

Sources of Support for the Phonological Awareness Screening Test (PAST)

David A. Kilpatrick, PhD

I have been asked to provide empirical support for the PAST and put it on www.thepasttest.com. I am aware that there have been recent criticisms of the PAST on social media. This document provides multiple lines of support for the test.

Below is a list of sources of empirical support for the PAST followed by an extended discussion of each.

- 1) The PAST is comprised of phoneme deletion and substitution test items. For decades, the research literature on phonemic awareness has shown that deletion and substitution tasks rank among the most valid and reliable ways to assess phonemic awareness.
- 2) The Phonemic Proficiency subtest from the Wechsler Individual Achievement Test–Fourth Edition (WIAT-4) was based on the PAST. The Phonemic Proficiency subtest has extensive, nationally stratified, and normed data supporting its reliability and validity.
- 3) The Alberta Department of Education (Alberta, Canada) has PAST data on 20,000 students demonstrating that it was their kindergarten screening battery’s best predictor of first-grade reading.
- 4) I have accumulated research data on the PAST across four school districts and over 1,000 students, showing the PAST has good to excellent reliability and validity.
- 5) The WIAT-4 Phonemic Proficiency subtest and PAST are the only current tests that assess *phonemic proficiency*, a skill central to the process of remembering written words.
- 6) Finally, critics of the PAST recommend using phoneme segmentation and blending. These two tasks have been shown for decades to have weaker correlations with reading than phoneme deletion and substitution tasks.

1) The phonemic awareness tasks used in the PAST have been shown to be reliable and valid in countless studies over the years.

The PAST uses phoneme deletion and substitution tasks. Over the last four decades, these phonemic tasks have been used in hundreds of studies and have been shown to display moderate to strong correlations with reading along with good reliability. In fact, manipulation tasks (deletion, substitution, reversal, and spoonerisms) have displayed the strongest correlations with reading compared to other phonemic tasks (e.g., segmentation, blending, isolation, rhyming; see Item 6 below). A variety of standard and experimenter-designed phoneme manipulation tasks have been used in the research literature. The results across these tasks have been fairly consistent. This shows that deletion and substitution tasks are very robust ways to estimate the underlying phonemic skills associated with reading. Below are some explicit quotes made by researchers about phoneme manipulation tasks over the last four decades:

In 1987: “The best predictor of word reading, in general, is deletion.” (p. 310) [compared to the phoneme blending and segmentation tasks they used]

Perfetti, C. A., Beck, I., Bell, L., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly*, 33, 283–319.

In 1990: “The correlations obtained between deletion and both the timed ($r = .74$) and the untimed measure of decoding ($r = .78$) as well as word attack on the CTBS ($r = .60$) exceeded those reported in the literature for any of the segmentation tasks.” (p. 243)

Lenchner, O., Gerber, M. M., & Routh, D. K. (1990). Phonological awareness tasks as predictors of decoding ability: Beyond segmentation. *Journal of Learning Disabilities*, 23(4), 240–247.

In 2000: “It was also not surprising that elision [i.e., deletion] and reversal tasks share more of the variance at this age level than do segmenting and blending. Both elision and reversal require more phonemic manipulation in that both tasks require both segmenting and blending of sounds as well as shifting the order of sounds. Therefore, for students older than K-2 whose phonological awareness has developed but still lags behind age-level peers, more difficult types of phonemic manipulation tasks are required to identify their phonological awareness problems.” (p. 121).

Kroese, J. M., Hynd, G. E., Knight, D. F., Hiemenz, J. R., & Hall, J. (2000). Clinical appraisal of spelling ability and its relationship to phonemic awareness (blending, segmenting, elision, and reversal), phonological memory, and reading in reading disabled, ADHD, and normal children. *Reading & Writing: An Interdisciplinary Journal*, 13, 105–131.

In 2001: “A deletion task was selected because research has shown that such a task ranks highly among phonological awareness tasks in predicting reading achievement.” (p. 40)

Catts, H. W. et al. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools*, 32, 38–50.

