

Unless otherwise noted, the publisher, which is the American Speech-Language-Hearing Association (ASHA), holds the copyright on all materials published in Perspectives on School-Based Issues, both as a compilation and as individual articles. Please see Rights and Permissions for terms and conditions of use of Perspectives content: <http://journals.asha.org/perspectives/terms.dtl>

Dynamic Assessment: Whys and Hows

Lynette Austin

Department of Communication Sciences and Disorders, Abilene Christian University
Abilene, TX

Abstract

Recent research indicates that speech-language pathologists (SLPs) typically do not implement dynamic assessment (DA) in evaluating the language skills of children from culturally and linguistically diverse backgrounds, despite recommendations from the American Speech-Language-Hearing Association (ASHA) and strong support in the literature. This article reviews different dynamic assessment formats and relevant research while making a case for DA implementation in school-based practices. Steps in designing dynamic assessments are outlined, and sample dynamic assessment write-ups from actual cases are provided.

Do you know this student? Sergei, a 2nd-grader formerly from Russia, has been referred for speech and language assessment due to teacher concerns about articulation, language skills, and academic progress. A non-English speaker on arrival to the United States at age 5, Sergei has received pull-out ESL instruction for the past 2.5 years. Recent English proficiency test scores now classify him as an intermediate speaker of English (still an English Language Learner, or ELL). Although Russian was his primary language when he entered school, his parents currently report that while Sergei still understands Russian he seems to be more comfortable speaking English. He usually responds in English to questions and comments made to him in Russian.

How are you going to test Sergei in a fair, unbiased way? What measures will help you identify how well he learns and understands language, rather than merely measure his English skills (which the English proficiency tests have already done)? How will you differentiate between the likely language differences you will see in his language skills and a true language disorder?

Let's consider typical strategies that might be employed in order to rule out a language impairment. Gillam and Peña (2004) emphasized that standardized testing has limited utility for assessing culturally and linguistically diverse (CLD) students, because results may reflect cultural or linguistic differences rather than disorders. These students typically score lower than their monolingual English-speaking peers on standardized, norm-referenced measures (Caesar & Kohler, 2007). Furthermore, normative samples usually do not include English language learners; therefore calculation of scores for these students is problematic because they are not represented in the norms. Language sampling and observation, other frequently used assessment tools, may be useful in identifying problem areas relative to the communication demands of the classroom, but again, these cannot easily identify the root cause of any difficulties noted.

Returning to our student Sergei, the speech-language pathologist (SLP) may decide to assess Russian language skills by using an interpreter/translator to translate a comprehensive

English measure into Russian. However, translation of English measures into other languages generally is ineffective, because tasks in one language often do not translate consistently into another (Caesar & Kohler, 2007). Lacking effective assessment tools and strategies, the frustrated SLP or campus referral committee may decide to wait on testing Sergei to give him more time to gain English skills before testing for language impairment. This “wait and see approach” may result in children with disabilities being denied needed services (Gillam & Peña, 2004).

The American Speech-Language-Hearing Association (ASHA) has addressed the challenge of conducting appropriate, non-biased assessment of students from CLD backgrounds, making the following clear: SLPs are to make effective use of alternative approaches to assessment including (among other tools) dynamic assessment (DA) in order to appropriately identify language impairment in children from CLD backgrounds (ASHA, 2004). However, a recent study indicated that, despite recommendations from knowledgeable leaders in the field, guidance and resources from ASHA, and (in recent years) relevant graduate coursework, SLPs tend to rely more on formal, standardized tests and scores than on any alternative procedures for determining CLD/bilingual students’ eligibility for language services (Caesar & Kohler, 2007). Authors of the study noted that, when respondents listed the formal and informal assessment tools generally used, DA was entirely absent from the list. The authors felt its absence was remarkable given that “a growing body of literature in the field of speech-language pathology . . . strongly advocates dynamic assessment (DA) as the most effective assessment procedure for evaluating CLD children (Caesar & Kohler, p. 196).” One possible reason for the limited implementation of DA is that SLPs may have difficulty envisioning how the technique may be utilized outside of research settings, in an effective and (especially) time-efficient manner. Therefore, the remainder of this article will explain briefly what DA is and provide examples of how it has been implemented in evaluation of CLD students in a school-based practice.

