

Designing an In-house Online English Placement Test — Implementation and Results

Bradley IRWIN*¹, Diego OLIVEIRA*¹ and Adrienne VERLA UCHIDA*¹

学部内オンライン英語レベル判定テストの設計—実装と結果

B.アーウィン*¹・D.オリベイラ*¹・A.ヴァーラ内田*¹

The following paper describes the process of creating an online English placement test for the English department at the Nihon University College of International Relations. The objectives of this paper are two-fold: to thoroughly explain and justify the design choices of the authors; and, to serve as a guide for other researchers who may consider designing their own placement tests. Issues relating to the elimination of technical problems, ensuring fairness and honesty, and the accurate assessment of language proficiency are identified and discussed. Moreover, using Google Forms as an appropriate platform for creating an online placement test is explained in detail. The paper ends with a brief discussion of the challenges of online collaboration and recommendations for future research.

本稿では、日本大学国際関係学部のオンライン英語レベル判定テストの作成プロセスについて記述する。本稿の目的は2つあり、著者のテスト作成上の選択を具体的に説明し正当化することおよび、レベル判定テストの作成を検討する可能性のある他の研究者に指導を与えることである。技術的な問題を排除し、公平性と誠実さを保ち、また言語能力の正確な評価に関連する諸問題を特定し、それらについて議論する。さらに、オンラインレベル判定テストを作成するための適切なプラットフォームとしてGoogleフォームを使用する方法について詳しく解説する。最後に、オンラインコラボレーションの課題と将来の研究に関する推奨事項について簡単に議論する。

Keywords: Online placement test, language proficiency test, language test construction, Google Forms

The 2020 and 2021 academic years were thrown into great turmoil as the COVID-19 pandemic spread throughout Japan. As a measure to reduce its spread, the vast majority of classes were moved online forcing teachers and students alike to engage in Emergency Remote Teaching (Milman, 2020; Hodges et al., 2020).

Unfortunately, conducting large-scale tests, which had been vital to the smooth operation of the English department at the Nihon University College of International Relations, were similarly affected by the cancellation of face-to-face activities. Proficiency tests, such as the TOEIC, TOEFL, and ACE tests, were initially cancelled outright before online versions were developed

*1 日本大学国際関係学部国際教養学科 助教 Assistant Professor, Department of International Liberal Arts, College of International Relations, Nihon University

and implemented. The TOEIC and TOEFL tests typically measure students' English reading and listening skills with multiple choice questions and take approximately 2 hours to complete. Similarly, the ACE test measures English language reading and writing skills with multiple choice test items but takes 1 hour to complete. While the online versions of the TOEIC and TOEFL tests were developed quickly, the ACE test was not. This presented the English department with a considerable problem because the ACE test had been used for several years as a placement test to stream incoming students in the appropriate first-year English classes that best suited their abilities. Not only was the ACE test used for first-year placement purposes, it was also used as a means to check student progress and to stream students into second year English courses. Therefore, pre-COVID-19, the ACE test was administered three times each academic year in April, July, and December, and was compulsory for all first-year students enrolled in English classes.

Because the ACE test played such a central role in creating classes with students of similar English abilities within the English Department, simply cancelling it was not a viable option. Therefore, it was decided that since the authors of the present study had experience designing and conducting online tests, they would work towards creating an in-house online English placement test.

Conducting an online test of such importance is not without its challenges. Similar to the concerns raised in the Organization for Economic Co-operation and Development (OECD, 2020) report on online exams, we wanted to create a test that was free from technical problems, ensured fairness and honesty, and above all, assessed English proficiency for placement purposes. Therefore, several key considerations went into the original design of the test.

