jefferson, 2008).

In relation to the context of VE or measuring IC, no studies were found using free association to measure the impact of VE or IC. There are studies using free association to evaluate education interventions, particularly in the field of moral education. Kamizono (2009) explained that free association can gather responses in words related to knowledge, feelings, and cultural backgrounds, both conscious and/or unconscious recalls, towards specific prompt words. Students' changes in perception after interventions of moral education lessons were explored (Kamizono & Moringa, 2012), and association maps were created from data gathered with free association for assessing the effectiveness of moral education classes (Araki, 2018).

With the wide applications of implicit assessments and free association, to the best of the authors' knowledge, there is no existing literature in the field of VE and IC using implicit assessments. Hence, this study could be the first to explore the applicability of examining changes in the internal outcomes component of IC from the case of a VE program with an implicit assessment of free association.

Methodology

Data in this study were gathered from 20 participants who completed both the pre- and post-VE free associations. This section first explains the background of the VE project that underpins this study and then explains how data were gathered and analyzed using the free association method.

Background of the VE

The VE project was held in a five-week seminar toward the end of the second semester (September to January) in the academic year 2022-23 at the Department of Foreign Languages at a private Japanese university.

Seminars were a compulsory unit of the curriculum for second year students majoring in English, and the students had been automatically assigned to one of the 16 seminars held simultaneously once a week for 90 minutes. Instructor A and Instructor B had decided to conduct VE together, so the two seminars taught by these instructors were turned into the format of a joint seminar consisting of 28 students. Instructor C was the VE coordinator of the Japanese university and took care of the communication between the Japanese university and the Taiwanese university though he did not directly teach the joint seminar. The theme of the VE project was sustainable tourism, which was the shared interest of both parties. All the weekly sessions were conducted asynchronously because of the scheduling differences between the two universities. The medium of communication between the Japanese and Taiwanese students was English. It should be noted again that this study focuses on only the Japanese students.

In week 1, an overview of the five-week seminar was given at the beginning. After the orientation, the instructors assigned the Japanese students into eight groups (three or four students each). The students joined two platforms: a team for the seminar on Microsoft Teams and a page on Flip. Each group then recorded a self-introduction video and uploaded it on the designated thread of Flip for Taiwanese students to watch. Next, the instructors asked the students to participate in pre-VE assessments. The students were asked to sign a consent form if they agreed to have their data used for research purposes, and all of them indicated consent. One of the instructors then gave a short lecture on sustainable tourism and explained the template the instructors had prepared for the Japanese students to use when making their respective group presentations on their hometown and sustainable tourism. A brief discussion session based on the lecture was held at the end of the class.

In week 2, the instructors first let each group know which Taiwanese group they would be working with. Each group then watched the self-introduction video made by their Taiwanese counterpart and left responses by text on Flip. The students then prepared for their presentation by gathering information and creating PowerPoint slides. The instructors walked around the classroom and provided content and

linguistic assistance as necessary. Those who could not finish the slides by the end of the class were asked to complete them by the following week.

In week 3, the students first practiced their presentation in groups. They then went to quiet places in different areas of the building and recorded their presentation using their smartphone. The maximum length of the presentation was set at 10 minutes in order to have each student speak for approximately 3 minutes. This was also due to the time limit on Flip. All the groups were able to upload the videos on Flip before the end of the class.

By week 4, the Taiwanese groups had uploaded their videos on the topic of their hometown and sustainable tourism, so in week 4, the Japanese students first watched the video in their group produced by the Taiwanese group they were corresponding with. The instructors then provided a template for making a response video, consisting of a thank you message, comments, and questions. The students prepared their responses based on the template, recorded them, and uploaded the video on Flip.

There was a one-month gap between week 4 and week 5 due to the winter vacation and public holidays. At the beginning of week 5, the students were immediately asked to take the post-VE assessment before they refreshed their memory of the project, as an attempt to capture what they had internalized. After the assessment, the students watched response videos uploaded by Taiwanese students and answered the questions by text. They then engaged in reflective writing about their VE experience and shared it with their classmates. A farewell video was then shot as a class and uploaded on Flip to conclude the VE project and the five-week seminar.

