STAMP 4S Content Development

The STAMP 4S is a test of language proficiency aligned to the ACTFL proficiency scale. It is intended for learners of world languages in the United States who are in middle school, high school, or college. Scores of 1-9 are reported on the STAMP proficiency scale, which aligns directly with ACTFL levels (including sublevels). Scores are used to track the growing proficiency of individuals and the progress of specific world language programs in these schools. It measures the four major language skills: Reading, Writing, Listening, and Speaking.

This paper will address the development of test content for the two groups of skills measured by the test: (a) Listening and Reading (receptive skills) and (b) Writing and Speaking (productive skills).

Content Development for the Reading and Listening sections of STAMP 4S

1. Construct

The construct for the Reading Comprehension test is the ability to comprehend a variety of written texts used for general communicative purposes in a target language. The construct for the Listening Comprehension test is the ability to comprehend a variety of spoken texts used for general communicative purposes in a target language.

2. Overview

The purpose of the test content is to provide evidence of a test-taker's comprehension of Target Language texts (whether written or spoken), targeted to a specific ACTFL major level (Novice, Intermediate, or Advanced). The tests proceed by asking test-takers to read or listen to level-appropriate texts and then answer level-appropriate comprehension questions based on those texts. It's important to note that although final scores on STAMP 4S also cover sub-levels (Low, Mid, and High) for each major level (Novice, Intermediate, or Advanced), that is a result of cut-score setting. At the development stage, passages are selected and items written to the major levels (Novice, Intermediate, and Advanced).

3. Text Characteristics

Texts that serve as target language samples on the STAMP 4S test must have certain characteristics. They must represent a specified variety of the target language (*eg.* Brazilian Portuguese). They must be texts that could reasonably be expected to be encountered in the actual, real-world target language environment. They must be texts that are natural, (*i.e.,* that *could* have been produced by a speaker of the target language for communicative purposes).

3.1 Text Sources

Written and spoken texts used on the Reading Comprehension and Listening Comprehension sections can be authentic, authentic-edited, or purpose-written by Target Language Experts (TLEs). Sources include both online and not. Whether the text is authentic or not, it must be judged by a Target Language Expert as being a text that could be encountered in the target language use environment.

3.2 Text Types

The text types used on the test depend on the proficiency level targeted and the skill. The types of texts used at each level are intended to closely align with the ACTFL proficiency descriptors.

- Reading texts at the Novice level include: lists; charts (*e.g.* schedules, calendars, tables, graphs, etc.); public signs; forms; greeting cards; public notices (*e.g.* classified newspaper ads, bulletin board notices, etc.); private notices (e.g. event invitations, birth announcements, change of address announcements, etc.), product labels.
- Reading texts at the Intermediate level include: classified and commercial ads; announcements; news briefs; simple instructions or directions; user reviews; formal and informal communication (e.g., email, written correspondence), informational texts; class notes (both content and organizational); student compositions; advice (*e.g.* advice columns, Q&A, horoscopes, etc.); very short narratives (e.g. diary entries, etc.).
- Reading texts at the Advanced level include: news stories, narratives, more complex formal and informal communication; more complex instructions or directions; explicitly expressed opinion; more complex informational texts; short biographies (e.*g.* book introduction, obituary, etc.), simple critical reviews.
- Listening texts at the Novice level include: lists (often embedded in Intermediate texts); extremely simple directions (e.*g.* classroom command, etc.), very basic informational exchanges; greetings
- Listening texts at the Intermediate level include: official and informal voicemails; announcements (*e.g.* loudspeaker announcements, outgoing messages, public service announcements); basic informational reports (e.g. weather or traffic), simple informal and formal conversations (e.g. conversations between friends or relatives, conversations between students and teachers, interviews, etc.), commercial advertisements, service encounters, basic instructions or directions, and simple presentations.
- Listening texts at the Advanced level include: more complex conversations; narratives; more complex instructions or directions; news stories, more complex presentations (including some simple basic speeches).

