

Volume 14, Issue 4, August 2020

# EBBP *briefs*

A scholarly forum for guiding evidence-based  
practices in speech-language pathology

THE USE OF DYNAMIC ASSESSMENT IN THE  
EVALUATION OF EARLY LITERACY SKILLS

KELLEY NELSON-STROUTS, MA, CCC-SLP  
UNIVERSITY OF KANSAS

W. MATTHEW GILLISPIE, PhD, CCC-SLP  
UNIVERSITY OF KANSAS

MINDY BRIDGES, PhD, CCC-SLP  
UNIVERSITY OF KANSAS MEDICAL CENTER

# EBP Briefs

---

## Editor

Mary Beth Schmitt  
*Texas Tech University Health Sciences Center*

## Editorial Review Board

Kelly Farquharson  
*Emerson College*

Erin Bush  
*University of Wyoming*

Angela Van Sickle  
*Texas Tech University Health Sciences Center*

Lisa Bowers  
*University of Arkansas*

Sherine Tambyraja  
*The Ohio State University*

## Managing Director

Tina Eichstadt  
Pearson  
5601 Green Valley Drive  
Bloomington, MN 55437

Cite this document as:

Nelson-Strouts, K., Gillispie, W. M., & Bridges, M. (2020). The use of dynamic assessment in the evaluation of early literacy skills. *EBP Briefs*, 14(4), 1–8.

---



For inquiries or reordering:  
800.627.7271  
[www.PearsonAssessments.com](http://www.PearsonAssessments.com)

Copyright © 2020 NCS Pearson, Inc. All rights reserved.

**Warning:** No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without the express written permission of the copyright owner.

**Pearson** and **WRMT** are trademarks, in the U.S. and/or other countries, of Pearson PLC or its affiliates. **DIBELS** and **Dynamic Indicators for Basic Early Literacy Skills** are registered trademarks of the University of Oregon.

**NCS Pearson, Inc. 5601 Green Valley Drive Bloomington, MN 55437**

Produced in the United States of America.

1.A

## Structured Abstract

**Clinical Question:** When assessing a kindergarten student in the area of early reading skills, does dynamic assessment provide additional classification accuracy over and above static, standardized assessments when determining if the student exhibits a difference in early learning experiences or a true disorder?

**Method:** Systematic Review

**Study Sources:** PsycINFO, ERIC, ASHAWire, forward search of previously selected articles

**Search Terms:** dynamic assessment OR dynamic test\* AND reading OR phonological awareness OR phonemic awareness

**Number of Studies Included:** 7

### Primary Results:

Dynamic assessment was found to add variance over and above static, standardized assessments to the prediction of students' performance on oral reading fluency, word identification, and letter/sound deletion tasks.

Dynamic assessment consistently increased the classification accuracy when used alone or in combination with static, standardized assessments of early literacy. Dynamic assessment consistently met acceptable criteria for sensitivity, specificity, and area under the curve.

Most authors of dynamic assessments of early literacy did not provide clear cut points or other guidelines for interpretation of scores. When provided, appropriate cut points varied based on the group being tested.

Rarely did studies of dynamic assessments of early reading include measures of modifiability, which has been shown to further improve DA's classification accuracy.

**Conclusions:** Six of the seven review articles showed either compelling or suggestive evidence for the use of dynamic assessment in the screening and evaluation of early literacy disorders of younger elementary-age students. Only one article was found to have equivocal evidence. No articles found strong evidence against dynamic assessment. Although dynamic assessments are not as straightforward to design and administer as static, standardized assessments, with support from the literature, clinicians can feel confident in choosing to incorporate them into evaluations of early literacy skills.