Avoiding Technical Problems

Online testing presents a unique set of challenges because of a wide array of technical problems that can arise. Mechanical breakdowns, internet connectivity issues, and server bottlenecks are among the most cited problems. Exacerbating potential technical problems, a recent study conducted by Irwin (2021) showed students accessed their online courses using a variety of devices (smartphones, tablets, laptops, and PCs) and internet data plans (no access, limited, and unlimited). This situation meant that the elimination of all technical problems would be out of the control of the test designers. Therefore, the goal for administering the test was to limit technical problems to the greatest degree possible. To achieve this, a simple test design was utilized within a system that could accommodate a wide variety of devices in suboptimal online conditions. The issues of testing environment and user experience are also closely connected to the notion of test fairness discussed in the next section.

Ensuring Fairness and Honesty

While fairness in testing includes issues such as cultural bias, test validity, equal access, and social consequences (Kunnan, 2004), our initial concern was to create an equal testing environment that offered a similar user experience to all test takers. In order to facilitate this, it was decided that the test would utilize a system that could accommodate a number of different devices ranging from desktop PCs to smartphones. Several methods were tested, including creating a basic website and using a third-party service, but it was decided that using Google Forms to administer the test would offer a robust user experience that was similar across devices. Like many Google productivity apps, Google Forms is very intuitive and user-friendly, thus lowering student anxiety. Students can take the

test by clicking on a link or scanning a QR code, entering their contact information, and then choosing their answers by clicking on their screen. Another benefit to using this system is that students already had Google Workspace for Education accounts provided by the university and many of them had already had experience using Google Forms in their online classes.

The issue of honesty was considered at great length. While a number of studies have shown that students are not more likely to cheat in online learning environments (Ladyshevsky, 2015; Peled et al., 2019; Peterson, 2019), particularly in low stakes testing situations (Grijalva et al., 2006; Munoz & Mackay, 2019), we decided to remind the students of the importance of academic honesty by including the following statement in both English and Japanese :

The purpose of this test is to evaluate your current English proficiency level for your English classes starting in April. Your score will be used to determine the level of English course you will take. Therefore, we ask you to take this test honestly and without cheating. Cheating will mean that you will be placed in an English class that does not match your ability. 英語オンラインプレースメントテストの目的は、あなたの現在の英語能力レベルを評価することです。スコアは、受講する英語コースのレベルを決定するために使用されます。したがって、試験は不正行為をせずに正直に受けてください。不正行為をすると、自分の能力に合わない英語クラスに入ることになります。

Therefore, we relied on the honour system and hoped that since the test was being used for placement purposes only, students would not be tempted to cheat.

Assessing English Proficiency

As the main purpose of the placement test was to stream students into appropriate levels of first- and second-year English language courses, adequately assessing the students' English proficiency was paramount. Therefore, it was decided that the test should include four constructs: listening, vocabulary, grammar, and reading. While each construct will be described in detail in subsequent sections of this paper, Table 1 offers a brief breakdown of the test sections.

Table 1

Breakdown of Test Sections and Questions for Each Construct

Section	Number of questions	Number of points	Percentage of test
Listening 1	3	15	12%
Listening 2	3	17	
Vocabulary	14	80	29%
Grammar	28	120	44%
Reading 1	3	12	15%
Reading 2	4	28	
Total	55	272	100%

The listening section represented 12% of the test and was divided into two parts with three questions each for a total of 32 points. The vocabulary section was composed of 14 questions for a total of 80 points, which represented 29% of the test. The grammar section contained 28 questions for a total of 120 points representing 44% of the test. Finally, the reading section, which represented the remaining 15% of the test, was divided into two parts for a total of seven questions for 40 points.

Only multiple-choice questions with four answer choices were designed for the test. These questions were scored differently to assess students' different levels of knowledge. To determine the difficulty level of questions within each construct, corpus-based word lists by English Profile (2015), which align with the Common European Framework of Reference (CEFR), and the New General Service List (Browne et al., 2013), which cover the most frequently used English words were compared. Questions that measured basic level knowledge of the language were scored three points, intermediate level six points, and advanced level eight points. The decision to utilize only multiple-choice questions had three reasons. First, as mentioned previously, students used different types of devices to take the test (Irwin, 2021). When utilizing Google forms certain question types such as matching questions, can be difficult to visualize in smaller devices such as smartphones and tablets. Second, it has been shown that multiple choice questions can be a fair way of evaluating test takers knowledge of a subject when questions are well constructed (Brady, 2005; Xu et al., 2016). Last, given the large number of test takers, open ended questions would have made it impossible to grade all the tests in a viable time frame.

