

# PEPPER GUIDE 1: DATA COLLECTION

## **Before the Recording Session**

Once instructional, clinical and/or research goals have been established and PepAssess and PepClass outputs selected (see Overview: Flow Chart and PG7), an assessment protocol should be assembled that includes relevant speech and/or language tasks. The Phonology Project assembled reference data using the tasks in the Madison Speech Assessment Protocol (MSAP; see Phonology Project website tab). Users may administer other tasks, but reference data in the Phonology Project technical reports are available only for the MSAP tasks.

When choosing assessment tasks it is important to consider age (chronological and/or mental), dialect, and any other known physical and/or personality characteristics that could influence the participant's ability to successfully complete each task. For example, the MSAP includes 4 sets of measures that are customized for 4 age groups: preschool, school-age, adolescent, and adult. Some factors that were taken into consideration when designing the MSAP for the 4 age groups include intelligibility, linguistic complexity, and response requirements. Detailed descriptions and instructions for administering the MSAP tasks are included in the [MSAP](#) tab on this website.

After assembling an assessment protocol that meets instructional/clinical/research goals and takes into consideration the participant factors mentioned previously, it is important to prepare and practice, preferably in the testing environment. Assemble all testing materials, such as test booklets, scoring forms, toys/games, books, reinforcers, playback equipment (if using pre-recorded stimuli), and recording equipment (audio and/or video). Choose recording equipment that will store a version of the speech signal that is as faithful to its original production as possible, and set up the recording environment to minimize or eliminate background noise, distortion, reverberation, etc. Remember that the goal is to obtain speech samples that are appropriate and efficient for phonetic transcription, and also for prosody-voice coding and acoustic analyses, if the latter two will be included in data reduction procedures. A summary of factors to consider when collecting and recording speech data can be found in Chapter 9 of *Clinical Phonetics (5th ed.; 2019)*. Another recent reference that includes methods for recording and analyzing speech is Ludlow, Kent, and Gray (2018): *Measuring Voice, Speech, and Swallowing in the Clinic and Laboratory*.

## During the Recording Session

This section covers general guidelines for setting up recording and playback (if using pre-recorded stimuli) devices. Guidelines we have found useful for obtaining high-quality conversational speech samples that can be used for speech, language, prosody-voice, and/or acoustic analyses are also included. Information and assessment materials specific to the administration of the MSAP can be accessed by clicking on the link provided in the previous section.

## Recording and Playback

Some of the instructions in this section pertain to the equipment used for the administration of the MSAP at the UW-Madison site and to specific MSAP tasks. For efficiency, abbreviations are used for MSAP tasks. If more detailed technical information on this equipment is needed, please refer to the original operating manual(s). Researchers and clinicians using other makes/models of recording or playback devices should consult the corresponding manuals for operational and technical information on the use of these devices.

Samples at UW-Madison have been most recently recorded on a Marantz CDR420 HD/CD recorder with an impedance-matched microphone. The recorder is placed on a different horizontal surface and maximally distant from the surface on which the microphone is placed to avoid recording any machine noise from the recorder (note the microphone and recorder placement in Panels A and B on p. 5). A session is recorded directly onto the hard drive of the Marantz CDR420 and then burned onto a CD-R (80-minute) disk.

The MSAP speech tasks with pre-recorded stimuli (the LST through the SPT) are played *through external speakers plugged into a laptop* (refer to Panels A and B on p. 5 for placement of the laptop and external speakers). **Due to inadequacies with the internal speakers in computers, all MSAP administrations of the audio stimuli are played through external speakers.** The stimuli are presented in PowerPoint. As shown in Panel A on p. 5, the participant can view the screen during those tasks that include pictures (LST, CWT, VT1, VT2, VT3). As shown in Panel B on p. 5, **the participant should NOT be able to view the screen during non-picture tasks (SRT, NRT, EST, RST, MWT1, MWT2, SPT) unless the visual representations (i.e., words or phonetic characters) of the stimuli have been removed from the PowerPoint slide show.** Instructions for copying and storing the PowerPoint files and playing the computerized section of the MSAP in PowerPoint can be found in the MSAP tab in the Phonology Project website.

## Recording and Playback Guidelines

(NOTE: The following instructions pertain to tabletop, or stand-mounted, microphones.)