# The Use of Dynamic Assessment in the Evaluation of Early Literacy Skills

**Kelley Nelson-Strouts, MA, CCC-SLP**  
**University of Kansas**

**W. Matthew Gillispie, PhD, CCC-SLP**  
**University of Kansas**

**Mindy Bridges, PhD, CCC-SLP**  
**University of Kansas Medical Center**

---

## Clinical Scenario

Julia is a speech-language pathologist (SLP) in her fifth year of working in an elementary school in a Midwestern suburban community. In January, Julia is approached by Mrs. Butler, one of the kindergarten teachers in her building, with concerns that her student, Will, does not seem to be showing progress in skills related to later reading ability. She notes Will struggles to identify letters, knows very few letter/sound matches, and cannot complete activities asking him to segment or rearrange sounds in words. Mrs. Butler recognizes the strong connection between language skills and reading abilities and is hopeful that Julia can provide insight into Will's situation. Mrs. Butler wonders if Will would benefit from individualized services to help him achieve these crucial early literacy skills.

Julia obtains consent for an evaluation from Will's mother. At that time, his mother shares that she has also been worried about Will's knowledge of letters and sounds. She says although none of her children attended preschool, Will's older siblings seemed to do better on reading activities by this point in kindergarten. She is committed to helping him learn to be a good reader. His mother confirms Will has never been evaluated for educational services before. Julia knows from reviewing his records that he qualifies for free and reduced lunch. Based on the Individual with Disabilities Education Act (IDEA; 2004), Julia knows she will need to evaluate Will in a way that takes into account any potential differences in his early learning experiences.

In her previous speech and language assessments, Julia primarily relied on standardized, norm-referenced measures supplied by her district to help her make diagnostic decisions. However, Julia recently attended a professional development session focusing on dynamic assessment (DA); she wonders if DA would be beneficial in her evaluation of Will's early literacy skills.

## Background Information

IDEA requires special education evaluations to discriminate between students with true disorders/disabilities and students who display the effects of differences in their learning experiences (IDEA, 2004). The American Speech-Language-Hearing Association (ASHA, n.d.) further recommends that SLPs use multiple sources of assessment to accomplish this successfully (ASHA, 2004). SLPs must become familiar with alternate assessment formats, including DA, but continue to rely on traditional evaluation formats over and above alternate procedures (Betz et al., 2013; Caesar & Kohler, 2009; Fulcher-Rood et al., 2019). Unfortunately, the psychometric properties of many commercially available standardized assessments rarely meet the minimum standards of validity (e.g., Bogue et al., 2014; Daub et al., 2019; Plante & Vance, 1994).

For even well-constructed, norm-referenced assessments to be clinically relevant, the student for whom the assessment is being used must be represented within the normative sample. Students such as those from lower socioeconomic status (SES) households or culturally or linguistically diverse (CLD) backgrounds are often not sampled in adequate numbers during the norming process (Austin, 2010). In fact, many static, standardized assessments (SSAs) have been found to be biased against the early learning experiences of students from lower SES backgrounds (e.g., Roseberry-McKibbin, 2007). Primarily relying on such measures to make diagnostic decisions for low-SES students contributes to overrepresenting this population in special education.

SSAs tell the assessor what the child already knows at the moment of assessment. DA, conversely, is based on social constructivist theory and has the potential to demonstrate how well a child responds to instruction or additional supports (Lidz & Peña, 1996). DA can take

the structure of a single round of test, teach, retest—often referred to as the sandwich model because instruction occurs between the traditional tests. The teaching here is referred to as the mediated learning experience (MLE). In this model, pretests and posttests provide information on the child's change in knowledge whereas the MLE provides information on what levels of support were needed for the child to obtain this change. DA also can be an on-going process, often called the cake model, where the testing and given levels of prompting form layers as the child progresses through the tasks (Grigorenko, 2009).

DA has been shown to be successful in assessing narratives (Henderson et al., 2018; Peña et al., 2006), vocabulary (e.g., Peña et al., 2001), and phonological awareness (e.g., Bridges & Catts, 2011). In the past two decades, authors have also demonstrated DA's particular effectiveness in appropriately assessing CLD students (Gillam & Peña, 2004; Laing & Kamhi, 2003) and those with other disabilities such as hearing loss (Mann et al., 2015). On the whole, research on the use of DA within the realm of speech-language pathology is still in the early stages.

