



Principles and Guidelines for Adopting or Creating Locally Developed Benchmark Assessments for Checkpoint B *Modern Languages**

Purpose

Checkpoint B Benchmark Assessments are selected by Local Educational Agencies (LEAs). They may be designed by the LEA itself, by regional or statewide educational consortia, or by commercial vendors.

The purpose of these Guidelines is twofold. They provide a framework to LEAs for constructing valid, reliable, and equitable locally developed Checkpoint B Benchmark Assessments aligned to the [NYS Learning Standards for World Languages \(2021\)](#) while allowing them flexibility to address specific local needs and priorities. Additionally, these guidelines provide a framework to regional or statewide consortia for constructing valid, reliable, and equitable Checkpoint B Benchmark Assessments aligned to the NYS Learning Standards for World Languages that will be made available for adoption by LEAs.

If an LEA makes the decision to adopt a commercial assessment, they may select from those already identified, evaluated and approved by NYSED for meeting the standards for the [4 + 1 Pathways requirements](#).

The guidance that follows pertains to those assessments developed by LEAs or regional or statewide educational consortia.

Guiding Principles

The following principles guided the development of these recommendations:

1. Maintaining a valid and reliable assessment of Checkpoint B in order to maintain the benchmark as the current requirement for the Regents Diploma with Advanced Designation that students can realistically attain after four years of continuous study;
2. Maintaining fidelity to the language functions, topics, and [performance indicators](#) for Category 1-2 Languages for Intermediate Low to Intermediate Mid and Category 3-4 Languages for Novice High to Intermediate Low from the revised [NYS Learning Standards for World Languages \(2021\)](#);
3. Considering pragmatic approaches and protocols for administration based on past practices that were shown to be successful, along with current research-based practices; and

**This guidance is for teachers of modern languages, other than American Sign Language.*

4. Considering overall assessment and task designs that are user-friendly and can be realistically administered, completed, and scored within the end-of-year timeframe.

Other Considerations

The following aims were also considered:

1. Establishing a foundation for the subsequent selection of Checkpoint C Benchmark Assessments; and
2. Serving as a milestone assessment en route to Seal of Biliteracy Assessments.

Guidelines

Other assessments. Guidelines for locally developed Checkpoint B Benchmark Assessments for American Sign Language (ASL) and Classical Languages are not addressed in this guidance document. Although ASL is a Modern Language, assessment guidance takes into account its emphasis on visual/signed modalities rather than oral/aural modalities. Guidelines for Classical Languages reflect the differences that exist between Modern and Classical Languages as articulated in the [NYS Learning Standards for World Languages](#) for Classical Languages and the [Proficiency Ranges and Performance Indicators](#) for Classical Languages.

Who needs to take this assessment. Three credits of a World Language remain a graduation requirement for all students in NYS who are pursuing a Regents Diploma with Advanced Designation except those students who complete a 5-unit sequence in the Arts or Career and Technical Education (CTE). Successful performance on a Checkpoint B Benchmark Assessment in tandem with passing the Checkpoint B course sequence fulfills the world language requirement of the Regents Diploma with Advanced Designation.

Administration. Checkpoint B Benchmark Assessments are conducted as an on-demand set of performance tasks or a test administered at the end of the course coinciding with the administration dates of Regents Exams. On-demand tasks, given under traditional protocols for Regents Examinations, ensure the integrity of assessment results and reduce the likelihood of plagiarism or cheating. The final score for the Checkpoint B Benchmark Assessment can only include scores for tasks which meet these criteria.

Assessment tasks for Interpersonal Speaking (and Presentational Speaking, if included) should be conducted prior to the exam administration date during the final ten weeks of the academic calendar including the exam window for the administration of the June Regents Examinations. Any sooner than this may not allow adequate instructional time for all students to have developed sufficient readiness to demonstrate the proficiency target. The Checkpoint B Benchmark Assessment is designed to be a summative benchmark assessment.

Technology support and safeguards. Individual LEAs may use technology tools to administer portions of the Checkpoint B Benchmark Assessment consistent with the nature of the standard assessed. LEAs may additionally use technology tools to score objective items and record scores on any assessment item. Technology tools must provide sufficient safeguards to guarantee the integrity of the assessment.

Scope and content. The Checkpoint B Benchmark Assessment will be targeted to the relevant performance indicators. For Category 1-2 languages, the target is Intermediate Mid with Intermediate Low considered passing. For Category 3-4 languages, the target is Intermediate Low with Novice High considered passing as shown in Table 1.