All speech tasks that will likely be transcribed (and optionally coded for prosody-voice and/or analyzed acoustically) after the assessment session should be recorded. Additionally, the examiner may choose to record the case history interview if it is convenient or useful to do so. Instructions presented here (some of which were introduced previously) should be followed carefully to maximize the quality of the recordings.

- 1. Record in a quiet environment.** Record speech samples in a quiet environment. Try to eliminate constant background noises such as electrical hum from fluorescent fixtures. If such noises cannot be eliminated, position the participant so that the pick-up pattern of the microphone is oriented away from the noise source. You can also try reducing the mouth-to-microphone distance, so that the speech signal is louder and the background noise relatively softer. Be alert also for any transient noises in the environment, such as those created from passers-by, automobiles, or aircraft; or those produced by the participant handling toys and materials, moving about, touching the microphone cord, and so forth. These can seriously interfere with the speech signal. Handle such potentially costly interferences by having the participant repeat the utterance obscured by noise. Consistent participant-produced noises should be eliminated indirectly by changing his/her position or the work surface. More directly, ask the participant to stop making the noise, because "it makes a noise on the tape recorder and that makes it hard to listen to your ideas later."
- 2. Carefully position the recording equipment.** Because the recording device may set up a resonance in the surface on which it rests, the external microphone should be placed on a different surface than the recorder. Place the microphone far enough away from the recorder to minimize any airborne noise emanating from it.
- 3. Adjust the distance of the microphone from the participant.** The participant's mouth should be no more than 6-8 inches from the microphone depending on his/her typical loudness level. Adjust the angle of the microphone to be slightly less than perpendicular to the participant's lips. A handy guideline is to aim the microphone at the participant's nose or slightly to the left or right of the participant's mouth.
- 4. Adjust and monitor the recording volume (gain).** Have the participant count or say some simple sentences while you adjust the gain on the recorder. For participants with more severe speech involvement, have them say some single words, for instance, by naming objects placed on the table. **Appropriate gain for recording should be -12 to -15 dB.** This should assure that vowels are not distorted, and that all the consonants are audible and undistorted.

5. **Make the participant comfortable with the recording equipment.** If the examiner is casual about the recording equipment, periodic adjustments of the gain and the placement of the microphone should not bother the participant. Inquiries about the recording equipment are not solicited but are handled directly and honestly when they do occur. When responding to the questions of an inquisitive child, for example, the examiner could say, "That's my recorder. It will help me remember all your good ideas." After a brief explanation, the examiner should return the participant's attention to the sampling as quickly and naturally as possible.
  
6. **Position the laptop (or other playback device) and speakers so that the participant can clearly see the screen and hear the stimuli.** In Panel A below, note that the laptop screen is clearly visible to the participant, and the external speakers are pointed toward her. After playing the first pre-recorded stimulus, ask the participant if the stimulus is loud enough (or too loud). Make volume adjustments as necessary to insure that the audio signal is comfortably audible for the participant. (Panel B demonstrates how to play the audio stimuli while the screen is not visible to the participant which, as noted previously, is necessary for the non-picture MSAP tasks.)

It is best to set up the recording and playback equipment and test out these procedures prior to recording and administering the MSAP to a participant. If potential problems are discovered and corrected before an actual assessment session, there will be fewer disruptions within the session, and speech data is less likely to be degraded or lost due to technical difficulties. The panels below provide examples of the set-up and positioning of the participant, examiner, and equipment during administration of the MSAP.



Panel A. The participant (in the green sweater) is positioned directly across from the computer screen so that she can easily see the pictures. Note that external speakers are used for playing the computerized stimuli, and the recording device (to the right of the examiner) is maximally distant from and on a different surface than the microphone.



Panel B. The computer screen is hidden for administration of the non-pictured tasks (SRT, NRT, EST, RST, MWT1, MWT2, SPT) in the computerized portion of the protocol.