## Clinical Question

Julia put herself at the center of the PICO framework (population, intervention, comparison, outcome; Straus & Sackett, 1998) to construct a clear clinical question.

- P – elementary school speech-language pathologist
- I – using dynamic assessment
- C – using standardized assessments
- O – appropriately identifying whether a student is exhibiting a difference in early learning experiences or a true early reading disorder

The proposed clinical question is: When assessing a kindergarten student in the area of early reading, does dynamic assessment provide additional classification accuracy over and above static, standardized assessments in determining if the student exhibits a difference in early learning experiences or a true disorder?

## Search for the Evidence

Julia began her search by establishing the following selection criteria for all studies returned:

## Inclusion and Exclusion Criteria

- Because of differences in early childhood education and kindergarten expectations across countries, studies had to be completed within the United States
- Participants had to be in early elementary school (Grades K–2)
- Participants had to be native English speakers
  - If bilingual students were included in the study, separate analysis had to be reported for a monolingual, English-speaking sample
- Participants could not have known additional disabilities
- Studies had to be published within the last 15 years (2004–present)
- The DA(s) had to focus on early literacy skills such as phonemic awareness and/or phonological awareness
  - Reading comprehension or other related language skills such as vocabulary could not be the primary target of the DA(s)
- DA(s) could be designed for either diagnostic or screening purposes
- Studies must have been peer-reviewed
- Studies had to be prospective and use at least one reference standard as a comparison to the DA(s) proposed

## Search Strategy

Julia selected three databases in which to conduct her search: PsycINFO, ERIC, and the ASHAWire. Her search terms were dynamic assessment OR dynamic test\* AND reading OR phonological awareness OR phonemic awareness. Her search initially returned 96 articles through the PsycINFO database; she kept seven after reviewing article summary pages and abstracts and applying the selection criteria (see Figure 1). The ERIC database initially returned 64 articles, and the ASHAWire database yielded an initial return of 186 articles. None of the ERIC and ASHA articles provided any additional studies for consideration. Julia then completed a forward search of the seven selected articles that failed to yield additional articles for review.

## Evaluating the Evidence

To evaluate the evidence, Julia borrowed aspects of the Critical Appraisal of Diagnostic Evidence (CADE) system outlined by Dollaghan (2007, pp. 81–104). Julia examined the reference standards each study used, the measures of classification accuracy they reported, whether any modifiability scales were included, the reported reliability of the DA, and whether the authors reported that the DA added variance above SSAs.

Specifically, Julia examined whether the studies included reference standards to compare with the DA. Reference standards are assessment procedures that are considered to be the best or most commonly used practice in a given area of interest. Some of the reference standards from the studies she chose included the Letter Name Fluency (LNF) subtest of the Dynamic Indicators for Basic Early Literacy Skills (DIBELS), the Word Attack and Word Identification subtests of the Woodcock Reading Mastery Tests-Revised (WRMT-R), and the Test of Word Reading Efficiency-2 (TOWRE-2). According to CADE, it is best practice to apply the proposed diagnostic procedure and the reference standard to all participants. Each study Julia included for her review had a clearly defined reference standard, and all participants in each study received the reference standards.

Additionally, CADE encourages the formulation of a diagnostic question with an outcome concerned with classification accuracy. This matched well with the PICO question Julia already devised, and she attempted to identify in each study at least one quantifiable measure of classification accuracy. Five articles provided such values. Examples of classification accuracy identified in the studies include sensitivity, specificity, and area under the curve (AUC). Sensitivity represents the proportion of individuals who truly have the disorder and test positive for a disorder on the assessment measure. Specificity is the reverse: the proportion of individuals who do not have the disorder and test negative for the disorder (Dollaghan, 2007). A measure is said to have acceptable sensitivity and specificity if it achieves 80% for each (Plante & Vance, 1994). The studies included in Julia's review that reported sensitivity values had a range of 80%–90.7%. Those that included specificity values reported between 63%–87.7%. AUC is another measure that quantifies a given assessment's ability to correctly classify true positives from false positives; values between .70 and .80 are acceptable, and values above .80 are

considered good (Compton et al., 2010). Of the four studies that reported AUC, the values ranged from .70 up to .953, indicating all were acceptable or good. One study (Cho et al., 2014) did not provide sensitivity, specificity, or AUC values but did report prediction of linear growth models to help explain DA's success in identifying students who would do well in Tier 2 instruction from those who would likely continue to struggle at that level of remediation.