3.3 Text Appropriateness

Texts used on the test must be comprehensible to a learner of the assessed variety of the target language, including with regard to accent, word choice, grammatical features, script, and font. The language used in these texts must also be accurate with regard to spelling or pronunciation, grammar, word choice, and punctuation or intonation.

The texts used on the test must have topics that are interesting and relevant to the test-taking population and should only include age-appropriate material. The topics should also be of interest to people in the target language culture and not appear to be significantly outdated.

Texts that cover inflammatory topics, either in US or Target Language culture, texts that demonstrate racial or religious prejudice, texts that include sexually explicit content, and texts that include vulgar language are <u>not</u> used on the test.

Texts used on the test should function as complete texts, i.e., they should stand alone as a comprehensible text without requiring background information or added explanation. Texts should, however, neither be common, universal knowledge (e.g. directions for how to turn on a computer), nor be relevant only to a very specific audience (e.g. a technical manual for repairing an engine).

4. Task Characteristics

Tasks that provide evidence of comprehension of the target language samples on the STAMP test must have certain characteristics. They must address significant elements of meaning from the target language sample and a test-taker's selection of a response option must provide some evidence of comprehension or lack of comprehension at the proficiency level targeted by the task.

4.1 Task Formats

All questions encountered on the Listening Comprehension and Reading Comprehension sections of STAMP 4S are selected-response (*i.e.,* the test-taker must select from among the set of answer choices provided). There are three formats used in these tests:

- Multiple Choice: there are four response options, each of which is a word, phrase, or sentence written in English; a test-taker selects an option by clicking on a radio button; the test-taker can change the selection by clicking on a different radio button.
- Picture Select: there are four response options, each of which is an image; a test-taker selects an option by clicking on an image; the test-taker can change the selection by clicking on a different image.
- Zone Click: there are three to six response options, each of which is an active zone within the same image; a test-taker selects an option by clicking on one of the active zones in the image; the test-taker can change the selection by clicking on a different zone.

4.2 Task Appropriateness

The first aspect of task appropriateness is level targeted. The element of meaning from the target-language text that is addressed by the comprehension question must be appropriate for test-takers at the major level targeted, as described by the ACTFL proficiency descriptors.

- Novice tasks address: recognition of known words and expressions; main idea; category of items in a list; information found in a table or graph
- Intermediate tasks address: main idea; important details
- Advanced tasks address: main idea; important details; sequence of events; cause and effect

The second aspect of task appropriateness is defensibility. Each question must have one and only one correct answer (key), and that answer must be clearly correct to those test-takers who fully understand the target-language text. The incorrect answer choices (distractors) must be clearly incorrect to test-takers who fully understand the target-language text.

The third aspect of task appropriateness is soundness. Each component of the task must function appropriately. The statement of the task (also known as the "stem") must clearly and concisely convey a comprehension task related to the target-language text and also help the test-taker focus on the ideas in the text that are addressed by the question. The correct answer (also known as the "key") must address author-intended ideas from the text, require comprehension of the text in order to choose (i.e., not be guessable), and **not** stand out from the other answer choices. The incorrect answer choices (also known as "distractors") must also require comprehension of the text in order to eliminate, be logical and plausible answers to the question posed by stem (in the context of the text), and **not** stand out from the other answer choices.

The fourth aspect of task appropriateness is independence. Answering any question on the test must be independent of answering any other question on the test. For questions pertaining to the same text, they must address different elements of meaning from each other, and they must not cue each other (*i.e.* no component of one question should make it easier to find the correct answer for any other question).

5. What the Test-taker Sees

Once test-takers have logged in to a particular test (i.e. target language and skill), they see an instructions screen, a practice question, and a "Start Test" screen. The instructions are given both in the target language and in English.

