

Identifying North Carolina Students at Risk of Scoring below Proficient in Reading at the End of Grade 3

A Publication of the National Center for Education Evaluation and Regional Assistance at IES



Identifying North Carolina Students at Risk of Scoring below Proficient in Reading at the End of Grade 3

Sharon Koon, Barbara Foorman, and Tara Galloway

August 2020

North Carolina policymakers are interested in which screening and monitoring assessment scores in grades K–3 best predict reading proficiency at the end of grade 3 in order to help educators identify at-risk students, place them in interventions, and monitor their progress. Since 2013, the state has routinely administered screening and monitoring assessments, but the percentage of students scoring proficient in reading at the end of grade 3 has remained largely unchanged.

This study used longitudinal student achievement data, provided by the North Carolina Department of Public Instruction, and classification and regression tree analysis to examine how well assessment scores from five points in time in grades K–3 predicted proficiency status on the state reading assessment at the end of grade 3. The score from only one of the assessments identified more than 80 percent of students who scored below proficient at the end of grade 3 as at risk and thus adequately identified at-risk students: the North Carolina Beginning-of-Grade 3 English Language Arts/Reading Test. That state-developed assessment measures the same content standards at the same difficulty level as the state reading assessment at the end of grade 3. The results suggest that more information is needed in order to use assessment data from grades K–2 to reliably identify students at risk of not achieving reading proficiency at the end of grade 3. Educators might want to consider supplementing screening and progress monitoring assessments with informal, curriculum-based assessments that measure student vocabulary, syntax, and listening comprehension skills because research has identified those skills as important predictors of reading comprehension.

Why this study?

More than half of North Carolina’s grade 3 students struggle with reading. The percentage of students scoring below proficient in reading at the end of grade 3 was 55 percent in 2017/18 (North Carolina Department of Public Instruction, 2018). That percentage has remained largely unchanged, at above 50 percent, since 2013/14, when North Carolina passed its K–3 Read to Achieve literacy act.

Read to Achieve required North Carolina schools to use interim assessments of reading proficiency in kindergarten–grade 3 to identify students at risk of not achieving reading proficiency at the end of grade 3. Districts administered screening and progress monitoring assessments at the beginning, in the middle, and at the end of each school year. This is meant to allow educators the opportunity to quickly identify and place students who may be at risk into interventions designed for improving reading proficiency and to track their improvement.

Hoping to accelerate their improvement efforts, the North Carolina Department of Public Instruction partnered with the Regional Educational Laboratory Southeast to strengthen the use of interim assessments in their state. Interim assessments can yield multiple scores, and teachers may be uncertain about which scores are most relevant to identifying students in need of intervention and which skill areas they are having difficulties in. This study identifies which scores provide teachers the most information about at-risk students. The department hopes to use that information to build decision trees in each grade for teachers to identify students in need of intervention.

For details on the study’s technical methods, access the report’s appendix at <https://go.usa.gov/xfAZa>.

Box 1. Key terms

Daze. A measure of a student’s reading fluency and reading comprehension.

First Sound Fluency. A measure of a student’s ability to segment the initial sound (phoneme) in a word. It is a test of early phonemic awareness.

Interim assessments. Screening and progress monitoring assessments administered at the beginning, in the middle, and at the end of the school year.

Letter Name Fluency. A measure of a student’s ability to name upper and lower case letters.

mCLASS 3D Reading assessments. The mCLASS 3D Reading assessments consist of Acadience™ Reading (Dynamic Measurement Group, 2018; formerly DIBELS Next®) measures and a set of reading comprehension passages called Text Reading and Comprehension (Amplify, 2015).

North Carolina Beginning-of-Grade 3 English Language Arts/Reading Test (BOG3 assessment). A state-developed assessment that establishes a baseline measure of students’ reading skills at the beginning of grade 3. It is an equivalent form of the state reading assessment at the end of grade 3. All students enrolled in grade 3 are expected to participate with or without accommodations in the administration of the BOG3 assessment.

