MICHIGAN LANGUAGE ASSESSMENT

MET

MET 2024 Test Report

Contact Information

All correspondence and mailings should be addressed to:

Michigan Language Assessment
Argus 1 building
535 West William St., Suite 310
Ann Arbor, Michigan
48103-4978 USA

T +1 866.696.3522
T +1 734.615.9629
F +1 734.763.0369

info@michiganassessment.org
MichiganAssessment.org

Table of Contents

1. Description of the Test

1.1 General Description

The Michigan English Test (MET) is a standardized, multilevel examination of general English language proficiency. Developed and produced by Michigan Language Assessment, the test covers the four language skills: listening, reading, writing, and speaking.

The Listening and Reading Sections measure listening, reading, grammar, and vocabulary skills in educational, public, and occupational contexts, with recordings and reading passages that reflect interactions in an American-English linguistic environment. The Writing Section measures an individual's ability to write in English in response to two different tasks, and the Speaking Section measures an individual's ability to produce comprehensible speech in response to a range of tasks and topics.

MET covers a range of proficiency levels from upper beginner to lower advanced: the A2 to C1 levels of the Common European Framework of Reference (CEFR; Council of Europe, 2001, 2020).

1.2 Intended Uses

MET is intended for English learners aged 16 or older at or above a secondary level of education who want to measure their general English language proficiency in a variety of linguistic contexts. The test is intended to support high-stakes decisions for higher education, immigration, and employment, and validation research focuses on these decisions. MET is accepted by a large number of colleges, universities, governments, professional organizations, and employers. For example, MET scores are used to support admissions decisions for universities that require evidence of English proficiency. MET is also accepted by the U.S. government and most boards of nursing from individual states to certify English proficiency for foreign-trained nurses wishing to obtain an occupational visa and nursing license to work in the U.S. Employers requiring evidence of English proficiency also use MET scores for this purpose. Organizations with questions about appropriate uses of MET scores should contact Michigan Language Assessment for advice on the validity of score use in their contexts.

1.3 Test Format

MET measures listening, reading, grammar, vocabulary, writing, and speaking skills. All sections are administered digitally. The Listening and Reading questions are multiple choice and have one correct answer. The Speaking and Writing Sections require test takers to produce spoken or written responses that are scored by trained and certified raters. Using Linear On the Fly Testing (LOFT) technology, unique digital forms are generated for each test taker. These forms pull from large item pools that are replaced periodically to ensure that questions are not overexposed. The specifications for these unique digital forms are the same for all test takers, so that even though the questions are different, the number and type of questions is consistent. Table 1 (next page) describes the format and content of MET.

1.4 Development and Administration

Michigan Language Assessment is committed to the excellence of its tests, which are developed in accordance with the highest standards in educational measurement. All parts of the examination are written following specified guidelines, and items are pretested to ensure that they function properly. MET is administered digitally and on demand, either in test centers or via a remote-proctored delivery platform. Michigan Language Assessment works closely with test centers and delivery partners to ensure that its tests are administered following rigorous security protocols and consistent procedures, that the test is fair and accessible to test takers, and that MET is open to all people who wish to take the exam. Test preparation resources are available on the Michigan Language Assessment website.

Table 1: Format and Content of MET

Section	Time	Description	Number of Items
Listening	35 minutes	**Part 1:** Short conversations are each followed by a question.	19 questions
		Part 2: Longer conversations between two people are each followed by several questions.	14 questions
		Part 3: Short talks are delivered by a single speaker and followed by several questions.	17 questions
Reading	65 minutes	**Grammar:** An incomplete sentence is followed by a choice of four words or phrases to complete it.	20 questions
		Single-text reading: Two extended reading passages are each followed by five questions.	10 questions
		Multiple-text reading: Two sets of three thematically linked passages are each followed by ten questions.	20 questions
Writing	45 minutes	Tasks require test takers to produce written language at the sentence, paragraph, and essay levels.	2 tasks
Speaking	10 minutes	Test takers respond to recorded prompts to complete structured speaking tasks that increase in difficulty and complexity.	5 tasks

2. Scoring and Reporting of Results

2.1 Explanation of Scoring

The MET Listening and Reading Sections are scored by computer at Michigan Language Assessment. Each correct answer contributes to the final score for each section, and there are no points deducted for wrong answers. A scaled score, ranging from 0 to 80, is calculated using Item Response Theory. This method ensures that scores are comparable across different LOFT forms and administrations, and that a given score represents the same level of ability each time.

The Writing and Speaking Sections are graded by human raters according to scales established by Michigan Language Assessment (see our website for the rating scales). All raters are trained and certified by Michigan Language Assessment specifically for the particular skill and rating scale.

2.2 Procedures for Reporting Scores

All test takers receive a scaled score from 0-80 for each test section, and an overall average score for all sections taken. The scores are also reported as CEFR levels. Table 2 shows the MET scaled scores that correspond to these CEFR levels. These correspondences are based on standard setting research conducted by Michigan Language Assessment (Papageorgiou, 2010; Michigan Language Assessment, 2014, Michigan Language Assessment, 2025).