Free association methodology

Participants were provided with a piece of paper folded into five slots. They were given one prompt word at a time in their native language, and were allowed 1 minute per prompt to write down as many words as they

could associate with in their native language by looking at the prompt word. After generating associations with one prompt word, participants were asked to fold the paper and write on the next slot, so that they would not see what they had written for the previous prompt. This was a measure taken to avoid potential influences from previously generated associations to the next prompt.

Since the theme of the VE project was sustainable tourism, the five prompts provided to the participants were tourism, hometown, sustainability, Taiwan, and Taiwanese people. Both the pre- and post-VE free associations followed the same procedure with the same prompts. The pre-VE free association was administered before the VE began, and the post-VE free association was administered in the final week of the VE. Referring to the Ebbinghaus' forgetting curve (Murre & Dros, 2015), by the time of the post-VE free association, participants were expected to have a vague memory of the VE project after one month of winter holiday. However, the finding which will be explained later in this article was the opposite, which could serve as evidence of the participants' learning outcomes with longer-lasting effects.

Associated words written by students on paper were transcribed into an Excel document and listed under their respective prompt words. Then, thematic analysis was conducted to first explore common linguistic themes among the associated words by one of the authors. Upon discovering emerging themes, the other author and an instructor involved in the VE project examined the words and their emerged themes together for triangulation. In cases where there were disagreements on the themes, they were resolved through thorough discussions. The following themes emerged: action (e.g., travel, dance), place (e.g., Osaka, hot springs), object (e.g., food, resources), people (e.g., friends, neighbors), concept (e.g., awareness, eco-friendliness), and feelings (e.g., comfortable, enjoyable). After the associated words are categorized into their respective themes, the number of associated words and the number of types under each theme were calculated. The same associated words generated by different participants under the same theme are calculated as "types" of words.

After the classification was completed, all the words were translated from Japanese into English considering the readership of this paper. Some of the examples mentioned below may appear to consist of more than one word; however, this was due to translation and the Japanese words in the original data were all judged as consisting of only one word.

Results and Analysis

Pre-VE free association

In the pre-VE free association, a total of 589 words were produced by the participants across the five prompts: 148 words (80 types) for tourism, 126 words (91 types) for hometown, 105 words (71 types) for sustainability, 124 words (77 types) for Taiwan, and 86 words (64 types) for Taiwanese. The median number of words produced by each participant was as follows: 7.5 words for tourism (min: 3, Max: 12), 6 words for hometown (min: 2, Max: 12), 4.5 words for sustainability (min: 1, Max: 14), 6 words for Taiwan (min: 2, Max: 14) and 4 words for Taiwanese (min: 1, Max: 8). None of the participants produced the same word more than once for any of the prompts. There were two students who consistently produced a small number of words and two students who produced a large number of words across the prompts, and therefore, the minimum number and the maximum number were mostly due to these students.

Among the five prompts, tourism turned out to be the least demanding prompt for the participants as 70% of them were able to write down words at the rate of more than 1 word per 10 seconds on average. At the same time, 28 words were written by at least two participants, and the total number of types was not significantly larger than other prompts. This indicated that the participants already had a certain group of words associated with tourism in their mind pre-VE and tended to write similar words for this prompt. On the other hand, Taiwanese was the most challenging prompt, and the participants needed more than 15 seconds on average to produce one word. They did not have prior experience studying with Taiwanese students and likely did not have clear ideas regarding the prompt. This was also evident in the nature of words they

produced (see below). The participants also struggled with thinking of words associated with sustainability, which resulted in the largest standard deviation in the number of responses produced by each participant (SD=3.5). Coming up with more words, however, did not necessarily reflect more knowledge of sustainability as the qualitative analysis below will show.

The prompts hometown and Taiwan elicited approximately the same number of responses from the participants. However, there was less overlap regarding the words produced for the former prompt. Although the number of words produced by more than two participants was 17 for both prompts, the total number of types was larger for hometown than for Taiwan. This showed that the prompt Taiwan was not more challenging than the prompt hometown but that the participants tended to hold similar impressions and images toward Taiwan.