Overall percentage of students correctly identified. The number of below-proficient students correctly identified as at risk plus the number of proficient students correctly identified as not at risk, divided by the total number of students.

Phoneme Segmentation Fluency. A measure of a student’s ability to segment a spoken word into its component parts or sound segments. It is a test of phonemic awareness.

Predictive ability. The percentage of below-proficient students correctly identified as at risk. This percentage is commonly referred to as sensitivity.

Research questions

The study addresses two research questions:

1. How well do scores from assessments administered from the middle of kindergarten to the beginning of grade 3 in North Carolina predict reading proficiency at the end of grade 3?
2. For assessments that correctly predict reading proficiency at the end of grade 3, what decision rules can educators use to identify students in need of intervention?

The study examined the predictive ability of scores from seven mCLASS 3D Reading interim assessments and one state-developed assessment, the North Carolina Beginning-of-Grade 3 English Language Arts/Reading Test (BOG3 assessment). It relied on data on students with assessment data from the middle of kindergarten in 2014/15 to the end of grade 3 in 2017/18. All the assessments are reliable measures of the curriculum taught in North Carolina schools (see table A2 in appendix A; North Carolina Department of Public Instruction, 2014). Administrations of the interim assessments at five points in time were examined:

- The middle of the year in kindergarten, selected because the timing allows students to settle into formal schooling and because it measures phonemic segmentation fluency, a major predictor of reading proficiency (Kaminski & Good, 1996).
- The beginning of the year in grades 1, 2, and 3, selected because the beginning of the year is when instructional decisions are made within a grade and allows the most instructional time for at-risk students.
- The middle of the year in grade 1, selected because it is the first time that oral reading fluency, a predictor of reading comprehension (Kim, Petscher, Schatschneider, & Foorman, 2010), is measured and because the

North Carolina Department of Public Instruction emphasizes oral reading fluency in statewide professional development.

The study developed a series of statistical models to determine the relationship between assessment scores and scores on the state’s grade 3 reading assessment. The models used in the study included 10 scores from the interim assessments as predictors: one each from the First Sound Fluency, Letter Name Fluency, Phoneme Segmentation Fluency, Daze (a timed maze box completion), and Text Reading and Comprehension assessments, two from the Nonsense Word Fluency assessment (Correct Letter Sounds and Whole Words Read), and three from the Oral Reading Fluency assessment (Words Correct Per Minute, Accuracy, and Retell). Table 1 indicates which interim assessment scores from which administrations of the assessments were included in the models.

The BOG3 assessment measures the same content standards at the same difficulty level as the state reading assessment at the end of grade 3. It was examined in combination with the beginning-of-the-year interim assessments in grade 3, which it is administered separately from. The BOG3 assessment yields a single composite score, which was the only predictor in the model that included that assessment.

The data sources, sample, and methods used in the study are summarized in box 2 and discussed in detail in appendix A.

Table 1. Interim assessment scores included in the analysis, by time of administration

Assessment score	Kindergarten middle of the year	Grade 1 beginning of the year	Grade 1 middle of the year	Grade 2 beginning of the year	Grade 3 beginning of the year
First Sound Fluency	●				
Letter Name Fluency	●	●			
Phonemic Segmentation Fluency	●	●			
<i>Nonsense Word Fluency</i>					
Correct Letter Sounds	●	●	●	●	
Whole Words Read		●	●	●	
<i>Oral Reading Fluency</i>					
Words Correct Per Minute			●	●	●
Accuracy			●	●	●
Retell					●
Daze					●
Text Reading and Comprehension	●	●	●	●	●

Source: Authors’ illustration.

Box 2. Data source, sample, and methods

Data source. The study used longitudinal student achievement data for 2014/15–2017/18 obtained through a memorandum of agreement between the Regional Educational Laboratory Southeast and the North Carolina Department of Public Instruction.