In 2005: “We selected phoneme deletion and spoonerisms to assess phoneme awareness because these two tasks have been found to be highly sensitive to individual differences in older children and adults.” (p. 110)

Caravolas, M., Volin, J., & Hulme, C. (2005). Phoneme awareness is a key component of alphabetic literacy skills in consistent and inconsistent orthographies: Evidence from Czech and English children. *Journal Experimental Child Psychology*, 92, 107–139.

In 2012: “The present article was designed to do two things. The first was to bring to the attention of researchers and practitioners what has already been reported but unheralded in existing research: From first grade and beyond, phonological segmentation tasks have weaker correlations with word-level reading skills than phonological manipulation tasks. Second, the present study provides explicit confirmation of what has already been previously reported.” (p. 162)

Kilpatrick, D. A. (2012). Phonological segmentation assessment is not enough: A comparison of three phonological awareness tests with first and second graders. *Canadian Journal of School Psychology*, 27(2), 150–165.

In 2020: “Our result is in line with [previous] findings . . . Indeed, the more complex phonological manipulation task predicts reading proficiency better than the less complex phonological segmentation task.” (p. 530)

Dorofeeva, S. V., et al. (2020). Complex phonological tasks predict reading in 7 to 11 years of age typically developing Russian children. *Journal of Research in Reading*, 43(4), 516–535.

The PAST is based on the Auditory Analysis Test (AAT) of Rosner and Simon (1971), the same test that formed the basis of the Elision subtest from the Comprehensive Test of Phonological Processing–Second Edition (CTOPP-2). The CTOPP-2 Elision subtest has become a standard measure of phonemic awareness in the reading research field and in school-based evaluations. A major difference is that the PAST supplements the original AAT’s phoneme deletion items with substitution items. This is significant given that recent

studies have indicated that phoneme substitution items correlate more strongly with reading than phoneme deletion (see Item 6).

Currently, there are no published studies on the PAST, though there are four either under review or in preparation (see Item 4 below). However, given that the PAST directly mimics the types of deletion and substitution tasks shown in the literature to be valid and reliable, it would seem reasonable to assume the PAST would yield similar results (and it does; see Item 4 below).

In sum, the PAST uses two of the most sensitive tasks designed to estimate the phonemic skills associated with reading. These phonemic tasks, in multiple forms and versions, have consistently been shown to correlate more strongly with reading than other phonemic tasks, such as segmentation, blending, and categorization. The extensive support for the tasks on which the PAST is based represents indirect empirical support for the PAST.

2) The nationally normed Phonemic Proficiency subtest from the Wechsler Individual Achievement Test–Fourth Edition (WIAT-4) provides indirect support for the PAST because it was based on the PAST.

The Phonemic Proficiency subtest from the WIAT-4 was based upon the PAST. I worked closely with the project director for that test battery, Kristina Breau, PhD, when designing that subtest (disclaimer: I elected to receive no financial compensation for my assistance with the WIAT-4 and have no financial interest in the WIAT-4 or any other test).

The Phonemic Proficiency subtest shares the key features of the PAST: (1) it uses phoneme deletion and substitution; (2) all of the correct responses are orthographically inconsistent with the prompts (e.g., “Say sweet without the /s/.” -> *wheat*); and (3) it uses the two-second timing procedure to determine if responses are instant or not instant (see Item 5 below).

The Phonemic Proficiency subtest was part of the WIAT-4 battery that was nationally normed and released in 2020. The Phonemic Proficiency subtest has nationally stratified data gathered on individuals ages 6 to 50. That subtest, modeled after the PAST but using different words and a different arrangement of items, consistently displays moderate to strong correlations with all of the word-reading subtests on the WIAT-4 across the entire age range. This is not surprising given that it uses deletion and substitution activities. As mentioned above, research has consistently shown that these tasks are among the most valid and reliable ways to assess the phonemic awareness skills associated with reading. According to Kristina Breau, PhD, the WIAT-4 project director:

The strong conceptual alignment between the PAST and the WIAT-4 Phonemic Proficiency (PP) subtest, combined with the observed correlations between the PP subtest and reading-related outcomes, provide meaningful preliminary evidence that interpretations based on PAST scores are likely to reflect important aspects of reading-related constructs. While formal validation studies are still needed, these findings offer a compelling foundation for considering the PAST as a theoretically and empirically supported measure within the broader context of reading assessment. (Kristina Breau, PhD, Pearson, Inc., personal communication, December 19, 2025)

3) The Alberta Department of Education found the PAST to be its best predictor of first-grade reading from their sample of about 20,000 students.