3. Interpreting MET Results

MET is a multilevel exam, covering a range of proficiency levels from A2 to C1 on the CEFR. Selected CEFR performance descriptors illustrating what test takers should be able to do at each level are available from the Michigan Language Assessment website.

When interpreting MET results, it is important to remember that MET estimates a test taker's true proficiency by approximating the kinds of tasks that may be encountered in real life. Also, temporary factors unrelated to a test taker's proficiency, such as fatigue, anxiety, or illness, may affect exam results.

When using test scores for decision-making, consider whether some sections are more important for the decision than others; if they are, consider those sections rather than relying solely on the overall average score. It is possible for a test taker to be at a higher language proficiency level in one language skill than in another. Therefore, the section scores may provide more relevant information than the overall score when interpreting the test results for use in decision-making. Additionally, check

Table 2: CEFR Level Equivalence of MET Scaled Scores

CEFR Level	Scaled Score
C1	64 – 80
B2	53 – 63
B1	40 – 52
A2	27 – 39
Below A2	0 – 26

the date the test was taken. Language ability can change over time, but Michigan Language Assessment acknowledges that different score users may have different concerns and needs for the recency of results. All past score reports are maintained by Michigan Language Assessment, enabling score users to determine their own requirements for the interval between test administration and score use.

It is also important to remember that test performance is only one aspect to be considered. Communicative language ability consists of both knowledge of language and knowledge of the world. Therefore, it is important to consider the factors other than language that may affect how

Table 3: List of First Language Backgrounds

Note: MET test takers may list English as their first language if they are from countries where English is an official language; however, for many such countries, such as the Philippines or Nigeria, people from these countries wishing to study in or immigrate to another country may still be required to prove their English proficiency.

Afrikaans	German	Punjabi
Albanian	Greek	Romanian
American Sign Language	Gujarati	Russian
Amharic	Hausa	Shona
Arabic	Hebrew	Sinhalese
Armenian	Hindi	Slovak
Bambara/Malinke	Hungarian	Somali
Bengali	Ibo (Igbo)	Spanish
Burmese/Chin	Italian	Swahili
Cambodian	Japanese	Swedish
Catalan	Kikuyu	Tagalog/Filipino
Chinese (Cantonese/Mandarin)	Korean	Thai
Creole	Kurdish	Tigre/Tigrinya
Croatian	Malayalam	Turkish
Dutch	Marathi	Ukrainian
English	Nepali	Urdu
Ewe	Norwegian	Vietnamese
Farsi/Persian	Polish	Yoruba
French	Portuguese	

well someone can communicate. For example, in the general context of using English in the workplace, the ability to function effectively involves not only knowledge of English, but also other knowledge and skills specific to the work.

4. Test-Taking Population

This section presents an overview of the test takers who took MET in 2024, providing demographic information for the test population. Every test taker was asked to complete a registration form, which asked for their gender, date of birth, and first language, as well as their purpose for taking the test and their educational level. Cases where information was not given or was not correctly given were treated as missing data.

Table 3 (previous page) lists the first-language backgrounds of the test takers, ordered alphabetically. The test takers represented 60 different first language backgrounds, but it should be noted that the test population primarily consisted of test takers whose first language was Spanish, Tagalog/Filipino, or Portuguese.

Tables 4 and 5 present the distribution of test takers by age and gender, respectively. Information about the purpose for taking the test and about the test takers' level of education is provided in tables 6 and 7. Most test takers are between 16 and 39 years old and taking the test for education-related or employment-related purposes.

Table 4: Distribution (in %) of MET Test Takers by Age

Age Range	% of Test Takers
≤12	0.25
13–15	5.37
16–19	15.89
20–22	12.47
23–25	12.18
26–29	11.72
30–39	25.88
≥ 40	15.04
Missing	1.20

Table 5: Distribution (in %) of MET Test Takers by Gender

Gender	% of Test Takers
Male	33.44
Female	53.44
Other	1.73
Missing	11.39

Table 6: Distribution (in %) of MET Test Takers by Purpose for Taking the Test

Purpose for Taking the Test	% of Test Takers
Education Program Admissions	9.00
Language Course Requirement	12.30
Scholarship	4.08
Nursing Credentials	15.03
Obtain Employment	9.92
Improve Employment	10.60
Personal Interest	16.11
Other	9.65
Missing	13.31

Table 7: Distribution (in %) of MET Test Takers by Level of Education

Level of Education	% of Test Takers
Primary	0.83
Lower Secondary	2.82
Upper Secondary	21.17
Undergraduate	30.46
Postgraduate	31.41
Missing	13.31

5. Test Statistics

5.1 Distribution of Results by CEFR Level

Table 8 lists the percentage of test taker scores by CEFR level. Most test takers are at levels B1 and B2.