Sample. The dataset consisted of students who took the state reading assessment at the end of grade 3 in 2017/18 for the first time and had data from the middle-of-the-year interim assessments in kindergarten in 2014/15 as well as data from the beginning-of-the-year and middle-of-the-year interim assessments in grade 1 in 2015/16, the beginning-of-the-year interim assessments in grade 2 in 2016/17, and the beginning-of-the-year interim assessments in grade 3 in 2017/18. These eligibility requirements resulted in a study sample of 91,855 students, or about 77 percent of the 120,029 grade 3 students statewide in 2017/18. The demographic characteristics of the statewide population and the study sample were very similar (see table A1 in appendix A). However, because the study sample excludes 23 percent of grade 3 students statewide, the study findings may not generalize to the statewide population of grade 3 students.

Methods. Associations between student scores on each assessment and proficiency status on the state reading assessment at the end of grade 3 were examined using classification and regression tree models. The dataset was split into a calibration dataset used to build the models, consisting of a random sample of 73,607 students (approximately 80 percent of the study sample), and a validation dataset used to test the models, consisting of the remaining 18,248 students (approximately 20 percent of the study sample). The statistical package *rpart* (Therneau, & Atkinson, & Ripley, 2019) for use in the R software environment was used for the analysis.

The study team tested six models. Five models included scores from a single administration of the interim assessments in grades K–3, and one model included scores from beginning-of-the-year interim assessments in grade 3 as well as a composite score on the North Carolina Beginning-of-Grade 3 English Language Arts/Reading Test. The study team did not test the BOG3 assessment individually because it is an equivalent form of the state reading assessment at the end of grade 3 and the predictive ability of this assessment alone is already demonstrated in technical studies. However, the predictive ability of the combination of scores from the interim assessments and the BOG3 assessment was unknown.

The primary criterion for evaluating each model was the percentage of below-proficient students correctly identified as at risk (referred to in this study as predictive ability). The percentage of proficient students correctly identified as not at risk and the overall percentage of students correctly identified (the number of below-proficient students correctly identified as at risk plus the number of proficient students correctly identified as not at risk, divided by the total number of students) were also considered. The study team used a threshold of 80 percent correct for all three criteria. That threshold is in line with previous research that has required values of between 80 percent (Piastra, Petscher, & Justice, 2012) and 90 percent (Compton, Fuchs, Fuchs, & Bryant, 2006; Jenkins, Hudson, & Johnson, 2007).

A more detailed description of the methodology is in appendix A.

Findings

This section presents key findings of the study. The assessment scores that were examined are specific to North Carolina standards and to a single cohort of students and thus may not generalize to nationally normed reading assessments or to other student cohorts. In addition, students may have been excluded from the study sample because they transferred in or out of the public school system during the study period, because they were absent during the testing sessions, or because they were exempt from testing due to learning disabilities or limited English proficiency. The demographic characteristics of the statewide population and the study sample were very similar (see table A1 in appendix A). However, because the study sample excluded 23 percent of grade 3 students statewide, including students with disabilities and English learner students not tested on the state reading assessment at the end of grade 3, the study findings may not generalize to the statewide population of grade 3 students.

The interim assessments did not adequately identify students at risk of scoring below proficient on the state reading assessment at the end of grade 3

The interim assessments, including each of the available scores at the five points in time, correctly identified 61–69 percent of students who scored below proficient on the state reading assessment at the end of grade 3 as at risk (table 2). Those percentages are below the study threshold for adequate predictive ability. The assessments performed better in identifying students who scored proficient as not at risk, correctly identifying 73–86 percent. The overall percentage of students correctly identified was below 80 percent for all seven assessments. The corresponding decision trees for the models that include these assessments are provided in figures A1–A5 in appendix A.

Only the model that combined the beginning-of-the-year interim assessments in grade 3 with the BOG3 assessment met the study threshold for adequate predictive ability for percentage of below-proficient students correctly identified as at risk and for overall percentage of students correctly identified. The next section discusses those findings in detail and provides the corresponding decision tree for that model. The rest of this section discusses detailed findings for the models that included a single administration of the interim assessments.

