

# **The *Student Oral Proficiency Assessment (SOPA)*: Overview and Appropriate Use**

This document provides an overview of the *Student Oral Proficiency Assessment (SOPA)* and explains appropriate use of the assessment and the resulting scores for decision-making within programs. The final section briefly summarizes key research studies about the SOPA.

## **I. Overview & Purpose**

### **What is the SOPA?**

The SOPA is a paired oral interview protocol that includes semi-scripted tasks that are administered by a trained administrator. The SOPA is scored holistically by a trained rater using rubrics for speaking and listening.

The SOPA was developed for use within US-based world language programs at the Pre-Kindergarten through Grade 8 (PK-8) levels. The assessment is flexible by design and can be adapted for use in any language and in a wide variety of language program types. The SOPA has a number of adjustments that can be made to align with a program's purposes for using the assessment and resources for administration.

See the *SOPA Manual* for a complete description of the assessment.

### **What does the SOPA assess?**

The SOPA is designed to assess a student's level of oral proficiency defined as "the ability to communicate verbally in a functional and accurate way in the target language" (Omaggio, 1986). SOPA results show students' proficiency in speaking and listening as described by the ACTFL Proficiency Guidelines.

### **How should results from the SOPA be used?**

The intended use of SOPA ratings is to determine a student's progress in developing oral proficiency in a language program. Appropriate uses of SOPA results include using the results to:

- Track student language growth
- Inform curriculum and instruction
- Provide feedback to instructors and students
- Evaluate a program
- Communicate learning outcomes
- Meet accountability requirements

The SOPA is intended to have a positive impact on users and stakeholders including programs, teachers, students, parents, and program administrators. The test format is intended to be engaging and meaningful to students and teachers, and the results are intended to be useful for programmatic purposes. Additionally, the SOPA meets the need for a flexible assessment tool that can be used across many languages, and particularly in languages where there are limited assessments or resources.

### Who can take the SOPA?

The SOPA is designed for children in grades Pre-Kindergarten through Grade 8 who are studying a world language. Students who take the SOPA may include: first-language speakers of English learning a world language; English/Multilingual Learners who are learning English in addition to the first language(s) they speak; and heritage learners of a world language. The assessment has been piloted and widely used with these populations.

## II. SOPA Description

### What is the format of the SOPA?

The SOPA takes approximately 20 minutes to complete and includes a series of tasks that increase in difficulty. In a typical administration, the SOPA is administered by a trained administrator to two students at a time and ratings are assigned by a trained rater who is also present at the time of administration. Performances by the two students are assigned a global rating using the SOPA Rating Scale. During the assessment, students are encouraged to “help” their partner in the target language, but the interviewer interacts with each student separately. Figure 1 presents an overall structure of the SOPA.

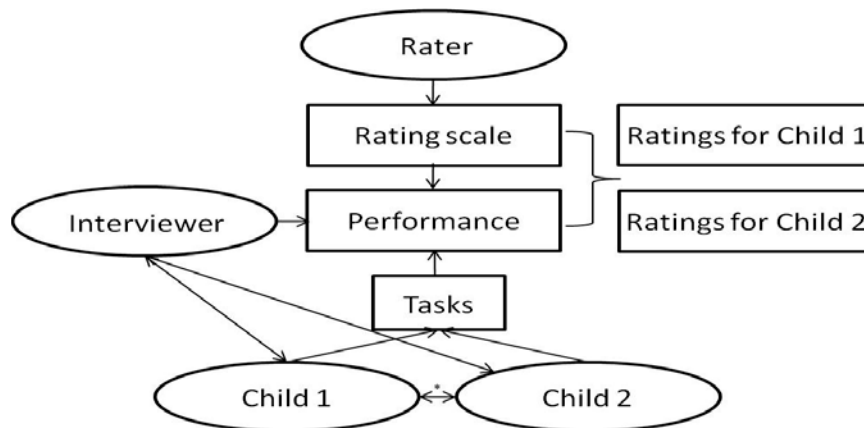


Figure 1. Structure of the SOPA

The assessment tasks follow the natural development of language skills, focusing first on receptive listening skills and then on productive speaking skills. This sequence allows the students to experience immediate success in their responses (receptive skills typically being less demanding than productive skills), thus building their confidence. The tasks in the SOPA are

designed with a hierarchy of functions to enable an administrator to assess the student's highest sustained oral proficiency level at a given point in time.