Definition of Dynamic Assessment

DA is “an assessment strategy that enables examiners to assess children’s language learning capabilities and their responsiveness to good language learning experiences within a very short period of time” (Gilliam & Peña, 2004, p. 2). DA has been described as a “mini” Response-to-Intervention (RTI) program, where the assumption is that use of the strategy can differentiate between two explanations for deficient performance: poor instruction (or limited opportunity for learning) versus disability (Fuchs, 2003).

Lidz and Peña (1996) described kinds of information that may be obtained via a dynamic approach to assessment. This includes insight into the child-as-learner (problem-solving strategies, error types, likelihood of self-correction) as well as insight into the modifiability of the child (to what extent the child is capable of change in response to intervention). Modifiability assessment asks the examiner to determine the level of effort required to facilitate change in the child, deciding whether that effort is greater than what would be expected from a typical learner. When a language disorder is identified, the knowledge described above should assist significantly in planning effective interventions for the student.

Two Models of DA

The research literature proposes several models of DA; two in particular have been implemented in the field of speech-language pathology. The first, the test-teach-retest method, has been described by some as the most adequate for distinguishing language differences from language disorders (Gutierrez-Clellen & Peña, 2001). Caesar and Kohler (2007) explained that test-teach-retest procedures examine children’s learning potential by providing them with mediated learning experiences (high quality instruction) following an initial testing session. The

child then is retested to identify the degree to which she or he has benefited from the instructional session(s). Theoretically, a child with good language-learning ability will benefit immediately and notably from intervention, whereas a child with a language disability will have difficulty learning even with moderate to maximum support. A large variety of different language tasks can be incorporated into this type of DA; tasks that have been employed include verbal problem-solving tasks, phonological and phonemic awareness tasks, naming tasks, and word definition tasks. In discussing task selection for DA, Hwa-Froelich, Westby, and Schommer-Aikens (2000) noted that clinicians designing a DA must use a teachable task. A number of grammatical forms, for example, can be tested, but are difficult to teach.

Generally speaking, the test-teach-retest model of DA should focus more on children's behavior during the mediated learning experience than on pre-test-to-post-test change. The extent of the child's improvement after mediation is not measured by looking at his performance compared to that of same-age peers. Rather, examiners observe how children learn to use and apply psychological tools or strategies in order to change when challenged in a learning situation. These strategies may include attention, memory, and concept formation (Peña et al., 2006).

Dynamic Assessment and Intervention: Improving Children's Narrative Abilities by Miller, Gillam, and Peña (2001) is a published test-teach-retest format for DA. The guide and kit provide a structured format for pre- and post-testing of narrative skills. The guide also outlines targeted interventions (over two intervention sessions) and gives measures for examining child learning and modifiability during the assessment. The assessment of modifiability asks the examiner to consider the intensity of examiner effort required to promote change in the child's levels of functioning compared to the amount of change that occurred (using two 5-point scales marking minimal to maximal). The authors of this tool noted that this and other test-teach-retest approaches often take longer to administer than the length of time required for a standardized test administration (Gilliam & Peña, 2004). However, recent research points to the possibility of using test-teach-retest formats reliably in "one-shot", single-session evaluations; this research will be further discussed later in this article.

A second DA model is the graduated prompting format. This method involves the presentation of a series of prompts that provide increasingly supportive cueing toward a desired response. The less assistance needed (as evidenced by fewer prompts required for success), the greater the child's modifiability (Hasson & Joffe, 2007). An example of this format was provided by Bain and Olswang (1995), who employed a prompt hierarchy in an attempt to assess children's readiness for producing two-term utterances. The prompts used in this study ranged from (minimal) indirect prompting, to (maximal) direct verbal shaping of the desired response.