Table 8: Distribution (in %) MET Test Takers by CEFR Level

Section	Below A2	A2	B1	B2	C1
Listening	2.96	3.52	30.39	39.52	23.61
Reading	3.76	6.97	34.55	35.44	19.28
Writing	1.77	4.38	34.88	41.65	17.33
Speaking	4.08	14.34	36.46	30.94	14.17
Average	2.05	7.75	35.11	38.94	16.15

5.2 Reliability Figures for Listening and Reading

Test scores are a numerical measure of a test taker's ability. *Reliability* refers to the consistency of the measurement. In theory, a test taker's test score should be the same each time the test is taken or across different forms of the same test. In practice, even when the test conditions are carefully controlled, an individual's performance on a set of test items will vary from one administration to another due to variation in the items across different forms of the same test or due to variability in individual performance. Among the reasons for this are temporary factors unrelated to a test taker's proficiency, such as fatigue, anxiety, or illness. As a result, test scores always contain a small amount of measurement error. The aim, however, is to keep this error to a minimum. For high-stakes exams such as MET, a reliability figure of 0.80 and above is expected and acceptable. In addition to monitoring reliability, the estimated variability in test taker performance is also monitored through the standard error of measurement (SEM) estimate.

Reliability and SEM estimates are obtained for the LOFT administrations over the course of the year. The reliability estimates are calculated by taking the average of the exam's conditional reliabilities at each available score point. The SEM estimates are calculated using the reliability estimates and the scaled scores. In this report, the reliability and SEM estimates are summarized as averages across the different individual MET LOFT administrations.

For the Listening Section, the average reliability estimate was 0.80, and the average SEM estimate was 5.52. For the Reading Section, the average reliability estimate was 0.81, and the average SEM estimate was 5.57. These values demonstrate that the reliability figures for both exam sections are above the minimally acceptable value of 0.80. Additionally, the SEM estimates as a proportion of the 80-point scale are reasonable. These values suggest good consistency of measurement for the MET Listening and Reading Sections.

5.3 Writing and Speaking Reliability and Rater Performance

The raters for the Writing and Speaking Sections are highly proficient speakers of English who are trained and certified according to standards set by Michigan Language Assessment. Rater performance is monitored through a routine audit process that double rates a set proportion of all Writing and Speaking Section performances.

The reliability of the Writing and Speaking scores is established by analyzing the composite scores awarded to the same test taker by two different raters randomly selected as a sample of performances. As shown in Table 9 (next page), MET raters demonstrate a very high degree of agreement and consistency as measured by the Intraclass Correlation Coefficient (ICC). Typically, ICC in the 0.7-0.9 range is considered indicative of a "very high" level of agreement and consistency among the raters, while ICC equal to or greater than 0.9 indicates is "excellent."

The SEM of the Writing and Speaking Sections are calculated using the ICC for Consistency reliability estimate and the scaled scores. The Writing Section had an SEM estimate of 4.25, and the Speaking Section had an SEM estimate of 4.70. These SEM estimates as a proportion of the 80-point scale are reasonable.

The results of these analyses for the Speaking and Writing Sections of MET provide evidence for high reliability of the raters' judgements. The high reliability of rating performance of MET raters assures that the scores produced by the judges are consistent and can be replicated with a high degree of reliability. These values suggest good consistency of measurement for the MET Writing and Speaking Sections.

Table 9: Intraclass Correlation Coefficients for Double-Rated Sample

Section	ICC for Agreement*			ICC for Consistency**		
	Value	Lower Bound	Upper Bound	Value	Lower Bound	Upper Bound
Writing	0.83	0.79	0.86	0.84	0.81	0.86
Speaking	0.83	0.80	0.85	0.83	0.81	0.85

* ICC values for Agreement obtained using a two-way random effects model for absolute agreement for a single rater/measure.

** ICC values for Consistency obtained using a two-way mixed effects model for absolute agreement for a single rater/measure.

6. References

Council of Europe (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Cambridge University Press.

Council of Europe (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion volume. Council of Europe Publishing, Strasbourg.

Michigan Language Assessment (2014). Linking the Common European Framework of Reference and the MET Writing Test, Michigan Language Assessment Technical Report, Michigan Language Assessment. Retrieved from https://michiganassessment.org/wp-content/uploads/2020/02/20.02.pdf.Res_.LinkingtheCommonEuropeanFrameworkofReferenceandtheMETWritingTest.pdf

Michigan Language Assessment (2025). Setting MET Speaking cut scores in alignment with the CEFR. Michigan Language Assessment Technical Report, Michigan Language Assessment.

Papageorgiou, S. (2010). Setting cut scores on the Common European Framework of Reference for the Michigan English Test, Michigan Language Assessment Technical Report, Michigan Language Assessment. Retrieved from https://michiganassessment.org/wp-content/uploads/2020/07/20.02.MET-Setting-Cut-Scores-CEFR.pdf