Kindergarten middle-of-the-year interim assessments. The middle-of-the-year interim assessments in kindergarten correctly identified 63 percent of students who scored below proficient on the state reading assessment at the end of grade 3 as at risk. They performed slightly better in identifying students who scored proficient as not at risk, correctly identifying 73 percent. And they correctly identified 69 percent of students overall.

Grade 1 beginning-of-the-year and middle-of-the-year interim assessments. The beginning-of-the-year interim assessments in grade 1 correctly identified a higher percentage of students who scored below proficient on the state reading assessment at the end of grade 3 as at risk (66 percent) than the kindergarten assessment did, but the middle-of-the-year interim assessments in grade 1 correctly identified a lower percentage (61 percent). Both grade 1 assessments correctly identified a higher percentage of students who scored proficient (79 percent and 86 percent) as not as risk and a higher percentage of students overall (73 percent and 75 percent) than the kindergarten assessments did.

Grade 2 beginning-of-the-year interim assessments. The beginning-of-the-year interim assessments in grade 2 were better than the middle of grade 1 assessments at correctly identifying below-proficient students but (slightly) worse than the beginning of grade 1 assessments. The grade 2 assessments were slightly better than both the grade 1 assessments at identifying students overall. The grade 2 assessments correctly identified 65 percent of

Table 2. Percentage of students correctly identified, by proficiency status and overall and by model, 2014/15–2017/18

Model	Percentage of below-proficient students correctly identified as at risk	Percentage of proficient students correctly identified as not at risk	Overall percentage of students correctly identified
Kindergarten middle-of-the-year interim assessments	63	73	69
Grade 1 beginning-of-the-year interim assessments	66	79	73
Grade 1 middle-of-the-year interim assessments	61	86	75
Grade 2 beginning-of-the-year interim assessments	65	85	76
Grade 3 beginning-of-the-year interim assessments	69	85	78
Grade 3 beginning-of-the-year interim assessments and BOG3 assessment	83	84	83

BOG3 assessment is the North Carolina Beginning-of-Grade 3 English Language Arts/Reading Test.

Note: Proficiency status refers to score on the state reading assessment at the end of grade 3. Results are based on the validation dataset ($n = 18,248$).

Source: Authors' analysis of data from the North Carolina Department of Public Instruction.

students who scored below proficient as at risk, 85 percent of students who scored proficient as not at risk, and 76 percent of students overall.

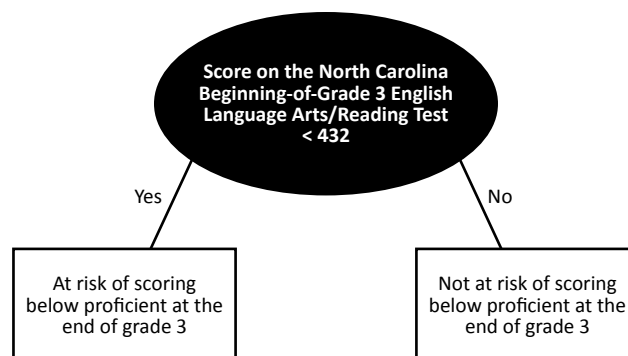
Grade 3 beginning-of-the-year interim assessments. The beginning-of-the-year interim assessments in grade 3 correctly identified 69 percent of students who scored below proficient on the state reading assessment at the end of grade 3 as at risk. This is the highest percentage of all the interim assessments, but it does not meet the study threshold for adequate predictive ability. As with the grade 2 assessments, the grade 3 assessments performed better in correctly identifying students who scored proficient as not at risk (85 percent) and correctly identifying students overall (78 percent).

The North Carolina Beginning-of-Grade 3 English Language Arts/Reading Test adequately identified students at risk of scoring below proficient on the state reading assessment at the end of grade 3

Although the model used in the analysis considered both the beginning-of-the-year interim assessments in grade 3 and the BOG3 assessment as potential predictors, only the BOG3 assessment was selected by the model to classify a student’s proficiency status on the state reading assessment at the end of grade 3. This could be because the BOG3 assessment is designed to serve as a pretest for the state reading assessment at the end of grade 3 and the two are thus better aligned with each other than with the interim assessments. The BOG3 assessment correctly identified 83 percent of students who scored below proficient as at risk and 83 percent of students overall.

