



MICHIGAN LANGUAGE ASSESSMENT

MET Go!

Development of the MET Go! Writing Test
Technical 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



© 2018 Cambridge Michigan Language Assessment

**MICHIGAN
LANGUAGE
ASSESSMENT**



Cambridge Assessment
English



UNIVERSITY OF MICHIGAN

TABLE OF CONTENTS

1. Introduction	1
2. Test Construct	1
2.1. Targeted CEFR Levels	1
2.2. Construct Definition	1
3. Task Development.....	3
3.1. Task Theory.....	3
3.2. Task Design.....	3
4. Rating Tool Development	4
4.1. Rating Scale Theory and Target Language Features.....	4
4.2. Rating Tool Design	5
4.3. Piloting the Checklist Scale	5
5. Interpreting Writing Test Scores	6
6. References	7

LIST OF TABLES

Table 1:	Overall Written Production.....	1
Table 2:	MET Go! Writing Test Tasks, CEFR Levels Targeted, and Linguistic Functions	4
Table 3:	Evaluation Criteria for the MET Go! Writing Test	5

LIST OF FIGURES

Figure 1:	Construct definition of the MET Go! Writing Test	2
Figure 2:	Components of the language knowledge and fundamental processes measured by the MET Go! Writing Test.....	3

1. INTRODUCTION

MET Go! is a multi-level test of English language ability designed for beginner to intermediate level learners of middle and secondary school age. Developed and produced by Michigan Language Assessment, the test covers the four language skills (listening, reading, speaking, and writing), assessing learners' ability in each area and assisting them as they progress in their learning.

The MET Go! Writing Test is designed to assess test takers' written English proficiency by evaluating their ability to perform a range of communicative functions in writing—narrating stories, describing personal experiences, and expressing and supporting opinions—on a variety of familiar school and everyday topics. It is assessed by a Michigan Language Assessment-certified writing rater and scored using a fit-for-purpose rating tool. The MET Go! Writing Test is intended to be useful in a variety of educational settings. The results can be used to monitor the progress of English as a Second Language (ESL) or English as a Foreign Language (EFL) learners, as well as for placement or diagnostic purposes to inform instructors of the strengths and weakness of the learners and areas where instruction is needed. Language programs can also use the test to certify whether or not learners have achieved the goals of a language course.

This report describes the development of the MET Go! Writing Test. It provides information on the development of the test construct, task types, and rating tool, as well as information on score interpretation.

2. TEST CONSTRUCT

2.1. Targeted CEFR Levels

The Common European Framework of Reference (CEFR) provides a common basis for evaluating the ability level of language learners. The framework identifies six broad levels of language ability, and offers illustrative scales and can-do statements that describe “what language learners have to learn to do in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act effectively” (Council of Europe, 2001, p. 1).

The MET Go! Writing Test targets writing at the A1-B1 levels of the CEFR. Both the CEFR (Council of Europe, 2001) and the CEFR companion volume (Council of Europe, 2018) were used by the MET

Go! Writing Test development team throughout the development process as references to inform the design of the test construct, task types, and rating tool.

The can-do statements from numerous CEFR scales were heavily referenced during development. These scales included the overall written production, creative writing, correspondence, coherence and cohesion, reports and essays, thematic development, grammatical accuracy, vocabulary control, vocabulary range, and orthographic control scales (Council of Europe, 2001, 2018). Table 1 summarizes the progression in overall written production from levels A1 to B1 for learners aged 11 – 15 (Council of Europe, 2001, 2016). As learners progress through each CEFR level they are expected to have mastered abilities described under lower levels of competence. The table shows that A1 level test-takers are able to write simple words or expressions to give information on personal topics. More proficient test-takers are able to write on an increasing range of topics using increasingly complex language (Council of Europe, 2001).

Table 1: Overall Written Production (Council of Europe, 2001, 2016)

CEFR Level	Descriptor
B1	Can write straightforward connected texts on a range of familiar subjects within his field of interest, by linking a series of shorter discrete elements into a linear sequence.
A2	Can write a series of simple phrases and sentences linked with simple connectors like 'and,' 'but' and 'because'.
A1	Can give information in writing about matters of personal relevance (e.g. likes and dislikes, family, pets) using simple words and basic expressions.