Concerns About DA

Concerns have been expressed about the reliability and validity of DA (Fuchs, 2003; Hasson & Joffe, 2007). In terms of reliability, examiners need to know whether, if the assessment was repeated, similar results would occur and similar conclusions be drawn. In terms of validity, examiners must be able to affirm that the assessment truly measures what it is intended to measure—the student's language-learning ability and modifiability. Hasson and Joffe outlined these issues and other criticisms that have been leveled at DA procedures, including questions regarding inter-rater reliability, the dependence on subjective examiner judgment for determining presence of an impairment, and the highly individualized nature of interventions employed to promote responsiveness. In a discussion regarding Response-to-Intervention (RtI) assessment formats (very similar to DA) Fuchs pointed out that responsiveness (change) following intensive instruction actually may be indicative of an impairment which leads to a need for such targeted intervention. The examiner may mistakenly identify such responsiveness as indicative of no impairment.

Solutions to many of these concerns may lie in some “standardization” of procedures; that is to say, using a replicable, structured process with clear parameters while carrying out the dynamic assessment. A drawback to this solution, as Hasson and Joffe (2007) pointed out, is that using standardized/structured procedures and formats might prevent the examiner from following children’s leads, and probing areas of difficulty. However, as they also noted, establishment of clear parameters for different types of dynamic assessments could help make findings more comprehensible to others, and improve reliability. As previously mentioned, the guide (Miller et al., 2001) is a relatively standardized, replicable DA format, which lends itself to scrutiny regarding classification accuracy and reliability. However, as also noted, this particular DA task requires a significant investment of assessment time to carry out.

In discussing the issues of task design, reliability, and validity, Fuchs (2003) suggested the following: first, when designing the assessment, consideration must be given to what constitutes a good response. How much change is enough? Some criteria must be specified. Second, the nature of the intervention used for the DA must be considered. Techniques that are amenable to conceptual and procedural standardization should be employed so that results would be similar if the results were conducted by different examiners.

Current Research on DA

In recent years, researchers have begun to propose and evaluate different DA formats and tasks. Three areas have been of particular interest to researchers. The first is classification accuracy: Does DA accurately identify language disability and, just as importantly, typical language ability? Second, researchers are concerned about reliability of DA. Third, researchers have begun to investigate whether DA can reliably be conducted in brief, single-session assessments, reflecting the criticism that DA is excessively time-consuming.

Classification accuracy of DA was addressed in a study by Peña and colleagues (2006). The authors were interested in how accurately the measure, Dynamic Assessment and Intervention: Improving Children’s Narrative Abilities (Miller, Gillam, & Peña, 2001), identified children with and without language impairment. They found that the modifiability rating (examiner judgment after the post-test narrative) demonstrated adequate sensitivity (or the ability of a measure to identify true impairment). The modifiability rating correctly identified 93% of the children with language impairment. Specificity of the measure (its ability to identify typically developing individuals) was lower, at 82%. When the authors looked at a combined score comprised of post-test measures of modifiability, lexical scores, and story components scores, 100% correct classification was achieved. The authors concluded that assessment of narrative skills along with clinician judgment of modifiability after a mediated learning experience resulted in excellent, unbiased classification.

A second study (Kapantzoglou, Restrepo, & Thompson, 2008) measured the classification accuracy of a brief, one-session dynamic assessment of word-learning skills with ELLs in the test-teach-retest format. In one 30- to 40-minute session, children were taught to identify and produce novel nonwords. Researchers found that after one phase of testing, teaching, and re-testing, language-impaired students were classified with moderate accuracy via their performance (sensitivity of 76.9%, specificity of 80%). While the sensitivity of the measure is still considered somewhat low, these results provided support for the utility of brief dynamic assessments for identifying language impairments in populations of English language learners. However, the researchers also noted that with repeated teaching sessions, classification accuracy decreased with the occurrence of more false negatives (more language impaired students responded correctly). The results shed some light on the question of intensity of the intervention in a DA; apparently, repeated instruction during the intervention phase may decrease the differences in performance between the two groups, lowering identification accuracy (Kapantzoglou et al., 2008).