Modifiability scores have been shown to provide additional information beyond the pretest and posttest scores for DA (Henderson et al., 2018; Mann et al., 2015). In essence, modifiability scores represent a way for the assessor to quantify the ease or difficulty with which the child learned the skill(s) taught during the MLE. In this way, modifiability can help differentiate students who score the same on the posttest but require varied levels of supports to achieve those scores. Modifiability is most often measured by Likert ratings of items such as child distractibility, degree of examiner effort, and transferability of skills. Of the studies reviewed, only two (Peterson et al., 2016; Peterson et al., 2018) included a modifiability scale. In these, the scores obtained via the modifiability scales were factored into their overall DA scores, allowing modifiability to be integrated into the calculations of classification accuracy and added variance of the DA overall.

Another appraisal point of interest as outlined by CADE is the reported reliability of each study's DA(s). No test can ever achieve validity without reliability. Three studies (Bridges & Catts, 2011; Cho et al., 2017; Peterson et al., 2016) reported good reliability values ( $\geq .90$ ) using inter-rater reliability coefficients. The other four studies (Cho et al., 2014; Compton et al., 2010; Fuchs et al., 2011; Peterson et al., 2018) presented evidence of 90% or greater examiner fidelity, which indicates the individuals administering the DA could consistently adhere to administration procedures. However, while strong examiner fidelity values like these would certainly aid the overall reliability score, they cannot be considered satisfactory stand-ins for true reliability values.

Finally, CADE asks evaluators to provide an overall judgment as to whether the evidence is compelling, suggestive, or equivocal (see Dollaghan, 2007, for specific descriptors of each). This rating provides a summation or "clinical bottom line" (p. 152) of each study and a way to compare diagnostic studies to one another. Overall, Julia rated three studies as compelling, three as suggestive, and one as equivocal (see Table 1). These findings strongly

support Julia's idea to incorporate DA into her evaluation of Will.

## The Evidence-Based Decision

After reviewing the evidence, Julia felt more confident in choosing to incorporate DA into Will's assessment plan. The studies she identified consistently showed DA provides evidence of classification accuracy in the area of early literacy. However, Julia felt unsure of how to best construct this type of assessment. She returned to the seven studies for ideas. Each study used its own author-created DA and nearly all provided copies of the DAs within the article appendixes or gave very detailed descriptions of the type and format of the DA questions. Many also provided guidelines on the MLE or an acceptable prompting sequence. These examples provided Julia with clear ideas of what to include in her own DA of early literacy skills.

While she was reviewing the DAs, Julia began to see additional benefits a DA might afford her as compared to SSAs. She was particularly drawn to the idea that DA sheds light on what her students can learn with supports instead of merely informing her of their present knowledge. For instance, Bridges & Catts (2011) walks the clinician through levels of prompting allowed during administration of the DA. Understanding how to modify administrations to suit each child is useful for clinicians. For example, Julia might learn one child is successful after simply repeating directions or a model, but another child might need picture supports to provide an accurate response. Knowing under which circumstances a student performs best will help her to plan interventions or make recommendations to other educators.

Although all studies gave examples of their DAs, very few provided cutoff scores for clinicians to use in their interpretation of DA results. Although this puts an additional burden on the clinician, cutoff scores vary dependent upon the population with which the DA will be used. For example, the Peterson et al. (2018) study demonstrated the cutoff score that was appropriate for the Caucasian sample did not meet adequate sensitivity/specificity for the Hispanic sample. This same study also used a measure of modifiability in the clinical decision-making, which proved to add significantly to the predictive validity of their DA. Such measures of modifiability help measure the child's effort to reach a given degree of learning and can be clinically informative.