Table 1 shows the frequencies and proportions of responses classified according to the six themes of actions, concepts, feelings, objects, people, and places. Regarding tourism, places constituted 37.8% of the data. Two in 3 words were general words (e.g., temples, hot springs, foreign countries, world heritage sites), while the rest were specific city names or country names (e.g., Osaka, Kyoto, Singapore). Objects also appeared frequently, and 17 words were related to transportation such as bullet trains and airplanes. A similar tendency was observed for the prompt Taiwan as the majority of words were classified into the categories of objects (37.9 %) and places (33.1%). Among the objects, 33 words were related to food, and only 3 words were on transportation methods. The prompt hometown also elicited various words related to places; however, unlike tourism and Taiwan, words relevant to people such as friends and family constituted nearly 20% of the responses.

For the prompt sustainability, words on actions (41.9%) and concepts (37.1%) were widely observed. Half of the action words (22 words) were on daily routines including running and studying, and all the concept words apart from five instances of SDGs (sustainable development goals) were general ideas (e.g., stability, persistence, durability, concentration, diversity). This meant that the participants did not have concrete ideas regarding sustainability pre-VE and had associated the

term with their daily lives, not social or environmental issues that were discussed in VE. For the final prompt, Taiwanese, words on feelings appeared most frequently. More than 75% were positive words (28 words), indicating that the participants held a positive attitude toward Taiwanese people in the first place. At the same time, seven words on feelings indicated uncertainty and most of the words on concepts were general and vague (e.g., rich, tall, flat), indicating the participants' unfamiliarity with actual Taiwanese people.

Table 1. Frequencies and Proportions of Each Prompt in the Pre-VE Free Association

	Tourism	Hometown	Sustainability	Taiwan	Taiwanese
Actions	22	9	44	8	3
	(14.9%)	(7.1%)	(41.9%)	(6.5%)	(3.5%)
Concepts	15	14	39	12	20
	(10.1%)	(11.1%)	(37.1%)	(9.7%)	(23.3%)
Feelings	5	14	9	10	36
	(3.4%)	(11.1%)	(8.6%)	(8.1%)	(41.9%)
Objects	39	10	10	47	3
	(26.4%)	(7.9%)	(9.5%)	(37.9%)	(3.5%)
People	11	23	2	6	17
	(7.4%)	(18.3%)	(1.9%)	(4.8%)	(19.8%)
Places	56	56	1	41	7
	(37.8%)	(44.4%)	(1.0%)	(33.1%)	(8.1%)
Total	148	126	105	124	86
	(100%)	(100%)	(100%)	(100%)	(100%)

Post-VE free association

In the post-VE free association, 658 words were elicited: 156 words (102 types) for tourism, 148 words (109 types) for hometown, 104 words (77 types) for sustainability, 150 words (93 types) for Taiwan, and 100 words (61 types) for Taiwanese. The median number of words elicited from each participant for the five prompts was as follows: 8 words for tourism (min: 2, Max: 14), 7.5 words for hometown (min: 3, Max: 12), 5 words for sustainability (min: 4, Max: 10), 7.5 words for Taiwan (min: 4, Max: 12), and 5 words for Taiwanese (min: 3, Max: 9). There was one

participant who wrote the same place name twice for tourism, but other than this one exception, the participants did not repeat the same word.

The total number of words produced for three of the five prompts, namely, tourism, hometown, and Taiwan was approximately the same, and more than 80% of the participants could write words faster than at the rate of 1 word per 10 seconds for these three prompts. The participants who only produced a small number of words consistently did so across most of the prompts. Considering the ratio of the number of words and the number of types, the prompt hometown elicited a wider range of words compared to the other two prompts, and indeed, there were only 16 words produced by more than two participants. This indicated that the three prompts in post-VE free association were not challenging for the participants but that they tended to come up with similar groups of words for tourism and Taiwan in comparison to hometown.

