



MICHIGAN LANGUAGE ASSESSMENT



MET

MET 2023 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
1.1 General Description	1
1.2 Intended Uses	1
1.3 Test Format	1
1.4 Development and Administration	1
2. Scoring and Reporting of Results	2
2.1 Explanation of Scoring	2
2.2 Procedures for Reporting Scores	2
3. Interpreting MET Results	3
4. Test-Taking Population	4
5. Test Statistics	4
5.1 Distribution of Results by CEFR Level	4
5.2 Reliability Figures for Listening and Reading	4
5.3 Writing and Speaking Rater Performance	5
6. References	6

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. MET is accepted by a large number of colleges, universities, governments, professional organizations, and employers for high-stakes decisions. In particular, international students use MET to certify English proficiency and apply to universities in the U.S., Canada, and other English-speaking countries. It is also accepted by the U.S. government and a large number of boards of nursing from individual states to certify English proficiency for foreign-trained nurses interested in obtaining an occupational visa and nursing license to work in the U.S.

1.3 Test Format

MET measures listening, reading, grammar, vocabulary, writing, and speaking skills. It is administered digitally as a 2-skill test that consists of Listening and Reading Sections or as a 4-skill test that additionally includes Writing and Speaking Sections. 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. There is a paper-based version available in limited circumstances as a 2-skill test consisting of the Listening and Reading sections, which follows the same specifications for number and type of questions as the digital version. 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. 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. The limited 2-skill paper-based version is administered four times a year, also following strict administration specifications. 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 a question.	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
		Multiple-text reading: Two sets of three thematically linked passages are each followed by ten questions.	20 questions
		Single-text reading: Two extended reading passages are each followed by five questions.	10 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 participate in a structured, multistage task.	5 stages

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 and paper forms and administrations, and that the ability required to receive a score remains the same from year to year.

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).

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

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

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 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.

Table 3: List of First Language Backgrounds

Afrikaans	Greek	Romanian
Albanian	Hebrew	Russian
American Sign Language	Hindi	Slovak
Amharic	Italian	Spanish
Arabic	Kannada	Swahili
Bengali	Korean	Swedish
Bulgarian	Kurdish	Tagalog/Filipino
Burmese/Chin	Latvian	Tamil
Catalan	Luo (Lwo)	Telegu
Chinese (Cantonese/Mandarin)	Macedonian	Thai
Creole	Malagasy	Tibetan
Croatian	Malayalam	Tigre/Tigrinya
Czech	Nepali	Turkish
Farsi/Persian	Polish	Ukrainian
French	Portuguese	Urdu
German	Punjabi	Vietnamese

4. Test-Taking Population

This section presents an overview of the test takers who took MET in 2023, 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 (native) language. Cases where information was not given or was not correctly given were treated as missing data. Note that the data encompass all formats and uses of the test; population differences that may exist between users of the paper and digital versions are not captured in all analyses.

Table 3 lists the first language backgrounds of the test takers. The test takers represented 48 different first language backgrounds, but it should be noted that the test population primarily consisted of test takers whose first language was Greek, Portuguese, or Spanish. Tables 4 and 5 present the distribution of test takers by age and gender, respectively. Table 4 shows that the majority of MET paper test takers were 16 - 19 years old (58.65%), while the digital test takers had an overall wider age distribution.

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

Age Band	2-Skill (Paper)	2-Skill (Digital)	4-Skill (Digital)
≤12	0.20	0.00	0.11
13 - 15	20.90	3.21	6.26
16 - 19	58.65	16.51	20.06
20 - 22	10.48	18.25	15.37
23 - 25	3.92	21.37	15.04
26 - 29	2.49	11.82	10.77
30 - 39	2.05	18.25	18.66
≥40	1.21	10.34	12.45
Missing	0.10	0.26	1.28

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

Gender	% of Test Population
Male	37.54
Female	48.62
Prefer to self-describe	0.13
Prefer not to say	1.58
Missing Data	12.13

5. Test Statistics

5.1 Distribution of Results by CEFR Level

Tables 6 and 7 list the percentage of test taker scores by CEFR level for the 2- and 4-skill versions of MET, respectively.

Table 6: Distribution (in %) of 2-Skill MET Test Takers by CEFR Level

Section	Below A2	A2	B1	B2	C1
Listening	2.46	4.38	36.33	37.40	19.43
Reading	3.02	7.03	31.83	34.90	23.22
Average	1.30	8.73	33.09	36.14	20.74

Table 7: Distribution (in %) of 4-Skill MET Test Takers by CEFR Level

Section	Below A2	A2	B1	B2	C1
Listening	4.95	5.51	32.11	35.89	21.54
Reading	6.97	10.50	34.54	31.69	16.30
Writing	3.72	6.86	34.48	37.99	16.96
Speaking	6.44	18.03	33.18	25.70	16.66
Average	4.16	11.94	33.83	33.66	16.41

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 each administration of MET. The statistical methodologies are as follows: For the paper-based MET, reliability estimates are calculated in Winsteps using the KR-20 (Kuder-Richardson Formula 20) method. The SEM estimates are then calculated using the reliability estimates and the scaled scores. For the digital MET, reliability estimates are calculated by taking the average of the digital exam's conditional reliabilities at each available score point. In this report, the reliability and SEM estimates are summarized as averages across the different MET administrations.

For the Listening Section, the average reliability estimate was 0.88, and the average SEM estimate was 3.90. For the Reading Section, the average reliability estimate was 0.86, and the average SEM estimate was 4.67. 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 very small. These values suggest good consistency of measurement for the MET Listening and Reading Sections.

5.3 Writing and Speaking 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 demonstrated in Table 8, MET raters demonstrate a very high degree of agreement and consistency as measured by the Intraclass Correlation Coefficient (ICC): 0.90 for Speaking, and 0.86 for Writing. 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 is “excellent.” The results of the reliability analyses for the Speaking and Writing Sections of MET provide evidence for high reliability of the raters’ judgements. 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.

Table 8: 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	0.86	0.83	0.89	0.86	0.83	0.89
Speaking	0.90	0.88	0.92	0.90	0.88	0.92

* 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/2014/12/MET-Writing-Linking-to-CEFR-20141106.pdf>

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/2014/12/MET_Standard_Setting.pdf