2.2. Construct Definition

The MET Go! Writing Test adopts the interactionalist approach to construct definition which considers performance as the results of traits, contextual features, and their interaction, and therefore, views performance as “a sign of underlying traits, and is influenced by the context in which it occurs, and is therefore a sample of performance in similar contexts” (Chapelle, 1998, p. 43). This perspective takes into account the role of contextual factors while allowing for

the generalization of test scores beyond the immediate testing instance. The test construct definition is illustrated in Figure 1 below.

The construct the MET Go! Writing Test aims to assess is defined as test-takers’ ability to communicate in written English across a range of situations beginner to intermediate level learners of middle and secondary school age might encounter in the course of routine daily/school life. The language knowledge that the test aims to measure is specified in previous applied linguistics research on components of language ability. Specifically, the three main components of language knowledge measured in this test—namely grammatical knowledge, textual knowledge, and sociolinguistic knowledge—are based on Bachman’s and Palmer’s (1996) framework of language ability. Grammatical and lexical knowledge involves knowledge of syntax and vocabulary to produce formally accurate sentences. Textual knowledge refers to the ability to produce explicitly marked relationships among sentences in written texts (knowledge of cohesion) and to produce well-developed and organized written texts (knowledge of rhetorical organization). Lastly, pragmatic knowledge, including functional knowledge and sociolinguistic knowledge, allows learners to create

written texts appropriate to a particular language use setting, i.e., to respond appropriately when asked to narrate a story, describe their personal experience, and state and support their opinion or preference following some writing conventions such as correct spelling, punctuation, and capitalization. Being aimed at the A1 to B1 levels, language tasks that require knowledge telling rather than knowledge transforming are employed (Scardamalia & Bereiter, 1987). Figure 2 summarizes the types of language knowledge and fundamental processes the test aims to measure.

In the course of responding to the test tasks, learners also need to use their strategic competence, defined as “a set of metacognitive components, or strategies, which can be thought of as higher order executive processes that provide cognitive activities” (Bachman & Palmer, 1996, p. 40). These strategies allow test-takers to assess the situation, decide how to respond to a question (i.e., goal setting), and decide the types of language knowledge and background knowledge to use to achieve that goal (i.e., planning).

In summary, the MET Go! Writing Test is intended to measure the ability of 11 – 15 year old test-takers to write cohesive, simply connected English to accomplish tasks that require knowledge-telling writing, such as