Julia knows she will need to consider the general trends of Will's performance on the DA. For example, if Will makes strong gains from pretest to posttest and achieves a high modifiability score, Julia would feel more confident in recommending short-term general education interventions for Will. Conversely, if Will's pretest to posttest gains are meager and his modifiability score is low, Julia would be more inclined to suspect Will has an early literacy disorder and recommend individualized services to best address his needs.

Although a DA can provide guidance in the diagnostic decision-making process as outlined in the above examples, no single assessment method should ever stand alone when considering whether to classify a child as having a disorder or disability (ASHA, 2004). The information garnered from a DA is intended to be synthesized into the evaluation as a whole. This idea was supported consistently throughout the seven studies Julia reviewed. Julia now sees how DA can play an integral and important part in any comprehensive evaluation and although research on the use of DA in many areas of speech-language pathology is still emerging, Julia feels assured by the level of support she found for DAs of early literacy.

## Authors' Note

Kelley Nelson-Strouts is a second-year doctoral student with a primary research interest in how to best evaluate nonmainstream students in the area of reading disorders. She received her MA from the University of Kansas in 2016. Before returning to KU to pursue her PhD, she worked as school-based certified speech-language pathologist in the Kansas City area.

Dr. W. Matthew Gillispie is a clinical associate professor and speech-language pathologist at the University of Kansas. His clinical and scholarship interests include the relationship between language and literacy development and clinical considerations when working with individuals and families from underrepresented backgrounds.

Dr. Mindy Bridges is an assistant professor in the Hearing and Speech Department at the University of Kansas Medical Center; she is also a certified speech-language pathologist. Her research interests include the relationship between language and reading development and disorders as well as the use of dynamic assessment to assess early literacy skills.

## References

- American Speech-Language-Hearing Association. (n.d.). *Dynamic assessment*. Retrieved March 20, 2019, from <https://www.asha.org/practice/multicultural/issues/Dynamic-Assessment/>
- American Speech-Language-Hearing Association. (2004). *Preferred practice patterns for the profession of speech-language pathology* [Preferred practice patterns]. <https://www.asha.org/policy/PP2004-00191/>
- Austin, L. (2010). Dynamic assessment: Whys and hows. *Perspectives on School-Based Issues, 11*(3), 80–87. <https://doi.org/10.1044/sbi11.3.80>
- Betz, S. K., Eickhoff, J. R., & Sullivan, S. F. (2013). Factors influencing the selection of standardized tests for the diagnosis of specific language impairment. *Language, Speech, and Hearing Services in Schools, 44*(2), 133–146. [https://doi.org/10.1044/0161-1461\(2012/12-0093\)](https://doi.org/10.1044/0161-1461(2012/12-0093))
- Bogue, E. L., DeThorne, L. S., & Schaefer, B. A. (2014). A psychometric analysis of childhood vocabulary tests. *Contemporary Issues in Communication Science and Disorders, 41*(Spring), 55–69. [https://doi.org/10.1044/cicsd\\_41\\_S\\_55](https://doi.org/10.1044/cicsd_41_S_55)
- Bridges, M. S., & Catts, H. W. (2011). The use of a dynamic screening of phonological awareness to predict risk for reading disabilities in kindergarten children. *Journal of Learning Disabilities, 44*(4), 330–338. <https://doi.org/10.1177/0022219411407863>
- Caesar, L. G., & Kohler, P. D. (2009). Tools clinicians use: A survey of language assessment procedures used by school-based speech-language pathologists. *Communication Disorders Quarterly, 30*(4), 226–236. <https://doi.org/10.1177/1525740108326334>
- Cho, E., Compton, D. L., Fuchs, D., Fuchs, L. S., & Bouton, B. (2014). Examining the predictive validity of a dynamic assessment of decoding to forecast response to Tier 2 intervention. *Journal of Learning Disabilities, 47*(5), 409–423. <https://doi.org/10.1177/0022219412466703>
- Cho, E., Compton, D. L., Gilbert, J. K., Steacy, L. M., Collins, A. A., & Lindström, E. R. (2017). Development of first-graders' word reading skills: For whom can dynamic assessment tell us more? *Journal of Learning Disabilities, 50*(1), 95–112. <https://doi.org/10.1177/0022219415599343>
- Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., Cho, E., & Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology, 102*(2), 327–340. <https://doi.org/10.1037/a0018448>
- Daub, O., Skarakis-Doyle, E., Bagatto, M. P., Johnson, A. M., & Cardy, J. O. (2019). A comment on test validation: The importance of the clinical perspective. *American Journal of Speech-Language Pathology, 28*(1), 204–210. [https://doi.org/10.1044/2018\\_AJSLP-18-0048](https://doi.org/10.1044/2018_AJSLP-18-0048)
- Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders*. Paul H. Brookes Publishing Co.
- Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B., & Caffrey, E. (2011). The construct and predictive validity of a dynamic assessment of young children learning to read: Implications for RTI frameworks. *Journal of Learning Disabilities, 44*(4), 339–347. <https://doi.org/10.1177/0022219411407864>
- Fulcher-Rood, K., Castilla-Earls, A., & Higginbotham, J. (2019). Diagnostic decisions in child language assessment: Findings from a case review assessment task. *Language, Speech, and Hearing Services in Schools, 50*(3), 385–398. [https://doi.org/10.1044/2019\\_LSHSS-18-0044](https://doi.org/10.1044/2019_LSHSS-18-0044)
- Gillam, R. B., & Peña, E. D. (2004). Dynamic assessment of children from culturally diverse backgrounds. *Perspectives on Communication Disorders and Sciences in Culturally and Linguistically Diverse Populations, 11*(2), 2–5. <https://doi.org/10.1044/cds11.2.2>
- Good, R. H., & Kaminski, R. A. (2010). *Dynamic Indicators of Basic Early Literacy Skills* (7th ed.). Dynamic Measurement Group.