The BOG3 model had the highest predictive ability and the highest overall percentage of students correctly identified. Teachers can use these scores, collected at the start of grade 3, to accurately predict 83 out of 100 students’ proficiency in reading at the end of grade 3. Based on this analysis, students who score below 432 are at risk of not achieving proficiency at the end of grade 3 (figure 1).

Figure 1. North Carolina students who score below 432 on the North Carolina Beginning-of-Grade 3 English Language Arts/Reading Test are at risk of scoring below proficient on the state reading assessment at the end of grade 3, 2017/18



Note: Results are based on the calibration dataset (n = 73,607).

Source: Authors’ analysis based on data from the North Carolina Department of Public Instruction.

Implications

The study findings suggest that an interim assessment that is well aligned with the state reading assessment at the end of grade 3, as the BOG3 assessment is, can better identify North Carolina students who are at risk of scoring below proficient in reading at the end of grade 3 than the mCLASS 3D Reading assessments can. It is possible that the mCLASS 3D Reading assessments are less aligned with the state assessments and this is why they performed worse in identifying at-risk students. Formal alignment studies may provide insight into the degree to which the interim assessments and the state reading assessment at the end of grade 3 have the same content focus or other alignment characteristics.

The inadequate ability of the interim assessments to identify at-risk and not-at-risk students means that a substantial proportion of students who need interventions may not receive them (Petscher et al., 2019) while the limited intervention resources available may be targeted to some students who do not actually need them. More information, such as student oral language skills, is needed to reliably identify students in grades K–2 as at risk of scoring below proficient on the state reading assessment at the end of grade 3. Several longitudinal studies starting as early as preschool (Storch & Whitehurst, 2002) and grade 1 (Kim et al., 2010; Quinn, Wagner, Petscher, & Lopez, 2015) have found vocabulary to be a significant predictor of reading comprehension in grade 3 or 4. One cross-sectional study from preschool to grade 3 found that vocabulary had an indirect effect on grade 3 reading comprehension through word recognition and listening comprehension (Language and Reading Research Consortium, 2015). A study of how well the Florida Assessments for Instruction in Reading predicted reading achievement in grades 1 and 2 using the same classification and regression tree methodology that the current study uses found that two oral language measures—vocabulary and listening comprehension—predicted reading achievement better than word identification did (Koon, Petscher, & Foorman 2014). In addition to average oral language skills predicting average reading comprehension, use of profile analysis has identified clusters of students in elementary school grades with particularly low oral language skills (Foorman, Petscher, Stanley, & Truckenmiller, 2017). Finally, there is a large degree of overlap in how oral language and word reading fluency predict reading comprehension in the elementary school grades (Foorman, Petscher, & Herrera, 2018; Lonigan, Burgess, & Schatschneider, 2018). Thus, it is important to measure oral language in the elementary school grades to improve the predictive validity of reading comprehension and to provide instructionally relevant information. In addition, strong oral language skills predict comprehension of complex text better than word reading fluency.

None of the predictors examined in this study include measures of oral language. Although the state reading assessment at the end of grade 3 does not assess oral language, it does assess reading comprehension, which oral language measures help predict, as described above. Thus, the North Carolina Department of Public Instruction might want to consider supplementing the interim assessment system with oral language measures to improve its ability to correctly identify at-risk students. In addition, educators can supplement progress monitoring assessments with informal, curriculum-based assessments to determine instructional objectives and differentiate instruction, which some teachers may already be doing.