As researchers have studied and developed structured formats for DA, the issue of reliability has begun to be addressed systematically. Procedural integrity was examined in one study (Glaspey, 2007) that purported to assess children's learning potential for producing change in phonological skills. The studied measure used a graduated prompt format with cue levels from 0 (spontaneous) to 3 (multiple cues given from tactile to verbal). Procedural reliability of the examiners for the measure (on a sample of 8 students with phonological impairment) was 89%. The following procedural errors were identified: skipping instruction (cue 1), failing to wait for spontaneous productions (jumping to elicitation), and ending cueing at the wrong place (too early or too late). The author suggested that procedural reliability could be increased by reducing the number of cues to be given and by developing a training program incorporating videos (Glaspey).

What clinical applications can be drawn from this preliminary research into dynamic assessment? First, DA must be carefully planned and designed in order to obtain reliable results. SLPs should spend time considering what task will be tested and taught and use procedures that can be replicated. If a graduated prompting format is used, the SLP will need to administer the prompts in a consistent hierarchy in order to draw conclusions about modifiability; again, pre-planning is essential. To increase consistency and procedural integrity, practice of the techniques to be implemented would be wise. Secondly, at least one study (Kapantzoglou, Restrepo, & Thompson, 2008), as well as this author's experience, suggests that a single-session DA may more easily identify children with typical language functioning than those with language impairment. Therefore, DA alone is inadequate for accurate identification. Certainly, this is true of any one assessment measure in any evaluation; in the case of evaluation of children from CLD/bilingual backgrounds, DA should be accompanied by other recommended assessments including portfolio assessment, parent/family interviews, and first and second language sampling. (ASHA, 2004). Finally, when considering how much intervention should be provided for purposes of differentiation between language difference and language impairment, it appears that less intensive intervention may provide the support necessary for typical learners to acquire a skill. Examiners should keep in mind that positive response after significant intervention, such as repeated instructional sessions and/or multiple cues and prompts, does not rule out language impairment.

The “Hows” of DA

Planning, Designing, and Conducting Assessments

SLPs who want to implement DA in their practices must begin by selecting appropriate tasks to be utilized with different age and grade levels. According to Roseberry-McKibbin (2001), phonological awareness and vocabulary development are productive targets for English language intervention with CLD students when a language impairment has been identified. As previously noted, a variety of narrative skills also have been targeted with CLD students. Logically, it would be reasonable to select these or other teachable grade-level academic language skills as tasks to be employed in DA; as a benefit, results are directly applicable to classroom language performance. Next, the SLP should develop procedures that can be implemented with each use of the task. For example, if the first-grade task selected for a test-teach-retest format is phonological awareness—syllable and phoneme segmentation—the SLP first will need to develop a number of test items (select compound and then multisyllabic words to segment into syllables, and single-syllable words to segment into phonemes). The next step is to devise an appropriate instructional intervention. The SLP in this case could choose to use colored blocks as visual cues while modeling the segmentation tasks, for example, then plan for guided practice until the child is able to complete the tasks unassisted. Finally, a post-test should be administered that mirrors the pre-test, but utilizes different items. A note about task selection: it is sometimes possible to pull tasks from standardized measures to utilize in a DA format; the case history write-ups to follow give examples of this strategy.

If a graduated prompting format is desired, again, the SLP begins with task selection and design of test items. A consistent hierarchy (to be utilized with each item) of prompts must then be planned, ranging from least supportive to most supportive. For example, a limited-support prompt may be as simple as the clinician repeating the item or question; a maximum-support prompt may have the child imitating the correct response. For each item, the clinician will note how “high up” the hierarchy of prompts she had to go in order to obtain a correct response from the child. A note about hierarchy design: although some graduated prompt formats in the literature contain 6- to 7-level prompting hierarchies, there is no evidence that these are more effective than hierarchies with 3 prompts for the purpose of dynamic assessment. See sample case C below for an example.