Planning and Test Design Preparation

During the initial planning phase, it was decided that the placement test would cover a wide range of topics and incorporate grammar and vocabulary items using CEFR vocabulary lists and CEFR "can do" statements. Additionally, as the authors had collaborated throughout the spring semester to design various grammar quizzes, a quiz bank with over 400 items ranging from basic, intermediate, and advance levels were available to adapt and utilize for the placement test. An example of adapting a test question would be as follows: Original - "There _____ any rice in the cupboard." Final - "There _____ any meat in the fridge." Moreover, various grammar and reading textbooks from a variety of publishers were consulted when making new test items for the different levels.

Google Forms Specifics

The test was divided into separate sections utilizing the Google Forms quiz template. All titles, explanations and commands were written in both English and Japanese to assure understanding from students of all levels.


The first section, consent and personal information, had a disclaimer explaining the objective of the placement test and the conduct expected from students taking the test. This section also included a short explanation on how to enter personal information in the form and spaces for students to enter their name and student number.

The next section, listening comprehension, started with a short explanation on how to listen to the audio of the section, an embedded YouTube video where students could listen to the passage, and questions about each listening passage.

Listening skills were measured utilizing two passages and six multiple choice questions. Questions in the first listening passage assessed basic level skills using a dialogue with a picture connected to the conversation. Questions from this passage asked test takers about the main topic of the dialogue and to identify locations in the picture connected to the dialogue. Questions in the second listening passage assessed intermediate-advanced level skills. Here test takers were asked comprehension questions about a longer listening passage and were not given any visual aids. The two listening sections together had a total of two questions scored three points, three questions scored six points, and one question scored eight points. Figure 1 is an example of a question from the listening section of the test.

Figure 1

Example of a Basic Level Listening Comprehension Question



Where has the man already looked? Choose the alphabet(s) that match the location(s) in the picture.

Where has the man already looked?

A	<input type="radio"/>
B	<input type="radio"/>
C	<input type="radio"/>
D	<input type="radio"/>

The following sections, vocabulary and grammar, included 14 and 28 questions respectively. Grammar and vocabulary questions were all sentence completion style questions. These questions asked test takers to choose the most suitable word or grammar structure to complete a given sentence. Grammar and vocabulary sections together had 22 questions scored three points, 13 questions scored six points, and seven questions scored eight points. Figure 2 is an example of a question from the grammar section of the test.

Figure 2

Example of Basic Level Grammar Question test.

There _____ any money in my bank account.

- is
- isn't
- are
- aren't

The final section of the test, reading comprehension, had two separate reading passages. The first reading passage was a poster advertisement of a car followed by three questions and a five-paragraph essay with four questions. The poster had a drawing of a car, the name of the car, information about the price and features of the car, and information about the car maker. Questions in the first reading passage assessed basic to intermediate level reading skills. The second reading passage was a short five paragraph academic essay in favor of allowing students to use smartphones in class. This passage intended to assess test takers intermediate to advanced level reading skills. To answer questions in this reading passage, test takers were required to have a general comprehension of the reading as well as being able to understand the different arguments being made by the writer. There was a total of seven questions in the entire section. Two questions had a score of three points, three questions a score of six points, and two questions a score of eight points.

Test Implementation

To date, the online English placement test has been conducted twice at the College of International Relations ; in December 2020 and April 2021. While the test items were nearly identical in terms of focus, the wording of questions were slightly altered.

The December 2020 Placement Test

The December 2020 placement test was conducted to stream students into second year English courses. Therefore, only those students who were planning on taking English courses the following academic year were required to take the test. The total number of participants was 399 (N=399).