A few years ago, I gave the Alberta Department of Education permission to use the PAST. George Georgiou, PhD, a reading researcher from the University of Alberta, helped the Alberta Department of Education implement the PAST as part of a larger screening battery for kindergarteners. In addition to the PAST, that battery included phoneme segmentation and blending subtests developed by Dr. Georgiou and his colleague Dr. Rauno Parrila, along with tests of letter-sound knowledge and rapid automatized naming.

Dr. Georgiou informed me that the Alberta Department of Education's data on approximately 20,000 students indicated that the PAST was the best predictor of first-grade reading compared to any of their other screening tests. This is clearly impressive empirical support for the PAST. While it only applies to the kindergarten level, it provides strong support for the PAST's validity, reliability, and practical usefulness for schools using the PAST at that level.

4) I have accumulated data on the PAST across four school districts and nine data-gathering efforts showing the PAST has very good validity and reliability.

Since 2004, I have been collecting data on the PAST. These were not studies of the PAST per se, but studies of other questions about reading that used the PAST as a measure of phonemic skills. Most of these results have been presented at scientific conferences and four are currently either under review at scientific journals or in preparation for submission to scientific journals.

As of Spring 2025, a colleague and I have accumulated PAST data on 1,002 students, primarily from Grades K-5, but also high school and college students. Our findings are entirely unsurprising. They are consistent with the large body of existing studies using phoneme deletion and substitution tasks (see Item 1 above). The PAST displays moderate to strong correlations with various word reading and nonsense word reading subtests, both timed and untimed, as well as oral reading fluency. In addition, the PAST strongly correlates with the CTOPP-2 Elision subtest (a phoneme deletion task), which has become a standard phonemic task in the research literature. Finally, the PAST showed impressive predictive validity. In four studies with different student samples, it predicted later reading ability after 2 ½ months, 4 months, 4 years, and 8-10 years.

In two of our data sets we measured reliability. It must be noted that you cannot have good validity estimates without reliability. This means that the moderate-to-strong validity estimates for the PAST provide indirect evidence of its reliability.

Our first test of reliability was conducted with 56 kindergarteners. The PAST displayed a test-retest reliability estimate of .82. That is considered acceptable, but two considerations must be noted. First, the tests were given 2 ½ months apart. Most test-retest reliability measurements occur a few days or perhaps a week apart. Second, an alternative form of the PAST was given at the second administration, making this an estimate of combined test-reliability and alternate-form reliability. Of note is the fact that four Aimsweb subtests were given at both times as well. The PAST's .82 reliability after 2 ½ months, using an

alternate form, exceeded the reliability estimates of two of the Aimsweb subtests, it was identical to a third, and was lower than a fourth. Thus, the PAST's reliability over that time frame was comparable to Aimsweb.

To expand on this, we administered the PAST one week apart to 57 students from kindergarten to Grade 4, this time using the same form. The reliability score was .92, which is very good. We measured interrater agreement/reliability on 25 K-1 students. There was a 99.7% agreement regarding whether items were correct or incorrect. Ratings of item responses as either instant or non-instant (see more below) displayed 90.0% agreement.

5) Currently, the WIAT-4 Phonemic Proficiency subtest and the PAST are the only tests available that assess phonemic proficiency, a critical skill behind the memory mechanism that allows students to remember written words in long-term memory for later, effortless retrieval.

Since 1989, studies have shown that the speed of responding to individual phonemic awareness items provides a more valid assessment of the phonemic skills needed for reading than simply whether a student gets the item correct or incorrect. This finding has been demonstrated across multiple alphabetic languages, including English, Dutch, French, and Spanish, and across the K-12 range (Ashby et al., 2013; Breaux et al., 2023; Gallagher et al., 1996; Kitz & Tarver, 1989; Paige & Rupley, 2026; Pourcin et al., 2016; Serrano & Defior, 2008; Snowling et al., 1997; Vaessen et al., 2009, 2010; Vaessen & Blomert, 2010). None of these researchers ever provided a name for the concept of rapid responses to phonemic awareness items. So, in a 2015 book I wrote, I coined the term *phonemic proficiency*.