Reading Test Instructions:

- In this section of the test you will answer questions about the texts you see.
- Step 1: Read the situation that sets the scene or context.
- Step 2: Read the question and the possible answers. This will help you know what to look for as you read the text.
- Step 3: Read the text carefully (skim, scan, search for details and clues). Search for the information that will best answer the question.
- Step 4: Select the best answer for each question.
- Step 5: Click **NEXT** to move to the next question.

Listening Test Instructions:

- In this section of the test you will answer Listening questions.
- Step 1: Read the situation that sets the scene.
- Step 2: Read the question and possible answers. This will tell you what kind of information you are looking for as you listen to the monologue, conversation, or announcement.
- Step 3: Click [>] to listen. You may listen to the audio 2 times only. After two attempts, the button will become inactive.
- Step 4: Select the best answer to the question.

- Step 5: Click **NEXT** to move to the next question.

Once test-takers have seen the above-mentioned screens, they begin the test. They will see a series of 30 screens (each with a separate test question.item), the components of which are described below.

In addition to the target-language spoken or written text, which the test-taker is being asked to comprehend, and the comprehension question that assesses an aspect of the test-taker's comprehension, there are several other components of a test screen:

- Orientation Statement: this is a short statement, in English, that introduces the text type of the target-language text in order to give the test-taker an idea about where this text might be found, which, in turn, allows the test-taker to invoke schemata.
- Reading Background Image: this is an image in which the target-language text is embedded, and it is intended to resemble the text in its real-life context.
- Listening Context Image: this is an image that is intended to depict the speaker(s) heard in the spoken text.
- Direction Line: this is an instruction that is dependent on the task format; it tells the test-taker how to answer the question (e.g., to select the best answer, image, or part of an image).

Figure 1. *Sample Reading item in Spanish (Advanced Level)*

6. Quality Assurance

Quality assurance for the STAMP 4S test involves several cycles of review. These reviews can be divided into two major phases: content review and test review. The content review pertains to the components of test content mentioned above: the target-language texts, comprehension questions, and peripheral components such as audios and images. Depending on the component, the content is individually reviewed by Avant-external Target Language Experts or Avant-internal Content Development Team members. Any content that will be seen by test-takers must first be seen and approved by at least two different Target Language Experts. These Target Language Experts can be consulted on any aspect of any of the individual components, but every "item set" (the target-language text accompanied by any

comprehension questions based on that text and all peripheral components) must be reviewed as a whole. Reviewers are asked to judge whether the item set meets all of the appropriateness criteria mentioned above for texts, tasks, and peripheral components. Reviewers also ensure that each single passage and item developed are free of bias (bias review) and are fair (fairness review) to all groups of people, and that no groups are either offended or represented in a negative light by either passage or item.

Once the content, in the form of item sets, is deemed to be finalized, a Quality Control review is performed using a delivery system that replicates what the content will look like to test-takers. Finally, the Rasch-informed routing rules (for moving from module to module within our multi-stage adaptive design, seen below) and Rasch-informed scoring table are given a quality-control review to make sure they are functioning as intended and that all scores assigned at the end of a STAMP 4S administration are assigned correctly.