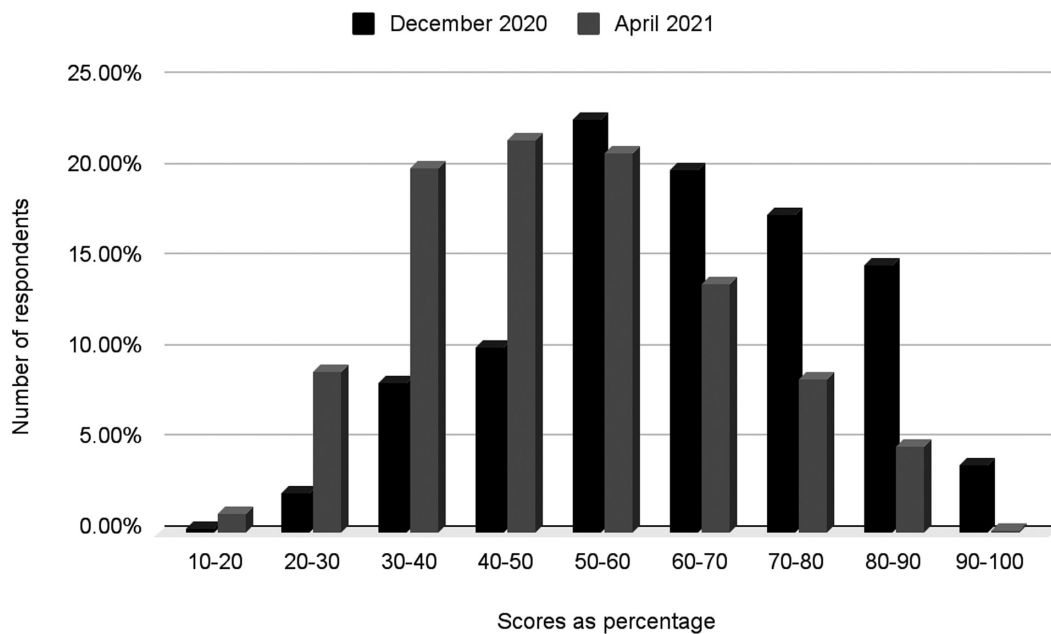
The April 2021 Placement Test

The placement test conducted in April of 2021 was to place in-coming first year students in the appropriate classes based on their language proficiency. Because the vast majority of first year students plan to study English as their second language, the number of students who took the test was 729 (N=729).

Figure 3 shows a comparison of the score distribution of the December 2020 and April 2021 tests. It is interesting to note that the distribution of scores for the December 2020 test skewed higher than the April 2021 test. There are several reasons that a difference like this may occur. One possible reason could be the smaller sample size for the December test. Only those students with an interest in English would have participated in that test, while the April test would have included students who may not have been as motivated to study English. Another reason may have been the delivery method. Students who took the December test had already been studying online since the beginning of the school year. They had already most likely taken tests or quizzes in an online format and would have been familiar with the procedure and method. The same cannot be said for the students who took the April 2021 test. Very few students in this group would have studied online or taken online tests in the school year prior to entering university because most high schools had continued to conduct face to face lessons during the 2020-2021 academic year.

Figure 3

Score Distribution of the December 2020 and April 2021 Placement Tests



Challenges Collaborating Online

Due to COVID-19 safety recommendations the entire process of creating the placement test was conducted online. Collaborating online had several challenges but also some benefits. First, all meetings to discuss the placement test had to be conducted by video conferencing. This created a challenge in communication especially in regard to sharing material necessary for the making of the test since we were unable to share physical materials such as books through video conferencing.

Another challenge when creating an online test was finding appropriate listening materials that were not only at the correct level, but also could be used without copyright infringement. Finding those materials for the listening section proved to be a time-consuming task. Ultimately, one track that was licensed under creative commons to be used for educational purposes was decided upon based on the length and content. However, the second listening section, a conversation, needed to be created and recorded in house. With no budget allocated to hire professional voice actors, two of the test creators, a male and female, decided to record the audio track and design the test items themselves. As both speakers were comfortable with speaking together online for the recording, the initial recording of the track was not overly complicated. However, post recording and editing the track for inclusion in the test proved to be the more complicated part of designing the listening section.