Proficiency Checkpoint and Benchmark Assessment Ratings	Category 1-2 Languages	Category 3-4 Languages
A Passing Rating for Benchmark Assessment	Novice Mid	Novice Mid
A Target Rating for Benchmark Assessment	Novice High	Novice High
B Passing Rating for Benchmark Assessment	Intermediate Low	Novice High
B Target Rating for Benchmark Assessment	Intermediate Mid	Intermediate Low
C Passing Rating for Checkpoint C	Intermediate Mid	Intermediate Low
C Target Rating for Seal of Bilingual	Intermediate High	Intermediate Mid

Table 1. Proficiency Checkpoints and Benchmark Assessment ratings by language category.

All Communication Standards (Standard 1 Interpretive Communication, Standard 2 Interpersonal Communication, and Standard 3 Presentational Communication) are assessed in the on-demand Checkpoint B Benchmark Assessment, although not all skills of these standards need to be included in the on-demand assessment. For example, Presentational Writing is assessed in the on-demand assessment, while Presentational Speaking could instead be assessed in a summative assessment as part of the course, and not as part of the on-demand assessment.

The content of the Checkpoint B Benchmark Assessment is intended to reflect a sufficiently broad representation of the four themes and 16 topics identified in the [NYS World Language Themes and Topics](#) within meaningfully integrated contexts.

Weighting of sections. The sections of the Checkpoint B Benchmark Assessment are weighted as follows:

Standard 1 Interpretive Communication – 40% Total Weight

- Interpretive Reading (20-25%)
- Interpretive Listening (15-20%)

Standard 2 Interpersonal Communication – 30% Total Weight

- Interpersonal Speaking (30%)

Standard 3 Presentational Communication – 30% Total Weight

- Presentational Writing (20-30%)
- (optional) Presentational Speaking (0-10%)

Items for the Culture Standards, Standard 4 Relating Cultural Practices and Products to Perspectives and Standard 5 Cultural Comparisons, are not addressed separately from the Communication Standards. Instead, items that assess the Culture Standards may be included in Standard 1 Interpretive Communication.

Items for all sections are aligned with the [performance indicators](#) for Checkpoint B for the language category, either Category 1-2 or Category 3-4 as follows:

- Category 1-2 Languages. With Intermediate Mid the target rating for Category 1-2 languages and Intermediate Low the passing rating, on-demand assessments should be designed to ensure that students with Intermediate Low proficiency can pass, as shown in Table 1. To accomplish this, in the Interpretive Reading and Interpretive Listening sections, approximately two thirds of the items are aligned to Intermediate Low proficiency with the remaining one third aligned to Intermediate Mid proficiency. For the Interpersonal Speaking, Presentational Writing, and Presentational Speaking sections of the benchmark assessment, the associated [scoring guide](#) includes criteria for Novice Mid, Novice High, Intermediate Low, and Intermediate Mid proficiency. The point values for Intermediate Low criteria, when totaled, constitute two thirds of the possible points for that section. The point values for Intermediate Mid constitute the remaining one third of the total possible points for the section.
- Category 3-4 Languages. With Intermediate Low the target rating for Category 3-4 languages and Novice High the passing rating, on-demand assessments should be designed to ensure that students with Novice High proficiency can pass, as shown in Table 1. To accomplish this, in the Interpretive Reading and Interpretive Listening sections, approximately two thirds of the items are aligned to Novice High proficiency with the remaining one third aligned to Intermediate Low proficiency. For the Interpersonal Speaking, Presentational Writing, and Presentational Speaking sections of the benchmark assessment, the associated [scoring guide](#) includes criteria for Novice Low, Novice Mid, Novice High, Intermediate Low proficiency. The point values for Novice High criteria, when totaled, constitute two thirds of the possible points for that

section. The point values for Intermediate Low constitute the remaining one third of the total possible points for the section.

English support for English Language Learners (ELLs). English-language *directions* in locally developed Checkpoint B Benchmark Assessments may be translated in writing into home languages for ELLs.

In addition, LEAs may provide ELLs with an oral translation of English Language *questions*, *answer choices*, and *prompts* within these assessments. All translations must be oral, direct translations of wording in English. Other than for directions as indicated above, written translations are not allowed. No clarifications or explanations may be provided.

Per CR Part 154, these testing accommodations are designated for ELLs. LEAs must make all necessary arrangements to make these accommodations available for ELLs who need them. Such testing accommodations are also afforded to former ELLs for two additional years after achieving proficiency on the NYSESLAT.

Assessment Format

This section contains guidelines for item development and design for use by LEAs and regional and statewide consortia in assessing each of the five [NYS Learning Standards for World Languages](#) for Checkpoint B Benchmark Assessments. Guidelines for Category 1-2 languages and Category 3-4 languages are listed separately. Beside the title of each section is the range of percentages for weighting each section as presented above.