7. Test Assembly

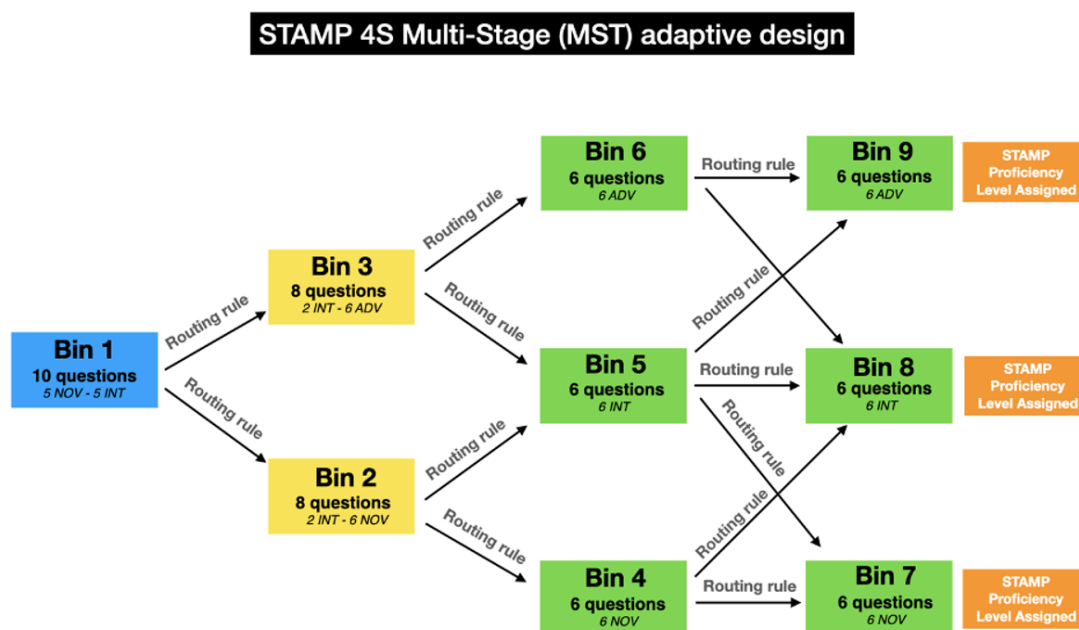Each test uses a computer adaptive, four-stage design, as illustrated below:

Figure 2. *Four-stage, adaptive STAMP 4S design.*

Item sets are assembled into the modules (also known as bins) based on the targeted level of the item set and some other distributional considerations. Item sets for the highest performing test-takers are assembled so that the test gets more difficult (*e.g.,* the difficulty of bin 9 is higher than that of bin 6). Item sets for the lowest performing test-takers are assembled so that the test gets less difficult (*e.g.,* the difficulty of bin 7 is lower than that of bin 4. In addition, item sets are distributed among the modules based on their characteristics (as mentioned above) such as text type and topic, in order to ensure that any given test-taker, regardless of the path they follow in the test, gets exposed to as wide a sample of text types, functions, etc, as possible given their proficiency level.

STAMP 4S Prompt Development: Speaking and Writing

8. Introduction

The Writing and Speaking skills of the STAMP 4S test are scored and reported on a scale of 1-8 within the 9-level STAMP proficiency scale, aligning directly with ACTFL levels Novice-Low through Advanced-Mid (including sublevels). This section addresses the development of the Writing and Speaking prompts used in the test.

9. Construct

The construct for the Writing and Speaking sections of the test is to measure the test taker's ability to produce coherent, cohesive, and well-formulated language in the target language in written and spoken form, respectively, addressing tasks that are precisely targeted at the ability level of the test taker.

10. Overview

The purpose of the productive sections of the STAMP 4S test is to deliver Writing and Speaking tasks/prompts that are appropriate for the test taker and will function to generate language that can be rated with a standard rubric. Upon completion of the Reading section, the STAMP 4S assessment system uses the test-taker's demonstrated Reading proficiency to deliver three Writing tasks/prompts that are appropriate for the test taker's level of ability. The same is true upon completion of the Listening section with regard to the three Speaking tasks/prompts that are delivered to the test-taker. These prompts have been developed to maximize language production for test-takers at different proficiency levels. They are developed with both topic and function in mind to assure the best possible language production from each test-taker. Three prompts in the Writing section and three prompts in the Speaking section are used to measure the breadth of topic control of each test-taker.

All Writing and Speaking tasks are replaced each December. This allows schools to do pre-testing in the fall and post-testing in the spring with new content for the productive components of the assessment. The next section outlines the process and procedures used to generate the tasks each year.

11. Prompt Creation Process

All writing and speaking prompts are developed using the same process and development team on an annual basis.