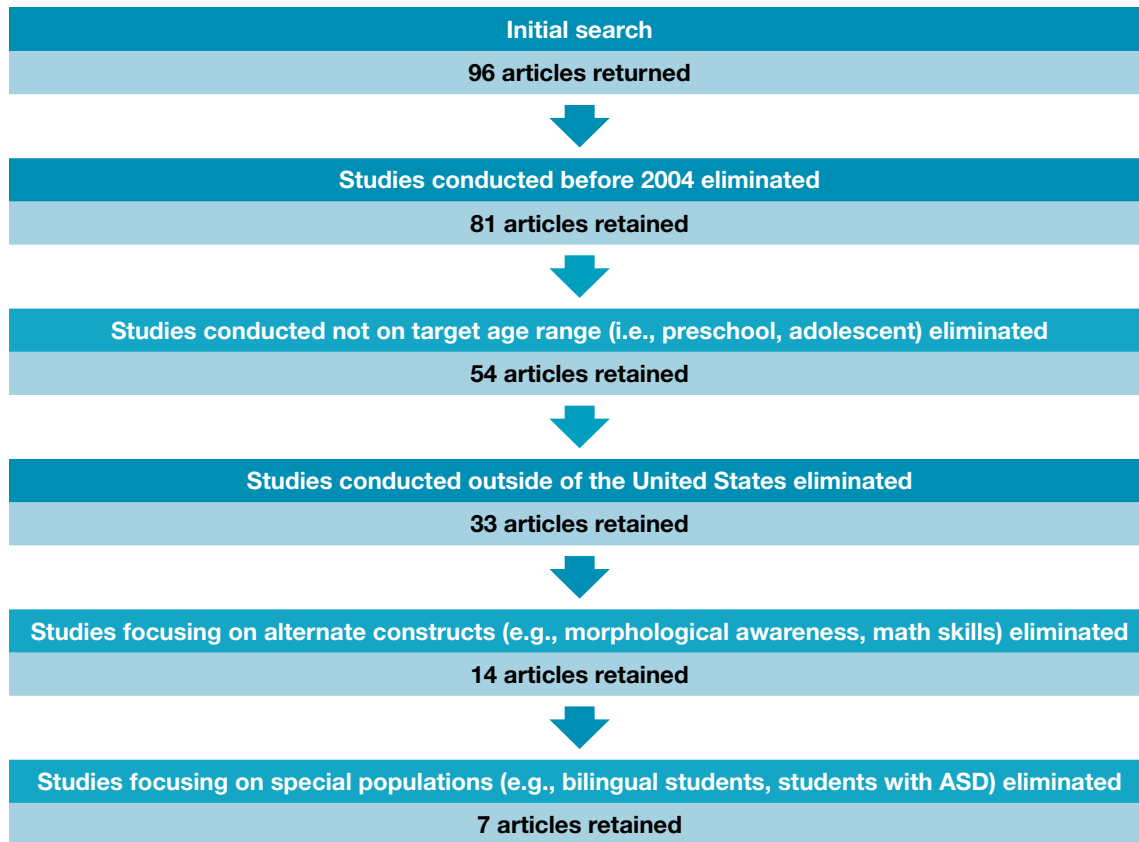
- Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities, 42*(2), 111–132. <https://doi.org/10.1177/0022219408326207>
- Henderson, D. E., Restrepo, M. A., & Aiken, L. S. (2018). Dynamic assessment of narratives among Navajo preschoolers. *Journal of Speech, Language, and Hearing Research, 61*(10), 2547–2560. [https://doi.org/10.1044/2018\\_JSLHR-L-17-0313](https://doi.org/10.1044/2018_JSLHR-L-17-0313)
- Individuals with Disabilities Education Act. (2004). *Evaluations, eligibility determinations, individualized education programs, and educational placements* [Statutes and regulations]. <https://sites.ed.gov/idea/statute-chapter-33/subchapter-ii/1414>
- Laing, S. P., & Kamhi, A. (2003). Alternative assessment of language and literacy in culturally and linguistically diverse populations. *Language, Speech, and Hearing Services in Schools, 34*(1), 44–55. [https://doi.org/10.1044/0161-1461\(2003/005\)](https://doi.org/10.1044/0161-1461(2003/005))
- Lidz, C. S., & Peña, E. D. (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*(4), 367–372. <https://doi.org/10.1044/0161-1461.2704.367>
- Mann, W., Peña, E. D., & Morgan, G. (2015). Child modifiability as a predictor of language abilities in deaf children who use American Sign Language. *American Journal of Speech-Language Pathology, 24*(3), 374–385. [https://doi.org/10.1044/2015\\_AJSLP-14-0072](https://doi.org/10.1044/2015_AJSLP-14-0072)
- Peña, E. D., Gillam, 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*(5), 1037–1057. [https://doi.org/10.1044/1092-4388\(2006/074\)](https://doi.org/10.1044/1092-4388(2006/074))