Standard 1 - Interpretive Reading (20 -25%)

Category 1-2 Languages. This section assesses the student’s ability to understand [authentic resources](#) at Intermediate Mid. For Checkpoint B, students identify the main idea and key information from short straightforward texts on a variety of familiar topics. Tasks can include answering multiple choice questions and responding to constructed response questions. This section may also contain items to assess Culture Standards 4 and/or 5. Interpretive tasks may include some items in English and some in the target language.

Category 3-4 Languages. This section assesses the student’s ability to understand [authentic resources](#) at Intermediate Low. For Checkpoint B, students identify the topic, main idea, and related information from short texts on familiar topics. Tasks can include answering multiple choice questions and responding to constructed response questions. This section may also contain items to assess Culture Standards 4 and/or 5. Interpretive tasks may include some items in English and some in the target language.

Standard 1 - Interpretive Listening (15 - 20%)

Category 1-2 Languages. This section assesses the student's ability to understand audio input from [authentic resources](#) at Intermediate Mid. As with Interpretive Reading, at this level, students identify the main idea and key information from short texts on a variety of familiar topics. Teachers may read transcripts or excerpts of transcripts from authentic resources as prompts for this section or may play recordings or portions of recordings of short radio ads, audio announcements, video clips, interviews, or podcast excerpts, all in their original forms. All students hear each audio text twice. Additional repetitions are given when necessary to comply with provisions of IEP or 504 specifications. Formats may include answering multiple choice questions or responding to constructed response questions. This section may also contain items to assess Culture Standards 4 and/or 5. Interpretive tasks may include some items in English and some in the target language.

Category 3-4 Languages. This section assesses the student's ability to understand audio input from [authentic resources](#) at Intermediate Low. As with Interpretive Reading, at this level, students identify the topic, main idea, and related information from short straightforward texts on familiar topics. Teachers may read transcripts or excerpts of transcripts from authentic resources as prompts for this section or may play recordings or portions of recordings of short radio ads, audio announcements, video clips, interviews, or podcast excerpts, all in their original forms. All students hear each audio text twice. Additional repetitions are given when necessary to comply with provisions of IEP or 504 specifications. Formats may include answering multiple choice questions or responding to constructed response questions. This section may also contain items to assess Culture Standards 4 and/or 5. Interpretive tasks may include some items in English and some in the target language.

Standard 2 - Interpersonal Speaking (30%)

Category 1-2 Languages. This section assesses the student's ability to exchange information, preferences, feelings, or opinions with another person about familiar contexts consistent with the performance indicators for Intermediate Mid. Students are individually assessed with a teacher serving as the conversation partner. As an example, the format could include spontaneous role play situations that are within the experience of the learners in which the students always play themselves and the teacher plays the role of a friend, teacher, or other adult. Another potential format is an interview, where both the student and teacher play themselves. Students choose from a selection of conversation prompts based on the [four themes and 16 Checkpoint B topics](#) and the [four interpersonal language functions](#) in Standard 2. Learners carry on two conversations which are evaluated on a [scoring guide](#) based on language of the [NYSED Master Rubrics](#) for Intermediate Mid.

Category 3-4 Languages. This section assesses the student's ability to request and provide information; express, ask about, and react with some details to preferences, feelings, or opinions with another person about familiar contexts consistent with the performance indicators for Intermediate Low. Students are individually assessed with a teacher serving as

the conversation partner. As an example, the format could include spontaneous role play situations that are within the experience of the learners in which the students always play themselves and the teacher plays the role of a friend, teacher, or other adult. Another potential format is an interview, where both the student and teacher play themselves. Students choose from a selection of conversation prompts based on the [four themes and 16 Checkpoint B topics](#) and the [four interpersonal language functions](#) in Standard 2. Learners carry on two conversations which are evaluated on a [scoring guide](#) based on language of the [NYSED Master Rubrics](#) for Intermediate Low.

Standard 3 - Presentational Writing (20-30%)

Category 1-2 Languages. This section assesses the student's ability to narrate personal experiences, present information, and state and support viewpoint(s) on familiar, concrete, and researched topics. The section could include two or three writing tasks representing contexts suggested by the [four themes and 16 Checkpoint B topics](#) and the [five presentational language functions](#), one of which may be in a read-to-write format.* Scoring of Presentational Writing is aligned to the performance indicators for Intermediate Mid and accomplished with a [scoring guide](#) based on language of the [NYSED Master Rubrics](#) for Intermediate Mid.