These studies all used phoneme deletion, substitution, or spoonerisms and timed item-level responses. Unfortunately, each of these studies used their own experimenter-designed assessments, none of which are available to educators. The exception was Paige & Rupley (2026), who used the WIAT-4.

The findings from these studies are consistent with Dr. Linnea Ehri's theory of orthographic mapping. Orthographic mapping is the memory process that begins after a word has been identified.¹ It explains how students store words in long-term memory for later, instant and effortless retrieval. Ehri's theory of orthographic mapping has undergone decades of rigorous scientific testing and has extensive research support (Ehri, 2005).

Orthographic mapping proposes that after a word has been identified, students use the sequence of phonemes in the word's pronunciation—a sequence already stored in memory—as the anchoring points in memory for the series of letters used to spell the word. After a few exposures, the written word becomes familiar and no longer needs to be sounded out. Thus, orthographic mapping makes it clear why phonemic skills are

¹ I have been hearing that the term "orthographic mapping" is being misapplied on social media to refer to sounding out words using phonetic decoding. This is not correct. Phonetic decoding is applied to unfamiliar words while orthographic mapping refers to a memory process that occurs *after* a word has been identified (Ehri, 2014).

important for reading and why poor access to phonemes interferes with reading skills. Access to phonemes is central to the written word-memory process.

There is another layer to this. If you are a skilled reader who did not struggle to learn to read, you have tens of thousands of familiar letter sequences (i.e., written words and word parts) stored in long-term memory ready for instant retrieval. The vast majority of those words—perhaps 99.9%—entered your long-term memory without any conscious effort. This is different from almost everything else we learned in school, which required conscious memorization. This indicates that the process of orthographic mapping occurs automatically and unconsciously. Once a new word is identified, the reader continues reading and does not stop to consciously memorize the newly encountered word. Given that this orthographic mapping/memory process is automatic, and access to phonemes is a central part of that process, it follows that access to phonemes must also be automatic.

This need for fast access to phonemes is consistent with the research mentioned above, which shows that the speed of response to individual phonemic test items is more strongly associated with reading than simply getting test items correct. In fact, based upon the WIAT-4 normative sample from ages 6 to 50, the correlation between instant responses on the Phonemic Proficiency subtest and the WIAT-4 Word Reading subtest is a little over .60 across the lifespan. By contrast, the correlation between word reading and correct responses that were not instant hovers around 0 (Breaux et al., 2023). This is consistent with orthographic mapping theory. The WIAT-4 data validates the practice of distinguishing between instant and non-instant responses, based upon a nationally stratified sample. In our studies (Item 4 above), the PAST displayed very similar results with the timing element as was found on the WIAT-4. This is not surprising because, as mentioned, the WIAT-4 Phonemic Proficiency subtest was based on the PAST.

As of 2026, apart from the PAST and the WIAT-4 Phonemic Proficiency subtest, phonemic awareness tests do not distinguish between fast and slow item-level responses. The timed phonemic segmentation tasks on the universal screeners, which provide an overall timing of all items, apparently do not provide the same degree of estimation of the phonemic skills needed for reading as item-level timing using manipulation tasks. The correlations between reading and those segmentation tasks are weaker and cease to be useful after first grade, which is presumably why they are discontinued from those batteries after first grade. By contrast, timed manipulation tasks maintain their moderate-to-strong correlations with reading across the entire K-12 level and into adulthood, as the normative data from the WIAT-4 show.

It is important to recognize that students can provide correct responses to phonemic awareness test items without having the phoneme skills needed for reading. They can get items correct, but slowly, using either mental spelling or slowly tracking through the words. Those makeshift task-related strategies do not appear to correlate with reading, while the quick responses—that is, responses suggesting instant access to phonemes—do. Distinguishing between fast and slow responses to individual items represents an important value of the WIAT-4 Phonemic Proficiency subtest and the PAST.

In sum, the PAST is one of only two current tests that assess a key phonemic skill that is central to the process of remembering words.