Taiwanese and sustainability were relatively demanding prompts for the participants although more than half of them managed to come up with at least 5 words. The words produced for sustainability were more diverse compared to those produced for Taiwanese, and the standard deviation of the number of responses per participant was smaller for sustainability (SD=1.5). This meant that the participants held different ideas about sustainability, whereas they had similar impressions of Taiwanese people.

The responses were classified into the six themes explained above, and the frequencies and proportions for each prompt are summarized in Table 2. As the table shows, over 1/3 of the words elicited by the prompts tourism and hometown were classified as places. Fifteen words for tourism and 14 words for hometown were specific place names. The category with the second largest number of words for tourism was objects (e.g., tickets, local mascot, famous product), followed by concepts (e.g., industry, time, expenses) and actions (e.g., travel, experience, walk). However, for hometown, people (e.g., childhood friend, neighbor) came second, followed by concepts (e.g., retro, safety, nature) and objects (e.g., sweet potatoes, green tea, taco rice). The

majority of words emerged for the prompt Taiwan were objects and places. More than half of the objects were food names, but there were words on movies (e.g., Spirited Away), culture (e.g., firecracker, lantern), and fashion items (e.g., glasses). Different types of buildings (e.g., temple, food stall, castle) rather than specific place names appeared frequently in the category of places.

Table 2. Frequencies and Proportions for Each Prompt in the Post-VE Free Association

	Tourism	Hometown	Sustainability	Taiwan	Taiwanese
Actions	23	11	45	5	3
	(14.7%)	(7.4%)	(43.3%)	(3.3%)	(3.0%)
Concepts	26	22	46	23	18
	(16.7%)	(14.9%)	(44.2%)	(15.3%)	(18.0%)
Feelings	8	12	9	15	55
	(5.1%)	(8.1%)	(8.7%)	(10.0%)	(55.0%)
Objects	32	20	3	56	7
	(20.5%)	(13.5%)	(2.9%)	(37.3%)	(7.0%)
People	10	26	0	4	14
	(6.4%)	(16.9%)	(0%)	(2.7%)	(14.0%)
Places	57	58	1	47	3
	(36.5%)	(39.2%)	(1.0%)	(31.3%)	(3.0%)
Total	156	148	104	150	100
	(100%)	(100%)	(100%)	(100%)	(100%)

Words on concepts and actions both constituted over 40% for the prompt sustainability. Concepts included words such as SDGs, environmental friendliness, energy conservation, ecosystem, and product performance, which were concrete ideas related to sustainability. More than 90% of the words on action (e.g., electricity generation, recycling, cooperation) were also closely relevant to sustainability of the environment and society. More than half of the words elicited by the prompt Taiwanese were on feelings. All but two words (a little scary, not punctual) were positive words such as friendly, kind, diligent, cheerful, and active), likely reflecting the participants' positive learning experiences with the Taiwanese students.

Comparison between pre- and post-VE free association

The total number of words increased from 589 to 658 between pre- and post-VE free association, and this was an 11.2% increase. Table 3 below summarizes the rate of increase of the number of words and types for each prompt. In the pre-VE assessment, the participants came up with the largest number of words for tourism (148), followed by those for hometown (126) and Taiwan (124). The gap between these two groups closed in the post-VE assessment (156, 148, 150), and this probably meant that most participants reached the upper limit of the number of words they could write within one minute although there were individual differences. Indeed, the number of words for tourism only increased by 5.4% despite the increase of nearly 30% regarding the number of types.

There was over 15% increase in the number of words for the prompts hometown, Taiwan, and Taiwanese, and this change can be regarded as a large increase considering the time limit the participants had for each prompt. The number of types also increased by approximately 20% for the former two prompts, indicating that the range of items associated with these two prompts became more diverse post-VE. The number of types decreased for Taiwanese, but this was mainly because the participants tended to write a group of complimentary words with higher frequency in the post-VE free association as described in the section above. Regarding the prompt sustainability, there was a slight decrease in the number of words; however, the participants were able to think of concepts and actions more closely associated with environmental issues as mentioned above. These results indicate that the participants' overall perception of the five prompts shifted through the VE experience.