### **Who administers and rates the SOPA?**

Administration of the SOPA typically requires two adults. One is a rater, and primarily takes notes and assigns ratings in addition to operating the recording equipment for rating verification. The other is an administrator, and primarily conducts the assessment. This separation of roles allows the rater to record as much as possible of each student's speech to facilitate assigning accurate ratings while the administrator focuses on eliciting language that will demonstrate the students' highest proficiency levels.

It is important that the SOPA be administered by individuals who are (a) trained in administering and rating the assessment, (b) experienced working with young students, and (c) fluent in the target language. Language specialists in schools and/or districts can receive training in SOPA administration and rating through the SOPA Training Program.

### **What is the design of the SOPA Rating Scale?**

The SOPA is a criterion-referenced assessment. This means that a student's performance rating is determined by the criteria on the rating rubric that most closely aligns with the evidences elicited during the assessment, showing what the student can do in the target language. Like the ACTFL Guidelines, the SOPA Rating Scale is not language specific. Rather, it describes language proficiency in terms of features that are generally found in most languages.

The SOPA Rating Scale describes performances in speaking and listening at the Novice, Intermediate, and Advanced levels of language proficiency. Each main level is divided further into three sub-levels, Low, Mid, and High, to distinguish differences within the main levels. There are nine sub-levels from Novice-Low to Advanced-High in the SOPA Rating Scale.

### **What is the rating process?**

The SOPA is locally scored by a trained rater. In a typical administration, the rater observes the test administration, takes notes, and assigns a global rating in speaking and listening to each of the two students at the end of the assessment. The global rating is a final, summary evaluation of a student's language proficiency in speaking and listening based on evidence across all tasks. It represents the highest level of language proficiency demonstrated during the assessment.

Raters must complete training before rating students on the SOPA. The SOPA Training Program includes rating practice in English as well as Spanish and Mandarin Chinese. If raters will rate the SOPA in a different language, they review training materials in English and then practice locally in the target language. Thorough training and local practice ensure that student performances are rated reliably and consistently with the SOPA Rating Scale.

### III. Local Use of the SOPA

#### How do we ensure the SOPA tasks are appropriate for our students?

Before the SOPA is administered, the SOPA administrator selects one of two tracks of the assessment to administer to students based on their estimated proficiency levels. This ensures that students are administered tasks at the appropriate level of difficulty to rate their proficiency level.

Additionally, the contents of the SOPA tasks are adapted to meet the needs of each program that uses them. All program-specific SOPA tasks reflect the structure and difficulty levels as set by the standard formats, but specific topics for the tasks are informed by the program's curriculum and the grade level(s) to be assessed. Extensive SOPA training materials, models, and task templates help ensure that intended task difficulty level and structure are maintained as the SOPA is adapted for local use, and this is all supported within the SOPA Training Program so that local users are equipped to adapt tasks and scripts to their contexts.

#### What training materials are available?

CAL provides a comprehensive, self-access SOPA Training Program for SOPA administrators and raters. The online training program is designed to support programs in using the test, adapting tasks to fit the program and student population, administering the SOPA reliably, assigning ratings, and reporting SOPA scores.

The SOPA Training Program is in English and includes extensive video examples of SOPA administrations that programs can view and then use to adapt materials into the appropriate target language. The SOPA Training Program also includes videos, scripts, rating practice, and materials in Spanish and Mandarin Chinese as two of the most widely administered languages for SOPA.

SOPA administrators and raters must be thoroughly trained to ensure the reliable administration and scoring of the SOPA within local programs.

### IV. Additional Background and Research

#### How was the SOPA developed?

The SOPA was developed by the Center for Applied Linguistics (CAL) using a format based on the ACTFL Oral Proficiency Interview (OPI). The ACTFL OPI is an interview format for adult language learners that consists of a warm up, level checks, and a wind down that is administered by a trained interviewer who also is responsible for assigning ratings.

The SOPA has some structural similarities to the ACTFL OPI, but with some important differences: 1) the SOPA is administered by a trained administrator to two students at a time, 2) ratings are assigned by a trained rater who is also present at the time of administration, 3) the

interview prompts are prior to the assessment to ensure consistency across administrations, and 4) the context and content of the SOPA is appropriate for young language learners.