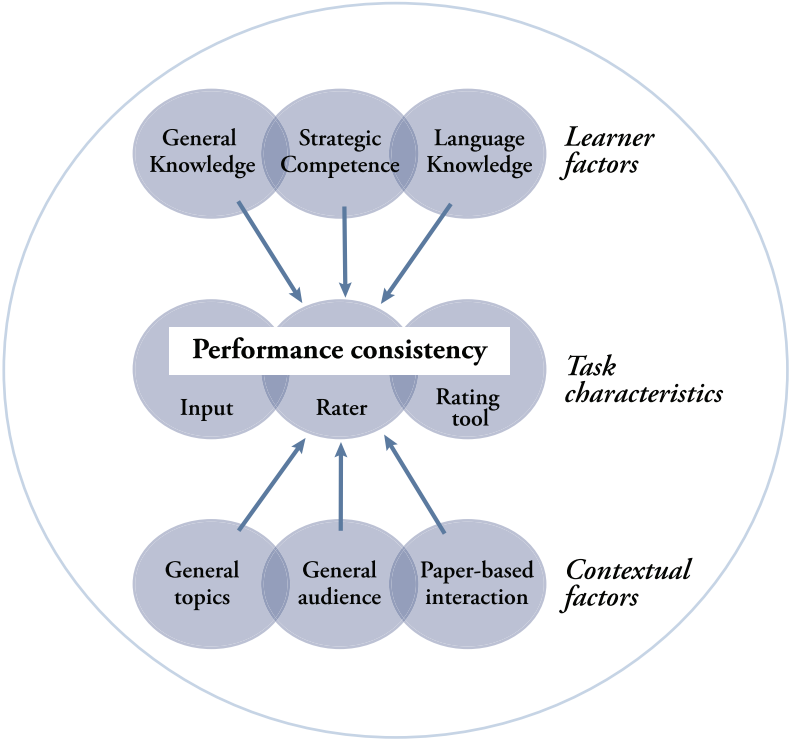


Figure 1: Construct definition of the MET Go! Writing Test

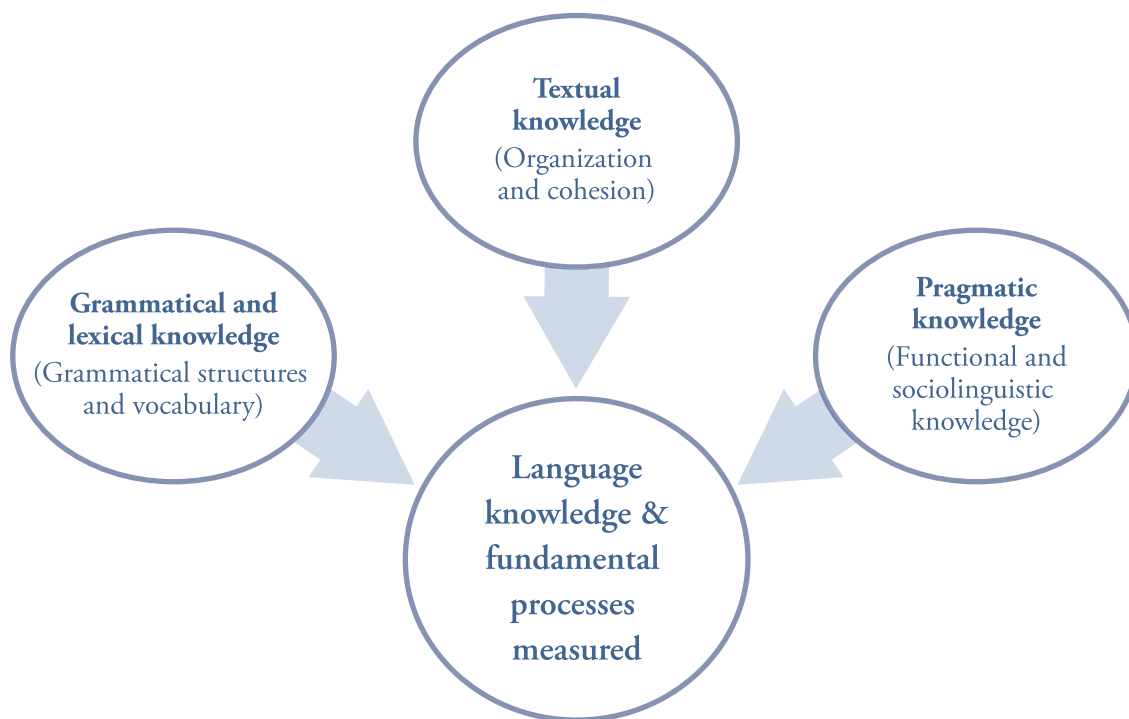


Figure 2: Components of the language knowledge and fundamental processes measured by the MET Go! Writing Test

short descriptions, narratives, and explanatory texts based on opinions and personal experience with familiar topics.

3. TASK DEVELOPMENT

3.1. Task Theory

The writing test tasks are designed to elicit written language representing a range of ability levels from beginner to low intermediate (CEFR levels A1-B1). Descriptors of these levels determined the linguistic functions that would be elicited in the test. For example, while an A1 level learner can produce phrases or sentences in response to a task, this learner cannot yet connect their ideas to create a more cohesive text (Council of Europe, 2001). An A2 level learner can produce a series of phrases linked by simple connectors, while a B2 level learner is able to produce straightforward connected texts (Council of Europe, 2001).