Table 3. Rate of Increase of the Number of Words and Types for Each Prompt

	Tourism	Hometown	Sustainability	Taiwan	Taiwanese
Words	5.4%	20.3%	-1.0%	18.7%	16.3%
Types	28.8%	19.8%	8.5%	21.1%	-4.7%

Chi-square tests conducted by SPSS Ver.29 revealed that the distribution of the number of words classified according to the six themes did not change significantly pre- and post-VE for any of the prompts [tourism: $\chi2(5)= 4.205$, $p = .520$, hometown: $\chi2(5)= 3.842$, $p = .572$, sustainability: $\chi2(5)= 6.352$, $p = .273$, Taiwan: $\chi2(5)= 4.317$, $p = .505$, Taiwanese: $\chi2(5)= 6.546$, $p = .257$]. This meant that the theme of words the participants associated with the prompts stayed the same after the VE project. However, the specific words elicited by free association changed between pre- and post-VE free association, and this was clear from an analysis of unique types produced for each prompt.

Table 4 shows the number of unique types observed in pre- and post-VE free association data. Unique types here refer to words that only appeared either in pre- or post-VE free association, and the percentage indicates the number of unique types out of all the types observed in the free association. Note again that words written by more than one participant for the specific prompt were counted as one type regardless of the frequency.

Table 4. Number of Unique Types of Words

	Tourism	Hometown	Sustainability	Taiwan	Taiwanese
Pre-VE	40	45	53	42	48
	(50.0%)	(49.5%)	(74.6%)	(54.5%)	(75.0%)
Post-VE	62	78	58	57	46
	(60.8%)	(71.6%)	(75.3%)	(61.3%)	(75.4%)

For the prompt tourism, there was a prominent increase in the number of concepts related to sustainable tourism. Pre-VE, the concepts elicited by free association tended to be general and vague such as holidays, hospitality, and business. After the VE, the participants instead wrote concepts such as awareness toward the destination, overcrowdedness, and overtourism, which were closely relevant to what they had learned through the VE project. In addition, the number of city and country names decreased, and the place names that had appeared in the presentations emerged more often. Similarly, all the concepts that appeared in the post-VE free association for hometown were unique

types and reflected what the participants had found out about their hometown through preparing their presentations (e.g., depopulation, local production and consumption, retro). Words related to people including family and friends emerged in free association both pre- and post-VE, with friends being the word of the highest frequency.

Regarding the prompt sustainability, the nature of words elicited by free association changed post-VE. As described above, the participants had not necessarily associated sustainability with environmental and social issues pre-VE, but they came up with words closely relevant to the environment and society post-VE. This meant that the VE project shifted the participants' perception of sustainability. Words produced for the prompt Taiwan became more specific and detailed post-VE as shown above, and this likely reflected the participants' increase of knowledge toward Taiwan through communicating with Taiwanese university students. The participants wrote positive, complimentary words with more certainty for the prompt Taiwanese post-VE, and the words on concepts that emerged post-VE referred to the relationship with Japan (e.g., similar to Japan, interest in Japan). Kind and friendly were written by nearly half of the students, reflecting a positive learning experience with Taiwanese students.

Discussion

Aligning with the various studies that show the positive impact of VE (Chu et al., 2024; Kilimci, 2010; O'Dowd, 2003; Siergiejczyk, 2020; Vinagre, 2010), the VE in this study also showed positive impact in helping students obtain content knowledge.

First, the number of words generated from free association in post-VE increased by 16.3% to 20.3% in three out of five prompt words (Taiwanese, Taiwan, Hometown), and then the types of words also increased significantly in three out of five categories by 19.8% to 28.8% (Hometown, Taiwan, Tourism). This could be interpreted as shifts in perception and deepened knowledge as students were able to intuitively associate more words and also more types of words with the prompts. It

must be mentioned again that the post-VE free association was administered after a near one-month winter break, and we were expecting students to have new short-term memories during the break to replace their short-term memories obtained from the VE. Thus, we believe it is a significant finding to see students associating more words than pre-VE with free association.