References

- Amplify. (2015). *mCLASS Reading 3D: Amplify Atlas book set technical manual, 2nd Edition*. New York, NY: Author.
- Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98*(2), 394–409. <http://eric.ed.gov/?id=EJ742190>.
- Dynamic Measurement Group. (2018). *Acadience™ Reading: Benchmark goals and composite score*. Eugene, OR: Author. Retrieved September 5, 2019, from <https://acadiencelarning.org/papers/AcadienceReadingBenchmarkGoals.pdf>.

- Foorman, B., Petscher, Y., & Herrera, S. (2018). Unique and common effects of oral language in predicting reading comprehension in grades 1–10. *Learning and Individual Differences, 63*, 12–23. <https://doi.org/10.1016/j.lindif.2018.02.011>.
- Foorman, B., Petscher, Y., Stanley, C., & Truckenmiller, A. (2017). Latent profiles of reading and language and their association with standardized reading outcomes in kindergarten through tenth grade. *Journal of Research on Educational Effectiveness, 10*(3), 619–645. <https://doi.org/10.1080/19345747.2016.1237597>.
- Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for service delivery in an RTI framework: Candidate measures. *School Psychology Review, 36*(4), 582–601.
- Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25*, 215–227.
- Kim, Y-S., Petscher, Y., Schatschneider, C., & Foorman, B. (2010). Does growth rate in oral reading fluency matter in predicting reading comprehension achievement? *Journal of Educational Psychology, 102*(3), 652–667.
- Koon, S., Petscher, Y., & Foorman, B. (2014). *Using evidence-based decision trees instead of formulas to identify at-risk readers* (REL 2014–036). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Educational Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from <http://ies.ed.gov/ncee/edlabs>.
- Language and Reading Research Consortium. (2015). Learning to read: Should we keep things simple? *Reading Research Quarterly, 50*(2), 151–169. <http://dx.doi.org/10.1002/rrq.99>.
- Lonigan, C., Burgess, S., & Schatschneider, C. (2018). Examining the Simple View of Reading with elementary school children: Still simple after all these years. *Remedial and Special Education, 39*(5), 260–273. <http://dx.doi.org/10.1177/0741932518764833>.
- North Carolina Department of Public Instruction. (2014). *Reliability of the North Carolina end-of-grade and end-of-course assessments*. Raleigh, NC: Author. Retrieved September 5, 2019, from <http://www.ncpublicschools.org/docs/accountability/testing/eogeo reliabilities14.pdf>.
- North Carolina Department of Public Instruction. (2018). *2017–18 North Carolina state testing results*. Raleigh, NC: Author. Retrieved August 7, 2019, from <http://www.ncpublicschools.org/accountability/testing/reports/archive>.
- Petscher, Y., Fien, H., Stanley, C., Gearin, B., Gaab, N., Fletcher, J., et al. (2019). *Screening for dyslexia*. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education, Office of Special Education Programs, National Center on Improving Literacy. Retrieved March 18, 2020, from <https://improvingliteracy.org>.
- Piasta, S. B., Petscher, Y., & Justice, L. M. (2012). How many letters should preschoolers in public programs know? The diagnostic efficiency of various preschool letter-naming benchmarks for predicting first-grade literacy achievement. *Journal of Educational Psychology, 104*(4), 945–958.
- Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development, 86*(1), 159–175.
- Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934–947. <https://doi.org/10.1037//0012-1649.38.6.934>.

Therneau, T. M., Atkinson, B., & Ripley, B. (2019). *rpart: Recursive Partitioning. R package version 4.1–15*. The Comprehensive R Archive Network. Retrieved March 18, 2020, from <http://cran.r-project.org/web/packages/rpart/rpart.pdf>.

REL 2020–030

August 2020

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-17-C-0011 by the Regional Educational Laboratory Southeast administered by Florida State University. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This REL report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Koon, S., Foorman, B., & Galloway, T. (2020). *Identifying North Carolina students at risk of scoring below proficient in reading at the end of grade 3* (REL 2020–030). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from <http://ies.ed.gov/ncee/edlabs>.

This report is available on the Regional Educational Laboratory website at <http://ies.ed.gov/ncee/edlabs>.