11.1 Prompt Development Team

The prompt development team consists of six individuals across three key roles as follows:

•**Prompt Writers:** Prompt writers are rater managers who have managed raters for at least 2 years and who have classroom teaching experience. Having classroom teaching experience is critical for this role so that they know the types of content and topics students would be expected to master at different proficiency levels.

•**Senior Prompt Reviewer:** The senior prompt reviewer is a senior content manager with over 5 years task writing experience, rating experience, and classroom teaching experience. The purpose of this role is to check the prompts for topic coverage, function appropriateness and level of engagement of each prompt targeted at each level.

•**Prompt Editor:** The prompt editor is a copy editor who checks each completed task for clarity and accuracy.

•**Language Specific Translators:** Translators for each test language to render each prompt into the target language.

11.2 Prompt Development Process

The prompt writing team of four rater managers begin the process by reviewing the prompts from the previous year and identifying those tasks that produced the best responses from the test-takers. As rater managers, they are each familiar with the test-taker responses and the prompts. Using these exemplar prompts, they then identify the characteristics of the prompts that they feel helped to produce good responses and discuss how to work these elements into future prompts.

The prompt writing team then reviews the topics and benchmarks and identifies the topics of focus for the writing process. The topics and benchmarks are posted on our website and are accessible to all concerned parties. https://avantassessment.com/stamp-4s-benchmarks-rubric-guide

Benchmark Level	Topics	Characteristics
NOVICE	• Self • Calendar/Time • Colors/Shapes • Pets/Animals • School/Classroom • Weather/Seasons • Clothing • Food/Beverage • Family/Friends • Home • Places/Geography • Community • Daily Routines • Shopping/Stores • Leisure/Activities	Students who are reading or listening at Novice proficiency are characterized by: • Reliance of learned phrases and basic vocabulary • Ability to recognize the purpose of basic texts • Can understand a core of simple, formulaic utterances

INTERMEDIATE	• Health
	• Holidays/Celebrations
	• Occupations/Professions
	• Transportation Travel/Vacations
	• Future plans
	• Culture
	• Contemporary Issues
	  • Current events
	  • Economics
	  • Literature
	  • Science
	  • Social Science
	  • History
	
	• *Plus more in-depth aspects of Novice topics*
		Students who are reading or listening at Intermediate proficiency are characterized by:
		
		• In reading, ability to understand the main ideas and explicit detail in everyday language
		
		• Ability to use language knowledge to understand information in everyday materials
		
		• Can follow short conversations and announcements on common topics and answer questions about the main idea and explicitly stated details

ADVANCED	• Arts • Politics • Religion • Math • *Plus more in-depth aspects of Novice and Intermediate topics*	Students who are reading or listening at Advanced proficiency are characterized by: • Can understand and use language for straightforward informational purposes • Can understand the content of most factual, non-specialized materials intended for a general audience • Can understand the content of most spoken factual, non-specialized language

Figure 3. *STAMP 4S topics and benchmarks.*

Once the prompt writing team has identified the focus topics, writing assignments are given to each of the four writers outlining the topics and target functions and they each develop 8-10 prompts and then

exchange the prompts that they have developed with one of the other prompt writers for review. When all prompts have been written, reviewed, and edited, the final versions are shared with the Senior Task Reviewer. If there are any edits or adjustment necessary, those will be completed by the reviewer and then all prompts are reviewed and checked for accuracy, fairness, lack of bias, and for clarity by the copy editor. Once the copy editor has completed the final check and review, the finished prompts are given to language specific translators so that they can render the prompts into the target language. Having the prompts delivered in both English and the target language assures that the test takers know what they need to do to complete the task, regardless of which language is their language of preference.

12. Prompt Examples

Following are examples of Writing and Speaking prompts and how they are presented inside of the STAMP test platform.

Figure 4. *Sample Writing Task/Prompt.*

Figure 5. *Sample Speaking Task/Prompt*