Next, evidence supporting the measurement of IC, specifically cultural knowledge and attitude, could be found from the prompts of Taiwan and Taiwanese. A 15% increase in the number of words for the prompts Taiwan and Taiwanese could be regarded as a large increase considering the 1-minute time limit the participants had for each prompt. Such an increase reflects students' expanded cultural knowledge towards both a foreign country and foreign nationals. The number of types of words also increased by 21.1% for the prompt Taiwan, indicating that the range of items associated with Taiwan became more diverse. However, the number of types of words decreased by 4.7% for the prompt Taiwanese. This was not interpreted as students reduced less knowledge for Taiwanese, but instead, a significant increase of words in the "feeling" category were associated in the post-VE, which reflects their experience of authentic interactions with Taiwanese partners. Students tended to write a group of complimentary words with higher frequency in the post-VE. We must emphasize once again that the data from this study were derived from an implicit assessment, therefore any observed changes in the post-VE free association is assumed to reflect internal outcomes of the intercultural communication experience. This kind of intuitive data is assumed to be closer to what students really think compared with data gathered with explicit assessments, where factors such as cognitive loading, social desirability bias, and Dunning-Kruger Effect, might interfere with deliberate responses.

In addition to an increased number of associated words, free association also revealed a shift of perspective. Take for example words associated with the prompt word "Taiwanese", Japanese students became less presumptuous, from pre-VE free association associated words such as "walk fast", "many females", "aggresive", "not wealthy", to post-VE free association associated words such as "English", "friendly", "bright",

and "wearing glasses". This kind of data suggests the possibility of reducing potential social desirability bias (Nederhof, 1985), and avoiding the Dunning-Kruger effect (Kruger & Dunning, 1999). First, from the authors' knowledge and experience of working with Japanese undergraduate students, we think that if we were to ask students to share explicitly how they would imagine Taiwanese people to be like, it would be rare to see responses such as they have "flat facial features", "wear glasses", or "not wealthy." These responses might be considered overly presumptuous or slightly foolish, which in the Japanese context, students would not likely share with others due to concerns about how others would perceive themselves. Hence, the free association method might have bypassed social desirability bias. As for circumventing the Dunning-Kruger effect, the number of words and the nature of words they could associate with the prompt words are straightforward evidence. They expanded their perspectives and knowledge about tourism, their hometown, sustainability, and Taiwan, and refined their understanding of Taiwanese people to be less presumptuous.

Conclusion

Free association has the potential to capture the internal outcomes in IC of students after experiencing intercultural communication. The application of free association in this study addressed the challenges of explicit assessments by intentionally reducing the cognitive of the assessment task to the minimum. Students only had to read one prompt word at a time, freely write any words that they associated with, and without having to conform to linguistic rules. Students not only retained but also deepened both content and cultural knowledge, where the post-VE free association was conducted after a one-month holiday, which allowed the findings to be even more significant when referring to Ebbinghaus' forgetting curve. The implication here is that the shifts in perception and knowledge have likely been internalized and are possibly long-lasting.

Additionally, free association might be able to mitigate the Dunning-Kruger effect in comparison to explicit assessments such as self-

evaluation or reflective writings. A one-month break from the VE led to the assumption that the majority of students would be unlikely to overestimate themselves since they did not engage in any tasks related to the VE during the break, and overestimation could lead to faster retrieval of associations, meaning an increased chance of generating more associated words. As a result, students associated more words and also the types of words, which is not likely to happen if there is a Dunning-Kruger effect.

Lastly, in the cultural context of Japan where social desirability is very common, students associated words with free association that might not otherwise be expressed with explicit assessments, for instance, their negative or presumptuous feelings towards a foreign culture. This finding implies that the social desirability effect might have been mitigated to a certain extent with free association. Therefore, we believe that using free association as an implicit assessment to observe changes in internal outcomes of IC in this study successfully reduced social desirability bias, Dunning-Kruger effect, and also the cognitive loading of the students, thus enabling us to gather intuitive data that might be more authentic or closer to what students were really thinking.

One of the limitations of the study was that the sample size was small with 20 participants. The study also took place in Japan, so it remains to be seen whether the findings are applicable to other cultural contexts. In addition, there still might have been some degree of conscious filtering involved during free association.