Interpretation of the results obtained is the final step in conducting the DA. When utilizing the test-teach-retest format, the clinician should ask the following questions: Did the student apply metacognitive strategies in learning the new task? Did she or he attend carefully, recall steps or procedures, and grow in confidence and performance level while practicing the task? What specific teaching strategies helped: repetition, additional time to respond, or modeling? How much effort was required on the part of the examiner to teach the task—minimum, moderate, maximum? If utilizing graduated prompts, the SLP will want to consider whether or not the student grew more skilled and confident at the task, as evidenced by decreased need for cuing.

The following are samples of actual dynamic assessment sessions conducted by the author. Both formats are demonstrated and interpretation of the data is provided. Readers should bear in mind that these were conducted in the context of a comprehensive evaluation utilizing multiple measures before determination of the presence or absence of a language impairment was made. It is hoped that these sample cases, along with the recommendations above, will provide school-based SLPs with a clear picture of how easily (with some planning) dynamic assessment can be incorporated into daily clinical practice.

Sample DA: Case A

The student has good language-learning ability (1st grader). The model is Model 1: Test-Teach-Retest.

Tasks were taken from Test of Problem Solving—Revised (Bowers, Barrett, Huisinigh, Orman, & LoGiudice, 1994). Portions were administered in DA format in order to determine how well Antonia was able to learn a new language-based task in his dominant language, English.

Antonia was able to answer 5 of 15 problem-solving questions correctly (33%) during the pre-test.

During the mediated learning experience, Antonia was instructed to look at missed scenarios and ask himself the following questions: Who is involved, what are they doing, and why are they doing it. All three previously administered scenarios were reviewed; the examiner modeled and prompted the question/answer strategy for the first two scenarios and then Antonia independently completed the question/answer task for the third scenario.

In the post-test, Antonia was able to answer 14 of 18 similar contextual problem solving questions correctly (78%), a significant improvement over the pretest performance. Moderate examiner effort was required to teach the questioning technique

Results indicate that Antonia is able to use language effectively to learn new skills.

Sample DA: Case B

The student has language-learning difficulties (kindergartener). The model is Model 1: Test-Teach-Retest.

Ahmed's language-learning ability was evaluated using DA procedures. Two types of tasks were utilized: phonemic awareness (identification of first sound) and segmentation of compound words into two parts. The tasks were selected because they form part of the kindergarten curriculum and should have been familiar to Ahmed.

In the pretest, Ahmed was given 5 trials on each task. He was not successful in any trial.

During the MLE portion, Ahmed was instructed using visual cues and manipulatives to identify the first sound of several CVC words. He was also instructed in segmentation of compound words using manipulatives. When Ahmed was able to give correct responses to the tasks, he then was tested to see if he could generalize the new skills. Significant examiner effort was required to obtain a correct, prompted response.

In neither case was he able to sustain the new skill, yielding zero correct responses when the examiner eliminated prompts. Significant examiner effort yielded little progress, indicating that Ahmed had difficulty learning new, language-based skills in his dominant language, English.

Sample DA: Case C

The student has good language-learning ability (3rd grader). The model is Model 2: Graduated Prompting.

Tasks from the Test of Problem Solving-3 (Bowers, Huisinigh, & LoGiudice, 2005) were used.

During the pre-test, Diego viewed 3 test scenarios and answered the related questions. He was easily able to answer questions requiring him to infer information not directly stated (3 of 3 correct). He demonstrated some difficulty with cause determination (2 of 3 correct), negative WH questions (1 of 2 correct), solutions (0 of 2 correct), and problem avoidance (0 of 2 correct). His total score on the pre-test was 6 out of 12 (50%).

The examiner provided prompts of varying levels to support Diego's ability to answer the questions. Level 1 prompts, slower repetition of item; Level 2 prompts, breaking complex sentences into simpler structures; and Level 3 prompts, explaining vocabulary or concepts.