With beginning language learners, test developers need to consider the extent to which young learners' skills in other areas could impinge on their ability to complete the task (McKay, 2006). Thus, the use of visual prompts may be appropriate, so that reading

ability does not become a factor in the elicitation of an appropriate writing sample. The use of a series of visuals has also been shown to be valuable as elicitation prompts for observing learners' narrative skills and their natural language use within and between sentences, with the linearity of events across visuals providing support and making it easier for test-takers to formulate a thread of meaning on the basis of the images and spark their imaginations to build a narrative (Bae & Lee, 2010). The amount and type of support given in the prompt can then be reduced or changed to elicit writing samples from learners who are further along in their learning.

3.2. Task Design

The MET Go! Writing Test is a paper-based test of written production. The test consists of three tasks accessible to both lower- and higher-level test-takers. Table 2 describes the purpose of each test task, the CEFR level targeted, and the corresponding linguistic functions.

Task 1, aimed at beginner and low-intermediate speakers (A1 and A2 on the CEFR), requires test-takers to give a description of a sequence of concrete, familiar events or narrate a story. This task is presented with a

Table 2: MET Go! Writing Test Tasks, CEFR Levels Targeted, and Linguistic Functions

Task	Description	Level(s) targeted	Linguistic functions
Task 1	Comic Strip Description/Narration	A1 – A2	Describe or narrate a sequence of concrete, familiar events
Task 2	Description of Personal Experience	A2 – B1	Describe an experience and elaborate with some detailed descriptions of events
Task 3	Express and Explain Preferences	A2 – B1	Express and explain a preference toward a familiar topic of personal interest

series of three visuals to set context for the narrative and support test-takers in terms of ideas and avoid too much written text which could be inaccessible to learners at lower levels. Task 2 targets more able writers (A2 and B1 on the CEFR) who are asked to describe a personal experience with supporting details. In this task, bulleted main points about the topic are provided to help test-takers plan their response. Task 3 also aims at more able writers (A2 and B1 on the CEFR) who are asked to present their opinion or preference on a topic and provide supporting details to explain their opinion or preference.

4. RATING TOOL DEVELOPMENT

4.1. Rating Scale Theory and Target Language Features

Tests need to have scoring criteria that cover the test construct, and a scoring approach that is fit for purpose. The scoring criteria “provides an operational definition of a linguistic construct” (Davies et al., 1999, p. 153) and represents the test developers’ view of the construct (Weigle, 2002; Knoch, 2009). For the MET Go! Writing Test, the work initially employed theoretical and intuitive methods (Council of Europe, 2001; Fulcher, 2003; Knoch, 2009).

A committee of experts determined the wording of the descriptors and levels in the rating tool. The level descriptors for each criterion were designed to be brief, clear, concrete, and detailed enough (with the absence of field-specific jargon) to sufficiently guide raters from varying backgrounds to rate writing responses consistently, and also allow them to make quick scoring decisions. Word count and length of each level’s performance descriptors were also considered, as the descriptors have to be concrete yet practical to be useful for raters (Luoma, 2004, p. 81).

Conceptualizations of writing ability, as narrowed down and instantiated in the construct, as well as the CEFR, (Bachman & Palmer, 1996) were the basis for identifying the criteria to be applied. Four groups of evaluation criteria emerged: Task Completion, Grammar, Vocabulary, and Mechanics.

Task Completion refers to the degree to which the test-taker addresses the task presented in the prompt, that is, the relevance of the response to the task. This criterion also focuses on cohesion and the extent to which the written text is organized in a logical order. At lower levels indicators of cohesion include the presence of simple connectors like “and”, “but”, and “because” (Council of Europe, 2001, p. 29). At higher levels, more complex discourse connectives are employed.