Given the limitations of this study, perhaps future studies could explore using free association in VE with a larger sample, in different cultural contexts, and on controversial topics to further investigate how free association might be able to avoid social desirability bias. Also, comparing findings from free association with explicit assessments might shed more light on how the Dunning-Kruger effect might interfere with gathering intuitive data, which could result in increased understanding of how conscious filtering may still have been involved.

Authors

CURTIS CHU, PhD, is an assistant professor at the Faculty of International Studies at Setsunan University, Osaka, Japan. He was the VE/COIL coordinator of the institution, and is currently on the executive team of the World Council on Global and Intercultural Competence. He was a lecturer at a public university in Taiwan, and has over 20 years of experience teaching English from elementary to graduate school levels. His interests include institution globalization, intercultural competence, virtual exchange, and TESOL.

MARIKO TAKAHASHI, PhD, is an associate professor at the School of Business Administration at Kwansei Gakuin University, Hyogo, Japan. Her current research interests include corpus-based analyses of features of Asian Englishes, the development of tasks and projects for EFL classrooms based on the approach of content and language integrated learning, and the analysis of English used in Japanese popular culture. She can be contacted at mariko.takahashi@kwansei.ac.jp

References

Araki, N. (2018). Dotoku kyoiku no tameno hyokaho: Rensoho o tsukatta kyoikuhyoka no hoho, renso map ni tsuite [The evaluation method for moral education: Use of the word-association maps]. *The Japanese Journal of Educational Practices on Moral Development, 11*(1), 50-101.

Archard, P., & O'Reilly, M. (2023). Qualitative research interviewing: application and use of free association. *Nurse researcher*, *31*(3), 22–29. https://doi.org/10.7748/nr.2023.e1875

Bazgan, M., & Norel, M. (2013). Explicit and implicit assessment of intercultural competence. *Procedia-Social and Behavioral Sciences, 76,* 95-99. https://doi.org/10.1016/j.sbspro.2013.04.080

Byram, M. (1997) *Teaching and assessing intercultural communicative competence*. Multilingual Matters.

Chu, C., DeBellis, C., Cartwright, C., Khan, F., Pu, C., & Tang, W. (2024a). The practice of assessing intercultural competence - A preliminary global study. In Deardorff, D. K., & Tatebayashi, M. (Eds.), *Current perspectives on intercultural and global competence* (pp. 36-62). STAR Scholars Press. https://doi.org/10.32674/jf1a4r48

Chu, C., Hooper, T., Takahashi, M., & Herke, M. (2024b). The effect of asynchronous virtual exchange on intercultural competence among undergraduate students. *Journal of Comparative & International Higher Education*, *16*(2), 213-226. https://doi.org/10.32674/jcihe.v16i2.5229

De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies* (pp. 179 -194). Guilford Press.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological bulletin*, *135*(3), 347-368. https://doi.org/10.1037/a0014211

Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education, 10*(3), 241–266. https://doi.org/10.1177/1028315306287002

European Commission: Directorate-General for Education, Youth, Sport and Culture. (2024). *Erasmus+ annual report 2023*. Publications Office of the European Union.

Fantini A. (2007) *Exploring and assessing IC* (CSD Research Paper No. 07-01). Washington University, Center for Social Development. https://doi.org/10.7936/K7TB16CX

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-235). Lawrence Erlbaum.

Freud, S. (1901). *Psychopathology of everyday life*. Basic Books.

Galton, F. (1880). Psychometric experiments. *Brain, 2*, 149-162.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit

association test. *Journal of Personality and Social Psychology, 74*, 1464-1480. https://doi.org/10.1037/0022-3514.74.6.1464

Griffith, R. L., Wolfeld, L., Armon, B. K., Rios, J., & Liu, Q. L. (2016). Assessing intercultural competence in higher education: Existing research and future directions. ETS research report. ETS.

Hollway, W. (2000). *Doing qualitative research differently: Free association, narrative and the interview method.* Sage Publication.

Hollway, W., & Jefferson, T. (2008). The free association narrative interview method. In L. M. Given (Ed.), *The SAGE encyclopedia of qualitative research methods* (pp. 296-315). Sage.