- Peña, E., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology, 10*(2), 138–154. [https://doi.org/10.1044/1058-0360\(2001/014\)](https://doi.org/10.1044/1058-0360(2001/014))
- Petersen, D. B., Allen, M. M., & Spencer, T. D. (2016). Predicting reading difficulty in first grade using dynamic assessment of decoding in early kindergarten: A large-scale longitudinal study. *Journal of Learning Disabilities, 49*(2), 200–215. <https://doi.org/10.1177/0022219414538518>
- Petersen, D. B., Gragg, S. L., & Spencer, T. D. (2018). Predicting reading problems 6 years into the future: Dynamic assessment reduces bias and increases classification accuracy. *Language, Speech, and Hearing Services in Schools, 49*(4), 875–888. [https://doi.org/10.1044/2018\\_LSHSS-DYSLC-18-0021](https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0021)
- Plante, E., & Vance, R. (1994). Selection of preschool language tests. *Language, Speech, and Hearing Services in Schools, 25*(1), 15–24. <https://doi.org/10.1044/0161-1461.2501.15>
- Roseberry-McKibbin, C. (2007). Assessment and intervention guidelines for service delivery to low-SES children. *Perspectives on School-Based Issues, 8*(3), 4–9. <https://doi.org/10.1044/sbi8.3.4>
- Straus, S. E., & Sackett, D. L. (1998). Using research findings in clinical practice. *BMJ, 317*(7154), 339–342. <https://doi.org/10.1136/bmj.317.7154.339>
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of Word Reading Efficiency-Second Edition (TOWRE-2)*. Pro-Ed.
- Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests—Revised/Normative Update*. American Guidance Service.