However, there were also benefits to collaborating online. The convenience of online Google Forms helped create the test faster and in fewer meetings. Because Google Forms allows different editors to work on the same document at different times and instantly updates changes made to the document, the authors could work at their most convenient time while at the same time keeping track of the changes made to the test.

Online Placement Test Limitations and Future Research

Our study found two main limitations to online placement tests. First, inequalities in results may arise because of the conditions of the internet environment or the device test takers use during the test. In our test, we tried to limit the influence of these factors as much as possible by allowing test takers to take the test at different times during the day and by utilizing only multiple-choice questions which are easier to visualize even in devices with small screens. Second, arguably the most significant limitation of online placement tests, is the difficulty to fully assess test takers' communicative competence. Designing test questions that measure language production, such as speaking or writing skills, in a completely online environment requires a high level of technological complexity and labour to evaluate the answers. In our test, due to the large number of test takers, and the fact that the test was fully designed in Google Forms, the focus was on language recognition and comprehension of test takers.

We would also like to measure test validity and reliability in the future. We hope that it will be possible to run a test-retest reliability measure if we can get the same group of students to take the test more than once. It may also be possible to complete an alternate form reliability measure by comparing student scores from a similar test.

References

- Brady, A. M. (2005) . Assessment of learning with multiple-choice questions. *Nurse Education in Practice*, 5(4), 238-242. <https://doi.org/10.1016/j.nepr.2004.12.005>
- Browne, C., Culligan, B. & Phillips, J. (2013). *The New General Service List* [Data set]. Retrieved from <http://www.newgeneralservicelist.org>
- English Profile (2015). *English Vocabulary Profile Online - American English* [Data set]. Cambridge University Press. Retrieved from <https://www.englishprofile.org/american-english>
- Grijalva, T. C., Kerkvliet, J., & Nowell, C. (2006). Academic honesty and online courses. *College Student Journal*, 40(1), 1-18.
- Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020, March). The difference between emergency remote teaching and online learning. *Educause Review*. Retrieved from <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-onlinelearning>
- Irwin, B. (2021). Emergency remote teaching - students' expectations and perceptions of online classes. *Studies in International Relations* 41, 113-120.
- Kunnan, A. J. (2004). Test fairness. *European language testing in a global context*, 27-48.
- Ladyshevsky, R. K. (2015). Post-graduate student performance in 'supervised in-class' vs.'unsupervised online' multiple choice tests : implications for cheating and test security. *Assessment & Evaluation in Higher Education*, 40(7), 883-897.
- Milman, N. B. (2020, March 30). This is emergency remote teaching, not just online teaching. *Education Week*. Retrieved from <https://www.edweek.org/ew/articles/2020/03/30/this-is-emergency-remoteteaching-not-just.html>
- Munoz, A., & Mackay, J. (2019) An online testing design choice typology towards cheating threat minimisation. *Journal of University Teaching & Learning Practice*, 16(3), 5.
- OECD (2020). Remote online exams in higher education during the COVID-19 crisis. *OECD Education Policy Perspectives* (No. 6). <https://doi.org/10.1787/f53e2177-en>.
- Peled, Y., Eshet, Y., Barczyk, C., & Grinautski, K. (2019). Predictors of academic dishonesty among undergraduate students in online and face-to-face courses. *Computers and Education*, 131, 49-59.
- Peterson, J. (2019). An analysis of academic dishonesty in online classes. *Mid-Western Educational Researcher*, 31(1), 24-36.
- Xu, X., Kauer, S., & Tupy, S. (2016). Multiple-choice questions : Tips for optimizing assessment in-seat and online. *Scholarship of Teaching and Learning in Psychology*, 2(2), 147-158. <https://doi.org/10.1037/stl0000062>