* The read-to-write format—interpreting an authentic text then writing in response to that text—aligns to the [performance indicators](#) (Category 1-2 languages) for Intermediate Mid which include writing in response to researched topics. It is, therefore, an appropriate format for Presentational Writing at Checkpoint B for Category 1-2 Languages.

Category 3-4 Languages. This section assesses the student's ability to present information about their life and activities, to state and make attempts to support preferences and opinions on familiar topics. The section could include two or three short writing tasks in familiar formats* representing contexts suggested by the [four themes and 16 Checkpoint B topics](#) and the [five presentational language functions](#). Scoring of Presentational Writing is aligned to the performance indicators for Intermediate Low and accomplished with a [scoring guide](#) based on language of the [NYSED Master Rubrics](#) for Intermediate Low.

* The read-to-write format—interpreting an authentic text then writing in response to that text—is not aligned to the [performance indicators](#) (Category 3-4 Languages) for Intermediate Low which call for students to write about their life and activities. It is, therefore, not an appropriate format for Presentational Writing at Checkpoint B for Category 3-4 languages.

Standard 3 - Presentational Speaking (Optional) (0-10%)

Presentational Speaking is not included in commercial proficiency assessments. LEAs and regional or statewide consortia may include Presentational Speaking as part of the Checkpoint B Benchmark Assessment. This part would be completed prior to the administration of the on-demand examination in the last ten weeks of the school year including the exam window for the administration of the June Regents. The task could be recorded using a computer application. To maintain integrity, the Presentational Speaking task should be administered to

all learners through electronic means simultaneously or a sufficient number of equally challenging tasks should be assigned randomly to students.

Category 1-2 Languages. A Presentational Speaking section would assess the student's ability to narrate personal experiences, present information, and state and support viewpoints on familiar topics in contexts suggested by the [four themes and 16 Checkpoint B topics](#) using one or more of the [five language functions of Standard 3](#) by speaking in response to a prompt. Students would be given a prompt, a short time to think and make some notes, and then record their Presentational speech sample. After sufficient time to prepare, Checkpoint B students are given 1½ to 2 minutes (1:30 – 2:00) to provide their speech sample. The Presentational Speaking task would be assessed on a [scoring guide](#) based on the language of the [NYSED Master Rubrics](#) for Intermediate Mid.

Category 3-4 Languages. A Presentational Speaking section would assess the student's ability to present information about their life and activities, to state and make attempts to support preferences and opinions on familiar topics in contexts suggested by the [four themes and 16 Checkpoint B topics](#) using one or more of the [five language functions of Standard 3](#) by speaking in response to a prompt. Students would be given a prompt, a short time to think and make some notes, and then record their Presentational speech sample. After sufficient time to prepare, Checkpoint B students are given 1½ to 2 minutes (1:30 – 2:00) to provide their speech sample. The Presentational Speaking task would be assessed on a [scoring guide](#) based on the language of the [NYSED Master Rubrics](#) for Intermediate Low.

Standard 4 - Relating Cultural Practices and Products to Perspectives (Optional)

Culture Standards are not assessed in commercial proficiency assessments. However, LEAs and regional or statewide consortia wishing to include assessment of Culture Standard 4, Relating Cultural Practices and Products to Perspectives, may include items in the Interpretive Reading and/or Interpretive Listening sections that ask students to describe cultural products and practices represented in [authentic resources](#). This could be done by developing multiple choice or constructed response questions or other task types as described for Interpretive Reading and/or Listening based on the same authentic resource(s) used to assess Standard 1, Interpretive Communication. Another possible format is graphic organizer completion which is scored according to a proficiency-based rubric using language from the [NYS World Language Master Rubric](#) for Novice High. These items can be related to, but should not require, specific prior knowledge of the cultural products or practices by the students in order for students to be able to answer correctly. Tasks may include some items in English and some in the target language.

Standard 5 – Cultural Comparisons (Optional)

Culture Standards are not assessed in commercial proficiency assessments. LEAs and regional or statewide consortia wishing to include assessment of Culture Standard 5, Cultural Comparisons, may include items in the Interpretive Reading and/or Interpretive Listening sections that ask students to compare cultural products and practices within the [authentic resources](#). This could be done by answering multiple choice questions or responding to a constructed response

question or other task types as described for Interpretive Reading and/or Listening based on the same authentic resource(s) used to assess Standard 1, Interpretive Communication. These items can be related to, but should not require, specific prior knowledge of the cultural products or practices by the students in order for students to be able to answer correctly. Tasks may include some items in English and some in the target language.