## The Phonology Project Speech Sampling Protocol for Young Children

The conversational speech sample serves several assessment needs. It is the primary data source for production phonology, including segmental and suprasegmental data. [Suprasegmental data are obtained from *The Prosody-Voice Screening Profile (PVSP)* (Shriberg, Kwiatkowski, & Rasmussen, 1990).] It can also be used to obtain language production data. The conversational speech sample may be analyzed for self-monitoring behaviors, which may be used as a source of information on comprehension phonology. Self-monitoring behaviors include occurrence of preplanning, prompted self-correction, and spontaneous self-correction, each of which has been defined in Shriberg and Kwiatkowski (1990; Table 3). For young children, a Structural Stage analysis (Miller, 1981) yields percentage and stage/age equivalent scores for the 14 grammatical morphemes and stage/age equivalent scores for simple and complex sentence development. Language production information is also available from the output of SALT.

All speech samples are conversational interchanges of experiences and ideas between the participant and the examiner. The examiner acts as both an interested listener and an interesting contributor. Interested listener activities include reacting to the participant's comments with interest, amusement, seriousness, and so forth as appropriate. Interesting contributor activities include sharing experiences and interests that will evoke spontaneous conversational comments from the participant, either because the examiner's contributions are consistent with the participant's own experience or because the examiner's comments are deliberately phrased to be in contrast to the participant's experience, thus requiring a clarification or some other response. Materials and conversational topics are changed as often as needed to keep the participant talking and to obtain representative proportions of the parts of speech, word shapes, and phonemes.

The participant's utterances are orally glossed in natural, conversational ways, with the participant given the opportunity to clarify utterances to increase intelligibility. For some studies, the examiner makes written notes on articulatory behaviors that may not be perceptible on the audio recording, such as lip rounding/unrounding gestures, unreleased stops, fricative distortions, and any facial gestures or groping behaviors that may accompany speech production. The examiner also may make written notes about the participant's general health, motivation, and physical state (e.g., whether congested, irritable, etc.) and any other aspect of the sampling session that may add to, clarify, or invalidate interpretation of the speech-prosody data.

Speech samples tend to range from 5 to 15 minutes in length and ideally contain a minimum of 100 multi-word utterances (i.e. one-word utterances excluded). The duration of the sampling period will be influenced by how verbal the participant is and by the diversity of the topics included in the conversational interchange. **Samples should be of adequate length to yield a minimum of 100 different word types (i.e., unique words). Additionally, if the sample will also be coded using the**

**PVSP and/or acoustically analyzed, the sample should contain at least 12 PVSP-codable utterances, with 6 of those utterances being 4+ words long.** There are additional guidelines at the end of this section for collecting a conversational speech sample that will be used for prosody-voice analysis.

NOTE: We discourage the use of toys during the collection of the conversational speech sample, as they can create background noises that interfere with the recording and, depending on the toy(s), may encourage the use of "unnatural" prosodic registers (character register, sound effects, etc.--see PVSP Register exclusion codes).

For young children, individual, realistic pictures of children engaged in a variety of typical daily-life activities can be used initially, as needed, to evoke conversation about topics such as birthdays; summer and winter activities; activities with mom, dad, siblings, and friends; morning, bedtime, and school routines; favorite playthings; favorite foods and eating out; activities related to pets; vacations; etc.. For children who do not respond adequately to the individual pictures or who are highly unintelligible, more explicit contextual cues can be provided via the use of books containing realistic pictures of children engaged in a variety of typical daily-life activities. Conversations with older children, adolescents, and adults can typically be accomplished easily without pictures, incorporating topics that are suitable for the participant's age and abilities (see next section).

Prior to introducing the individual pictures, the examiner identifies the conversational intent of the activity by saying, "This is the first time we have seen each other. I don't know anything about you and you don't know anything about me. Let's talk for a while. We can talk about what we like and do. Because it's sometimes hard to think of things to say when we don't know each other, I brought some pictures to help us think of things we might want to tell each other." The examiner then presents a picture and identifies a potential topic for conversation. For example, when presenting a picture of children making a snowman, the examiner might say, "This picture makes me think of all the things I like to do in the wintertime." To create the opportunity for the child to take the lead in developing the topic, the examiner pauses after introducing the picture. If the child does not respond, the examiner offers a comment from his/her own experience (real or contrived to be consistent with experiences the child may have had) and then prompts the child to respond with the question "What about you?" The child's response is pursued as a potential topic for extended conversation, whether or not it is relevant to the picture.