*Table 1. Appraisal of Diagnostic Evidence*

	Cho et al. (2017)	Bridges & Catts (2011)	Petersen et al. (2018)	Fuchs et al. (2011)	Petersen, Allen, & Spencer (2016)	Cho et al. (2014)	Compton et al. (2010)
What reference standard(s) was the DA compared to?	CTOPP Elision & RDN; WRMT-R/NU WA & Word ID	DIBELS ISF or author-developed Static Deletion Task (SDT)	DIBELS FSF & LNF	CTOPP RLN & RDN, WRMT-R/NU WA, author-developed measures of decoding	DIBELS LNF, FSF, & Composite; TOWRE SWE & PDE	CTOPP RLN & SM; WRMT-R/NU WA & Word ID; TOWRE PDE	TOWRE SWE & PDE; WRMT-R/NU PC, Word ID, & WA; ORF; District Running Records
Were all participants given the DA and the reference standard(s)?	Yes	Yes, half given DIBELS and half given SDT	Yes; groups differed on DA teaching methods	Yes	Yes; groups differed between DA teaching methods	Yes	Yes
Did the authors report measures of classification accuracy?	No	Yes, DA provided sensitivity 85%–86% and 63% specificity AUC > .70 for all models including DA	Yes, DA provided sensitivity and specificity > 80% AUC = .87 for Grade 2 Caucasian group	Not reported	Yes, DA provided sensitivity and specificity > 80% AUC = .83, .85 for the two DA teaching methods	Yes, DA significantly predicted linear growth during Tier 2 intervention	Yes, DA in extension model provided 90.7% sensitivity; 87.7% specificity AUC = .953
Did the DA include a measure of modifiability?	No	No	Yes	No	Yes	No	No
Did the authors report reliability?	Yes, inter-rater > .90	Yes, inter-rater reliability = .98	No, the authors only reported examiner fidelity	No, the authors only reported examiner fidelity	Yes, inter-rater = .96 (excellent)	No, the authors only reported examiner fidelity	No, the authors only reported examiner fidelity
Amount of variance added by the DA above and beyond the reference standard(s):	3.49% for Word ID	4%–9% above SDT; 2%–5% above DIBELS ISF	2%–7% for Grades 2–5 ORF	2.3% for Word ID	Not reported	3%–13%	Not reported
Advantage of using DA based on this study:	Suggestive	Suggestive	Compelling	Equivocal	Compelling	Suggestive	Compelling

*Note.* CTOPP = Comprehensive Test of Phonological Processing; RDN = Rapid Digit Naming; RLN = Rapid Letter Naming; SM = Sound Matching; WRMT-R/NU = Woodcock Reading Mastery Tests-Revised/Normative Update; WA = Word Attack; Word ID = Word Identification; PC = Passage Comprehension; DIBELS = Dynamic Indicators of Basic Literacy Skills; ISF = Initial Sound Fluency; FSF = Final Sound Fluency; LNF = Letter Name Fluency; TOWRE = Test of Sight Word Efficiency; SWE = Sight Word Efficiency; PDE = Phonemic Decoding Efficiency; SDT = Static Deletion Task; ORF = Oral Reading Fluency; DA = dynamic assessment; AUC = area under the curve.

*Figure 1. PsycINFO Search Process*



*Note.* ERIC and ASHA database searches did not yield additional sources for review.