The SOPA was originally developed by CAL in 1991 for use in immersion language programs. A non-immersion version of the SOPA was developed in 1996. Later, in 2001, a version of the SOPA called the ELLOPA (Early Language Listening and Oral Proficiency Assessment) was adapted for young language learners in grades PK-2. Most recently, in 2024, CAL released a new, self-access training program for the SOPA, streamlined the assessment protocol to increase usability and reliability for language programs, and updated the scoring rubric to provide greater clarity for scoring. Because the 2024 update to the SOPA can be used for grades PK-8, the ELLOPA is no longer needed or supported as a separate assessment.

### **What research supports the use of the SOPA?**

Since its initial development, several research studies have been conducted to support the SOPA and provide evidence for the appropriateness of using the SOPA in language programs.

Several key studies are summarized below. Note that evidence from the ELLOPA, an earlier version of the SOPA for young language learners, is included here as the updated (2024) SOPA subsumes this assessment.

#### ***Rhodes, Thompson & Snow (1989)***

---

Prior to the development of the SOPA, this research study comparing Foreign Language Elementary School (FLES) and immersion programs provided valuable information about world language teaching and learning in these contexts and the particular focus of each type of program. This study identified themes, topics, and language functions addressed in these two types of programs. SOPA and ELLOPA tasks were developed with direct input from immersion and FLES teachers and through surveys of immersion and FLES language curricula. Since the initial creation and field-testing of assessment tasks for SOPA (1991, 1996) and ELLOPA (2001), care has been taken to ensure that all SOPA and ELLOPA tasks continue to be representative of the PK-8 world language teaching domain.

#### **SOPA Validation Studies**

---

The validity testing for the non-immersion and immersion versions of the SOPA took place from 1997 to 1999 and validation for the ELLOPA in 2001 and 2002. Overall, the analyses of the data supported the SOPA and ELLOPA as valid instruments for assessing listening comprehension and speaking proficiency in young language learners across languages.

#### **SOPA Interrater Reliability Study**

---

This study investigated interrater agreement on administrations of the ELLOPA and SOPA. Interrater reliability measures the degree to which two raters assign the same rating to the same student response. The higher the rater agreement is, the more reliable the ratings are. If two raters' scores are the same, the scores are considered to have exact agreement. If two raters'

scores differ by one raw score point, the scores are considered to be adjacent. In the study, recordings of the ELLOPA were double rated by local raters and by CAL raters.

As an example of results, ratings for 57 Chinese language ELLOPAs administered to kindergarten students showed the exact agreement ranged from 77% to 86% and the adjacent agreement ranged from 14% to 23%. There was no discrepant agreement. The Pearson correlations between the raw score assigned by the first and second rater ranged from 0.77 to 0.87. These results indicate that the raters were very consistent in their ratings. Ratings in other grade levels showed similarly reliable results, and a similar study conducted in Spanish also showed very high interrater reliability between local raters and CAL raters.

### **Correlation of SOPA Results with Other Measures**

---

Correlational analyses were conducted between ratings from the SOPA/ELLOPA and teachers' observation ratings on their students' class performance of each of the traits in the SOPA Rating Scale. Teachers' observation ratings of student proficiency were collected using a standard observation form.

As an example of results, in French immersion programs ( $N=76$ ) moderate significant correlations were found between teacher observation ratings and SOPA scores for all traits in the SOPA rating scale. These results suggest that there are moderate relationships between the ratings on the same traits from two different methods used to assess students' oral fluency in the target language. Therefore, the construct measured in the SOPA appears to tap into a similar construct of oral proficiency as measured in the teacher observation assessment.

### **Usability of the SOPA Training Program**

---

The SOPA Training Program was piloted with three programs in summer 2024 with a total of 20 participating educators who reported working with English, Spanish, and Mandarin Chinese. Participating educators were asked to complete the SOPA Training Program, complete rating practice exercises using 25 pre-rated SOPA sample videos in a given language, and provide feedback.

A total of 16 participants completed and submitted the results of their rating practice exercises. When analyzing participants' ratings, if the participant rated the performance to be the same ACTFL proficiency level as the rating determined by CAL raters, it was considered to be an exact agreement. If the participant's rating and the CAL raters' rating were only one level apart, it was considered to be an adjacent agreement.

Results showed that, on average, participants achieved 71% of exact and adjacent agreement with CAL's Speaking ratings and 74% of exact and adjacent agreement with CAL's Listening ratings. This provides evidence that the SOPA Training Program can reliably train educators to rate the assessment with accuracy. These results also provide evidence that educators can complete the training in English and then successfully use the SOPA for other languages.