Throughout the conversational interchange, YES/NO and narrow-answer WH questions are used to identify topics that engage the child. Once identified, a topic is pursued with comments and open-ended questions to prompt additional commentary from the child until the topic is exhausted or a new topic has evolved. A typical interchange might include the following. The examiner asks, "Do you have a brother?" The child's "Yes" response identifies this as a possible topic for conversation. The examiner waits briefly for the child to offer comments. If the child does not make a

comment the examiner tries to suggest a general topic by saying, "I bet you do lots of things with your brother." If the child does not respond, the examiner suggests a specific topic by commenting on an experience (real or contrived) that is presumed to be consistent with the child's experience, such as "I used to play baseball with my brother." The pictures are used, as needed, to identify new topics. With some children very few pictures are required, because their comments suggest other topics that can be pursued.

Although the evocation techniques are the same whether a set of individual pictures or a book is used, when a book is used with a highly unintelligible child the examiner begins by evoking a response regarding what the child sees in the picture to create a shared referent. The examiner uses questions and comments such as "Look what's happening here," "What happened here?" and "What will happen now?" Such questions and comments are followed by questions that invite the child to comment on the activity in the picture as it relates to his/her own experience. Questions are frequently phrased to contain alternative answers from which the child can select a response (e.g., "Do you ride on a school bus or walk to school with your friends?"). This helps maximize the number of intelligible words that can be glossed, and encourages the production of multi-word utterances.

The examiner routinely orally glosses many or most of the child's utterances during the conversational sample. Although glossing is essential only for utterances that are likely to be unintelligible during later glossing and transcription, routine glossing during the assessment session establishes glossing as part of the examiner's interactive style and avoids calling attention to the child's least intelligible utterances. Techniques for glossing are best explained by example. Given the following utterance from the child, "We went to the zoo yesterday and saw an elephant," the examiner might respond with a) a reaction comment (e.g., "What fun. You went to the zoo yesterday and saw an elephant."), b) topic elaboration (e.g., "You saw an elephant at the zoo yesterday. I remember seeing some monkeys the last time I went to the zoo."), c) topic assumption (e.g., "I saw an elephant at the zoo, but not yesterday."), or d) a prompt to continue (e.g., "You saw an elephant at the zoo yesterday. What was happening?"). As above, for ease and reliability of subsequent glossing and transcription of the recorded sample, the examiner's glosses should be easily audible and consistent in volume.

### **The Phonology Project Speech Sampling Protocol for Older Children and Adults**

The use of Individual, realistic pictures of children/people engaged in a variety of typical daily-life activities that are often used initially to evoke conversation about various topics with young children should not be necessary with older children and adults. With older participants, it is helpful to have a list of potential topics that can be



referred to as necessary. Potential topics include vacations (where they've gone, what they've done and seen, favorite places, etc.); how they acquire spending money (allowances, jobs, gifts, etc.); collections (stamps, baseball cards, bottle caps, etc.); favorite sports (to play, to watch, favorite teams and players); hobbies and how they spend their free time; family (siblings, grandparents, family activities, etc.); pets (type of pet, who takes care of it and how, habits and characteristics of the pet, tricks it can perform, etc.); and their daily environment (where they live, what their bedroom/house/yard/neighborhood are like, etc.). Depending upon the age and experiences of the participant, topics related to school, employment, children, and personal/professional goals could also be introduced.

If the participant seems old enough to understand the assessment situation, at least in a general sense, the examiner could begin by saying, "I want to get a sample of your speech, so we'll just talk for a while." Then the examiner can ask a question or introduce a topic to begin the conversation. The speech sample is introduced in this way so that the older child or adult will understand that there is a purpose to the conversation and that it isn't a waste of his/her time. Even though it may seem somewhat artificial or contrived to begin this way, as long as the examiner acts and speaks in a natural, relaxed, and conversational manner, the participant usually does as well.

To ensure that the speech sample is as natural and conversational as possible, the examiner should act as both an interested listener and an interesting contributor. The examiner should react appropriately to the participant's comments and be sensitive to opportunities to contribute ideas, ask questions, and change topics without monopolizing the conversation. Because the speech sample needs to be conversational and as typical of the participant's conversational style as possible (in a prosodic sense), the examiner should *avoid* topics or situations that could result in the participant's taking on a character or narrative register (i.e., telling a story or describing a book, movie, TV show, or picture). If the participant has begun talking in a character or narrative register, the examiner should get him/her out of it as soon as possible! The examiner should also let the participant finish an utterance completely before speaking to minimize conversational interruptions and eliminate "overtalk." Words on which the examiner talks over a participant often have to be discarded from prosody-voice and acoustic analyses.