Diego was able to respond correctly to 4 of 6 missed items with minimal prompting (Level 1, slower repetition). He required a level 2 prompt to answer correctly on one item requesting him to explain how to avoid a future problem. On one item he was unable to improve his response (this item required comprehension of the English subjunctive mood: "could have"). Total correct responses with minimal prompting: 11 of 12 correct.

On this measure, Diego responded very favorably to minimal support, improving his performance significantly. He demonstrated high modifiability with little examiner effort. His rapid improvement with minimal scaffolding was more reflective of a student with typical linguistic ability rather than a student with language impairment.

References

- American Speech-Language-Hearing Association. (2004). *Knowledge and Skills Needed by Speech-Language Pathologists and Audiologists to Provide Culturally and Linguistically Appropriate Services* [Knowledge and Skills]. Available from www.asha.org/policy
- Bain, B., & Olswang, L. (1995). Examining readiness for learning two-word utterances by children with specific expressive language impairment: Dynamic assessment validation. *American Journal of Speech-Language Pathology*, 4(1), 81-91.
- Bowers, L., Barrett, M., Huisinigh, R., Orman, J., & LoGiudice, C. (1994). *Test of Problem Solving—Revised*. East Moline, IL: Linguisystems.

- Bowers, L., Huisinigh, R., & LoGiudice, C. (2005). *Test of Problem Solving—3*. East Moline, IL: Linguisystems.
- Caesar, L., & Kohler, P. (2007). The state of school-based bilingual assessment: Actual practice versus recommended guidelines. *Language, Speech and Hearing Services in Schools*, 38, 190-200.
- Fuchs, L. (2003). Assessing intervention responsiveness: Conceptual and technical issues. *Learning Disabilities Research & Practice*, 18(3), 172-186.
- Gillam, R. B., & Peña, E. (2004). Dynamic assessment of children from culturally diverse backgrounds. *Perspectives on Language Learning and Education*, 11, 2-5.
- Glaspey, A. (2007, November). *Procedural reliability of dynamic assessment in phonological disorders*. Poster session presented at the annual convention of the American Speech-Language-Hearing Association, Boston, MA.
- Gutierrez-Clellen, V. F., & Peña, E. (2001). Dynamic assessment of diverse children: A tutorial. *Language, Speech & Hearing Services in the Schools*, 32(4), 212-224.
- Hasson, N., & Joffe, V. (2007). The case for dynamic assessment in speech and language therapy. *Child Language Teaching and Therapy*, 23(1), 9-25.
- Hwa-Froelich, D. A., Westby, C. E., & Schommer-Aikens, M. (2000). Assessing language learnability. *Perspectives on Communication Disorders and Sciences in Culturally and Linguistically Diverse Populations*, 6(3), 1-6.
- Kapantzoglou, M., Restrepo, M. & Thompson, M. (2008, November). *Can dynamic assessment of word learning skills differentiate English language learners with and without language impairment?* Poster session presented at the annual convention of the American Speech-Language-Hearing Association, Chicago, IL.
- Lidz, C., & Peña, E. (1996). Dynamic assessment: The model, its relevance as a nonbiased approach, and its application to Latino American preschool children. *Language, Speech and Hearing Services in Schools*, 27, 367-372.
- Miller, L., Gillam, R. B., & Peña, E. D. (2001). *Dynamic assessment and intervention: Improving children's narrative skills*. Austin, TX: PRO-ED.
- Peña, E., Gilliam, R. B., Malek, M., Ruiz-Felter, R., Resendiz, M., Fiestas, C. & Sabel, T. (2006). Dynamic assessment of school-age children's narrative ability: An experimental investigation of classification accuracy. *Journal of Speech, Language and Hearing Research*, 49, 1037-1057.
- Roseberry-McKibbin, C. (2001). *The source for bilingual students with language disorders*. East Moline, IL: Linguisystems.