6) Finally, critics of the PAST recommend using phoneme segmentation and blending. It has been shown for decades that these two tasks tend to have weaker correlations with reading than phoneme deletion and substitution tasks.

I have read several hundred research reports on phonemic awareness and reading over the last 28 years. I am unaware of the basis for the claim that phoneme segmentation assessments should be preferred over phonemic manipulation tasks. My best attempt at explaining the origin of this notion is that some educators and some in the school psychology field mistakenly assume segmentation must be the best measure of phonemic awareness because that is the type of task found on the universal screeners popular in schools (i.e., DIBELS, Acadience, Aimsweb, EasyCBM). However, that would be a faulty inference. Such tasks conveniently fit the one-minute format used in universal screeners. But they have weaker correlations with reading than other more complex phonemic awareness tasks.

The quotes in Item 1 above show that researchers have explicitly acknowledged that segmentation tasks have weaker correlations with reading than more complex phonemic awareness tasks (e.g., deletion, substitution, spoonerisms).

An interesting example of the disparity between segmentation and manipulation is the original CTOPP, which was published in 1999. The CTOPP's phoneme segmentation task, called the Segmenting Words subtest, was not part of the regular battery but was placed in a supplemental battery. This is presumably because it had such weak correlations with word reading (.35) and nonsense-word reading (.25) while their phoneme manipulation task, the Elision subtest, had much stronger correlations with word reading (.53 and .47, respectively). This is very noteworthy and more powerful than a single study because these results represent nationally stratified data from ages 7 to 24. The CTOPP normative sample was consistent with many previous studies indicating that phoneme segmentation tasks have weaker correlations with reading than phoneme manipulation tasks. The weak results on the Segmenting Words subtest may be why it was left off the updated version (the CTOPP-2). It was replaced in the supplementary battery with the Segmenting Nonwords subtest. However, nowhere in the CTOPP-2 manual is any data provided on the correlation between Segmenting Nonwords and reading.

Other nationally stratified datasets have been consistent with the findings from decades of studies showing that phoneme segmentation displays weaker correlations with reading than phoneme manipulation. Kristina Breaux, PhD, the project director for the Kaufman Test of Educational Achievement–Third Edition (KTEA-3), recently analyzed the phonemic tasks that make up the KTEA-3's Phonological Processing subtest. That subtest is comprised of separate tasks including deleting, segmenting, blending, sound matching, and rhyming. In Grades K to 5, deleting sounds had the strongest correlations with reading compared to the other phonemic tasks while blending had the weakest (Breaux & Kilpatrick, 2026).

The Tests of Dyslexia (TOD) is a normed assessment battery of reading and reading-related language skills. After the norming data were gathered and analyzed, the test's author sent me an email saying, "The Phonemic Manipulation test was much stronger than

blending and segmenting at predicting reading” (Nancy Mather, PhD, personal communication, June 9, 2023).

There is thus data to address this issue from three nationally stratified normative samples: the CTOPP, the KTEA-3, and the TOD. The data from each of these normative samples are consistent with decades of research showing that phoneme segmentation and blending display weaker correlations with reading than phoneme manipulation tasks.

Deletion vs. Substitution. There is one further bit of indirect support for the PAST. As mentioned above, the PAST and Elision subtests were both derived from the Rosner and Simon (1971) AAT, which is made up of deletion items only. Back in the 1980s, my mentor Dr. Philip J. McInnis made his modification of the AAT and included substitution items. The PAST followed McInnis’ lead on this. Two recent studies have shown that McInnis was prescient when it was discovered that substitution items tend to correlate more strongly with reading than deletion items. The first was Paige and Rupley (2026) who used the WIAT-4 with 161 first and second graders. Second, Kristina Breaux examined the WIAT-4 normative sample and found that at most grade levels, substitution items correlated more strongly with reading than deletion items (Breaux & Kilpatrick, 2026). Thus, while phoneme deletion displays stronger correlations with reading than segmentation and blending, phoneme substitution tends to correlate more strongly with reading than deletion. The PAST is made up of both deletion and substitution items. This further supports the comment made above that the PAST makes use of the most sensitive tasks designed to estimate the phonemic skills associated with reading.