The examiner should make a mental note of any unusual or abnormal laryngeal or resonance qualities and ask the participant (or parent/caregiver) if he/she has a cold or allergies to determine possible reasons for that quality and also to determine if that quality is most likely temporary (i.e., related to the cold or other upper respiratory disturbance). It is not necessary to tell the participant why the question is being asked (don't say, "I'm asking because your voice sounds really weird today").

During the speech sample, the examiner should focus as much as possible on the conversation, not on taking notes or doing something else that could be distracting or

that could make the participant feel uncomfortable or self-conscious. The examiner can make mental notes during the conversation and written notes after the session.

An oral gloss of the speech of an older child or adult should not be necessary unless intelligibility is poor. It is helpful to gloss/repeat what the participant says if the examiner doesn't understand or needs clarification of an utterance or idea. The examiner should also gloss if he/she feels that the utterance may not be understood later from the recording.

### **Additional Guidelines for Speech Sampling for Prosody-Voice Coding**

*Portions of this section are taken from [Phonology Project Technical Report No. 7](#). For more detailed information, please refer to this technical report as well as the [PVSP manual](#).*

Valid PVSP coding results depend upon a natural conversational speech sample that contains at least 12 "codable" multiword utterances. A "codable" utterance does **not** meet the criteria for one or more *exclusion codes* in the following categories: Content/Context, Environment, Register, and States. Following are some general suggestions for maximizing the number and quality of codable utterances in a conversational speech sample.

- If the conversational sample is part of a larger assessment protocol, the examiner should complete one or more other more structured tasks first in order to make the participant more comfortable with the testing environment and with the examiner, to establish a rapport between the participant and the examiner, to determine potential topics of conversation, and to identify speech error patterns that could help the examiner verbally gloss utterances that might be partially or completely unintelligible to someone unfamiliar with the participant's errors.
- The examiner should model and maintain a natural speaking style and refrain from using a character or play register, exaggerated stress, an unusually fast or slow speech rate, etc.
- The examiner should avoid asking questions or introducing topics that could lead the participant to count or to recite the alphabet or other memorized verbal content, to say "I don't know" repeatedly, to read aloud, to sing, or to provide mostly one-word responses.
- The examiner should refrain from encouraging the participant to imitate her speech and/or prosody.
- The examiner should continuously monitor environmental noise(s) that could interfere with the recording signal and insure that a consistent lip-to-microphone distance is maintained throughout the sample.

- The examiner should avoid conversational topics or tasks that encourage the participant to whisper or to use a character register, a narrative register (describing a picture, book, movie, etc.), or sound effects.
- The participant should not eat or chew gum during the assessment session.

## References

- Ludlow, C.L., Kent, R.D., & Gray, L.C. (2018). *Measuring voice, speech, and swallowing in the clinic and laboratory*. San Diego, CA: Plural Publishing, Inc.
- McSweeny, J. (1998). *Procedures to obtain extended conversational speech samples for prosody-voice analysis*. (Tech. Rep. No. 7). Phonology Project, Waisman Center, University of Wisconsin-Madison.
- Miller, J. (1981). *Assessing language production in children: Experimental procedures*. Boston, MA: Allyn and Bacon.
- Miller, J., & Iglesias, A. (2017). *Systematic Analysis of Language Transcripts (SALT), Research Version 18* [Computer Software]. Madison, WI: SALT Software, LLC.
- Shriberg, L.D., Kent, R. D., McAllister, T., & Preston, J. L. (2019). *Clinical Phonetics*. (5th ed.). Boston, MA: Pearson Education.
- Shriberg, L. D., & Kwiatkowski, J. (1990). Self-monitoring and generalization in preschool speech-delayed children. *Language, Speech, and Hearing Services in Schools*, 21, 157–170.
- Shriberg, L.D., Kwiatkowski, J., & Rasmussen, C. (1990). *The Prosody-Voice Screening Profile*. Originally published by Communication Skill Builders, Tucson, AZ; rights returned to The Phonology Project, University of Wisconsin-Madison, Madison, WI.