Grammar refers to the ability of test-takers to use grammatical resources of a language. It includes grammatical accuracy, complexity, and the ability to use correct verb tenses to convey ideas. According to Savignon (2005), the findings of studies conducted on learners’ language proficiency and task performance have shown that learners of higher proficiency levels have higher awareness of grammatical accuracy and demonstrate more accurate language performance. Generally, there is evidence that the error rate decreases as learners’ proficiency advances (Alexopoulou, Michel, Murakami, & Meurers, 2017). According to the Council of Europe (2001), a B1 learner can communicate clearly in familiar contexts with reasonable accuracy and generally good control of the language despite first-language related errors, while an A2 learner can use some simple structures correctly but still makes basic mistakes related to verb tenses or subject-verb agreement.

Vocabulary encompasses complexity of range and accuracy, and refers to how test-takers use their lexical resources to convey meaning. Previous research has found a relationship between proficiency level and

lexical diversity (Alexopoulou et al., 2017; Yu, 2009). Specifically, the measure of textual lexical diversity increases steadily from the very early levels to B1 (Alexopoulou et al., 2017). According to the Council of Europe (2001), A1 level learners are expected to have basic lexical knowledge of isolated words and phrases related to concrete situations, while higher-level learners are expected to have sufficient vocabulary repertoire to express more complex thoughts in a wider range of topics.

Mechanics refers to test-takers' ability to conform to mechanic conventions of standard written English, such as spelling, punctuation, and capitalization, when composing their responses. According to the Council of Europe (2001), A1 and A2 level learners should be able to write basic, short words with reasonable phonetic accuracy while a B1 learner is expected to produce writing with spelling and punctuation accurate enough to be intelligible most of the time.

4.2. Rating Tool Design

Given the desire to provide test takers with feedback to support their learning, it was determined that conventional scoring approaches used in large-scale testing—such as holistic and analytic rating scales—were not sufficiently fit for purpose. With that in mind, an approach used more frequently in teaching and learning situations—checklists—was chosen.

In L2 assessment, checklists are commonly used by teachers (and possibly classmates) for continuous assessment of class performances, pieces of work, and projects throughout the course. They can also be used for summative assessment at the end of the course (Banerjee & Wall, 2006; Council of Europe, 2001). Additionally, checklists, such as the CEFR “can-do” self-assessment checklists or the portfolio assessment checklists developed by the American Council on the Teaching of Foreign Languages, are very popular for students' portfolio assessment.

In the course of developing checklist items, it was noted that raters can consistently identify performances that were in between a “yes” and a “no”. The decision was thereby taken to make checklist items with three options where appropriate, which has the added benefit of providing finer grained information about performance. Because discrete observations are made about many aspects of test takers' writing performances, information becomes available about their strengths and weaknesses. Multiple indicators were developed based on the test tasks and the four aspects of writing ability previously identified. Table 3 provides a summary, though it should be noted that some of the items below are represented by multiple checklist items.

4.3. Piloting the Checklist Scale

A pilot study was conducted to ensure that the MET Go! Writing test checklist functioned as expected

Table 3: Evaluation Criteria for the MET Go! Writing Test

Criteria	Descriptions of features
Task Completion	<ul style="list-style-type: none"> Relevance of response to task • Cohesion and logical progression of the response • Richness of the response (i.e. elaboration, supporting details)
Grammar	<ul style="list-style-type: none"> Use of grammatical resources to add meaning • Verb tense • Accuracy • Complexity
Vocabulary	<ul style="list-style-type: none"> Use of lexical resources to add meaning • Accuracy • Complexity of range
Mechanics	<ul style="list-style-type: none"> Use of standard mechanics conventions • Spelling • Punctuation • Capitalization

in terms of its usability, meaningfulness, and ability to distinguish appropriately between test-takers at different levels. Thirteen raters with a background in linguistics/ TESOL scored written responses from 2,041 test-takers from various countries, such as Argentina (188), Bolivia (151), Brazil (135), Colombia (121), Greece (45), Mexico (1,070), Paraguay (34), Peru (140), South Korea (92), and Uruguay (60). Each test performance was scored by one to four raters using the rating tool.