Response to Critics

I can only assume that those who criticize the PAST and recommend phoneme segmentation are unfamiliar with key details from the decades of research on phonemic awareness described above. As mentioned, my best guess is that they have concluded from the widespread use of phoneme segmentation tests in schools that segmentation must be the most valid way to assess phoneme awareness.

If the PAST’s critics were familiar with the phonemic awareness research literature and yet remained critical of the PAST, they would have pointed out specifically how the PAST falls short of similar phoneme manipulation tasks that consistently demonstrate reliability and validity. There are only three differences between the PAST and similar tests that have been shown for decades to be valid: 1) the actual prompt words used, 2) the order in which items are presented in terms of linguistic complexity, and 3) the number of total items on the test. Nearly all deletion and substitution tasks vary on these three differences yet yield similar results in research studies. So, unless it can be shown that the PAST represents a very poor implementation of those highly sensitive phonemic awareness tasks, we have no reason to believe that the PAST “is not a good test,” as one critic claimed online. On the contrary, it can be seen from the above that there are several reasons to believe that the PAST is a good test of phonemic awareness.

Compiled May 2026

References

- Ashby, J., Dix, H., Bontrager, M., Dey, R., & Archer, A. (2013). Phonemic awareness contributes to text reading fluency: Evidence from eye movements. *School Psychology Review*, 42(2), 157–170.
- Breaux, K. C., & Kilpatrick D. A. (2026, July 21). *Phoneme deletion, segmentation, and blending: Which best estimates the latent phonemic skills needed for word-level reading?* [Paper presentation] Society for the Scientific Study of Reading 33rd Annual Conference, Rotterdam, The Netherlands.
- Breaux, K. C., Kilpatrick D. A., & Kuang, H. (2023, July 21). *Beyond phonemic awareness: Phonemic proficiency and word-reading skills*. [Paper presentation] Society for the Scientific Study of Reading 30th Annual Conference, Port Douglas, Queensland, Australia.
- Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9(2), 167-188.
- Ehri, L. C. (2014). Orthographic Mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning, *Scientific Studies of Reading*, 18(1), 5-21.
- Gallagher, A. M., Laxon, V., Armstrong, E., & U. Frith (1996). Phonological difficulties in high-functioning dyslexics. *Reading and Writing: An Interdisciplinary Journal*, 8, 499-509.
- Kitz, W. R., & Tarver, S. G. (1989). Comparison of dyslexic and nondyslexic adults on decoding and phonemic awareness tasks. *Annals of Dyslexia*, 39, 196-205.
- Paige, D. D., & Rupley, W. H. (2026). Phoneme automaticity: A test of the phonemic proficiency hypothesis. *Education Sciences*, 16(2), 286.
- Pourcin, L., Pourcin, L., Sprenger-Charolles, L., El Ahmadi, A., & Colé, P. (2016). Reading and related skills in grades 6, 7, 8 and 9: French normative data from EVALEC. *European Review of Applied Psychology*, 66(1), 23-37.
- Rosner, J. & Simon, D. P. (1971). The Auditory Analysis Test: An initial report. *Journal of Learning Disabilities*, 4(7), 384-392.
- Serrano, F., & Defior, S. (2008). Dyslexia speed problems in a transparent orthography. *Annals of Dyslexia*, 58, 81–95.
- Snowling, M., Nation, K., Moxham, P., Gallagher, A., & Frith, U. (1997). Phonological processing skills of dyslexic students in higher education: A preliminary report. *Journal of Research in Reading*, 20(1), 31-41.
- Vaessen, A., & Blomert, L. (2010). Long-term cognitive dynamics of fluent reading development. *Journal of Experimental Child Psychology*, 105, 213–231.
- Vaessen, A., Bertrand. D., Tóth, D., Csépe, V., Fásca, L., Reis, A., & Blomert, L. (2010). Cognitive development of fluent word reading does not qualitatively differ between transparent and opaque orthographies. *Journal of Educational Psychology*, 102(4), 827–842.
- Vaessen, A., Gerretsen, P., & Blomert, L. (2009). Naming problems do not reflect a second independent core deficit in dyslexia: Double deficits explored. *Journal of Experimental Child Psychology* 103, 202–221.