Juniper E. F. (2009). Validated questionnaires should not be modified. *The European Respiratory Journal*, *34*(5), 1015–1017. https://doi.org/10.1183/09031936.00110209

Kamizono, K. (2009). Death consciousness in Germany, China and Japan by association method. *Bulletin of Faculty of Education, Nagasaki University, Educational Science, 73,* 1-16.

Kent, G. H., & Rosanoff, A. J. (1910). A study of association in insanity. *American Journal of Psychiatry*, *67*(1), 37-96.

Kilimci, S. (2010). Integration of the internet into a language curriculum in a multicultural society. *Turkish Online Journal of Educational Technology*, 9(1), 107-113.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134. https://doi.org/10.1037//0022-3514.77.6.1121

Lai, E. R. (2011). Metacognition: A literature review. *Always Learning: Pearson Research Report*, *24*, 1-40.

McDowell, W. S. (2004). Exploring a free association methodology to capture and differentiate abstract media brand associations: A study of three cable news networks. *Journal of Media Economics, 17*(4), 309–320. https://doi.org/10.1207/s15327736me1704_5

Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PLOS ONE, 10*(7), e0120644. https://doi.org/10.1371/journal.pone.0120644

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, *15*(3), 263-280. https://doi.org/10.1002/ejsp.2420150303

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402-407. https://doi.org/10.3758/BF03195588

O'Dowd, R. (2003) Understanding the 'other side': intercultural learning in Spanish-English e-mail exchange. *Language Learning & Technology, 7*(2), 118-144.

O'Dowd, R. (2021). What do students learn in virtual exchange? A qualitative content analysis of learning outcomes across multiple exchanges. *International Journal of Educational Research, 109*, 101804. https://doi.org/10.1016/j.ijer.2021.101804

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277-293. https://doi.org/10.1037/0022-3514.89.3.277

Project Implicit. (2011). Overview of the implicit association test. Retrieved from https://implicit.harvard.edu/implicit/education.html.

Rosset, E. (2008). It's no accident: Our bias for intentional explanations. *Cognition, 108*(3), 771-780. https://doi.org/10.1016/j.cognition.2008.07.001

Schnabel, K., & Asendorpf, J. B. (2013). Free associations as a measure of stable implicit attitudes. *European Journal of Personality*, *27*(1), 39-50. https://doi.org/10.1002/per.1890

Sercu, L. (2004). Assessing IC: a framework for systematic test development in foreign language education and beyond. *Intercultural Education*, *15*(1), 73-89. https://doi.org/10.1080/1467598042000190004

Sercu, L. (2010). Assessing IC: More questions than answers. In A. Paran & L. Sercu (Eds.), *Testing the Untestable in Language Education* (pp. 17-34). Multilingual Matters.

Siergiejczyk, G. (2020). Virtual international exchange as a high-impact learning tool for more inclusive, equitable and diverse classrooms. *European Journal of Open, Distance and E-Learning*, *23*(1), 1-17. https://doi.org/10.2478/eurodl-2020-0001

Stewart, A. L., Thrasher, A. D., Goldberg, J., & Shea, J. A. (2012). A framework for understanding modifications to measures for diverse populations. *Journal of Aging and Health*, *24*(6), 992-1017. https://doi.org/10.1177/0898264312440321

Stevens Initiative. (2023). 2023 Virtual exchange impact and learning report. The Aspen Institute. https://www.stevensinitiative.org/resource/2023-virtual-exchange-impact-and-learning-report/

Szalay, L.B., & Deese, J. (1978). *Subjective meaning and culture: An assessment through word associations*. Routledge.

Vinagre, M. (2010). Intercultural Learning in asynchronous telecollaborative exchanges: A case study. *Eurocall Review, 17*, 37-47. https://doi.org/10.4995/eurocall.2010.16324

Zotzmann, K. (2015). The impossibility of defining and measuring intercultural competencies. In Rivers, D.J. (Ed.) *Resistance to the Known*. Palgrave Macmillan. https://doi.org/10.1057/9781137345196_8