Results from exploratory factor analysis showed that the checklist items measured the same underlying construct, but only correlated moderately, indicating that they were measuring distinct aspects of the ability being measured. Multi-facet Rasch measurement analyses indicated that the checklist generally functioned as intended. All of the criteria included in the final rating tool functioned well, although a small number of changes were made as a result of the pilot study. Specifically, a criterion for Task 1 about details being relevant to the story was removed due to its high correlation with the other two Task 1 criteria, and the overall coherence criterion was replaced with individual coherence criteria for Task 2 and Task 3 to obtain a more accurate measure of the test-takers coherence. Additionally, while a formal survey of the raters was not conducted, the raters did report positive attitudes toward the checklist during informal discussions about its use, generally agreeing that the checklist was easy to understand and apply, and that the descriptors were clear, distinguishable from one another, and appropriate for rating test-takers' responses. They also reported high confidence in their scores when using the rating tool.

results can help them recognize their strengths and weaknesses and make decisions about which strategies can best help them improve their English. For ESL/EFL instructors, these results can help them place students into appropriate classes, monitor the progress of students in a class, and provide diagnostic information to identify areas where instruction is needed.

5. INTERPRETING WRITING TEST SCORES

MET Go! Writing Test scores are intended to reflect test-takers' ability to write successfully in English demonstrating knowledge of a range grammar of grammar structures, vocabulary, and mechanics conventions to produce cohesive text that is connected, coherent, and intelligible. Test takers who complete the writing test will receive a score report that includes a scaled score (0-52) and CEFR level (Below A1-B1) based on their overall performance on the writing test, as well as personalized feedback in the form of a performance descriptor statement and a recommended learning activity based on their performance on the different parts of the writing test. For test takers, these

6. REFERENCES

- Alexopoulou, T., Michel, M., Murakami, A. & Meurers, D. (2017). Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing natural language processing techniques. *Language Learning*, 67, 180-208. doi:10.1111/lang.12232
- Bachman, L. F. & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bae, J., & Lee, Y-S. (2010). The validation of parallel test forms: 'Mountain' and 'beach' picture series for assessment of language skills. *Language Testing*, 28(2), 155–177.
- Banerjee, J., & Wall, D. (2006). Assessing and reporting performances on pre-sessional EAP courses: Developing a final assessment checklist and investigating its validity. *Journal of English for Academic Purposes*, 5(1), 50-69.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, U.K: Press Syndicate of the University of Cambridge.
- Council of Europe. (2016). *Collated representative samples of descriptors of language competences developed for young learners aged 11-15 years*. Retrieved from <https://rm.coe.int/1680697fc9> on June 7, 2018.
- Council of Europe. (2018). *Common European Framework of Reference for Languages: learning, teaching, assessment. Companion Volume with New Descriptors*. Retrieved from <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T, & McNamara, T. (1999). *Dictionary of Language Testing*. Cambridge: Cambridge University Press.
- Fulcher, G. (2003). *Testing second language speaking*. Harlow, UK: Pearson Longman.
- Hasselgreen, A., & Caudwell, G. (2016). *Assessing the language of young learners*. Bristol, CT: Equinox Publishing Ltd.
- Knoch, U. (2009). *Diagnostic writing assessment: The development and validation of a rating scale*. New York: Peter Lang.
- Lumley, T. (2005). *Assessing second language writing: The rater's perspective*. New York: Peter Lang.
- McKay, P. (2015). *Assessing young learners*. Cambridge: Cambridge University Press.
- McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press.
- Savignon, S. J. (2005). Communicative Language Teaching: Strategies and Goals. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 635-651). Mahwah, NJ: Lawrence Erlbaum Associates.
- Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Advances in Applied Psycholinguistics, Volume 2: Reading, writing, and language learning* (pp. 142-175). Cambridge: Cambridge University Press.
- Upshur, J. & Turner, C. (1995). Constructing rating scales for second language tests. *ELT Journal*, 49(1), 3–12. doi: 10.1093/elt/49.1.3
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Yu, G. (2009). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, 31(2), 236–259. <https://doi.org/10.1093/applin/amp024>