Methodological Brief: The incomparability of campus-specific Benchmark Running Records

Courtney Thrash, M.A. and Erin Baumgartner, Ph.D.   June 2020

The comparability of various assessments of literacy achievement and growth adopted for use by the Houston Independent School District (HISD) was the focus of this study. Renaissance 360 is the universal screener used by campuses throughout the district. In contrast, Benchmark Running Records (BRR) are campus-specific, and each campus is able to select which BRR to use. Although results from various BRRs are generally treated as comparable by the district, this study suggests the results from various BRRs may not always be comparable, and the district may want to use caution when comparing BRR results across campuses that use different BRR assessments.

Key Findings

When looking at all English Benchmark Running Records and Renaissance 360, the matching proficient and not-proficient rates were similar. Seventy-six percent of students who were proficient on Renaissance 360 were also proficient on their campus' BRR assessment and 80% of students who were not proficient on Renaissance 360 were also not proficient on their campus' BRR assessment.

For English assessments, the match rate between Renaissance 360 and Benchmark Running Records varied based on the Benchmark Running Record used. Students who took the Scholastic and DRA2 BRRs were significantly more likely to have matching proficient results when compared to those that took the Next Steps and F&P BRRs. In contrast, students who took the DRA2, Next Steps, and F&P BRRs were more likely to have matching not-proficient results when compared to those that took the Scholastic BRR.

Spanish Benchmark Running Records were more like to have matching proficient rates than matching not-proficient rates. Ninety-four percent of students who were proficient on Renaissance 360 were also proficient on their campus' BRR assessment, while only 37% of students who were not proficient on Renaissance 360 were not proficient on their campus's BRR assessment.

For Spanish assessments, the match rate between Renaissance 360 and Benchmark Running Records did not vary based on the Benchmark Running Record used. Both the Sistema and Oral Reading Records BRRs provided similar matching proficient and not-proficient rates when compared to Renaissance 360.

Study Purpose

Campuses across the HISD utilize a variety of literacy assessments to measure student growth and achievement, including Renaissance 360 and BRRs. Renaissance 360, a universal screener, is used districtwide and taken by all kindergarten through 12th grade students at the beginning, middle, and end of the school year (BOY, MOY, and EOY, respectively).

BRRs also are used by campuses to measure student literacy performance for kindergarten through 5th grade students. Schools are permitted to choose which BRR they want to use for student monitoring, and at the end of each of the assessment periods (BOY, MOY, and EOY), a student's level of proficiency on the BRR is reported to the district. Both Renaissance 360 and BRRs are considered formative assessments.

Although student proficiency is reported to the district, at the time of this study (Fall 2018), the district did not maintain information about which BRR each campus used. As a result, student proficiency on BRRs was treated in the same manner across all students and campuses, without evidence that the tests were comparable.

To learn more about the comparability between BRRs across campuses, this study sought to collect data on which BRRs were being used by campuses around the district, and how their results compared to the results of Renaissance 360. The purpose of this methodological brief is to examine the comparability of literacy BRRs utilized by campuses in HISD.

Method

This analysis utilizes student-level, administrative data from HISD for the 2018-19 school year. These include assessment data from English and Spanish Renaissance 360 and BRRs at the BOY, MOY, and EOY. For Renaissance 360, the early literacy assessment is used for students in kindergarten and first grade, and the reading assessment is used for students in second through fifth grade. Data from a survey of school administrators was also used. This survey occurred in fall 2018 and asked about literacy assessments used on their campus (50 of 178 campuses with K-fifth-grade students participated). Table 1 shows the most frequently used BRRs across the district.

Table 1. Benchmark Running Records Assessments[1]	
English	**Spanish**
Scholastic Benchmark Book Running Record and Comprehension Assessment (Scholastic)	Sistema de evaluación de la lectura (Sistema)
Developmental Reading Assessment 2 (DRA2)	Spanish Oral Reading Records (Oral Reading Records)
Fountas & Pinnell Benchmark Assessment (F&P)	
Next Steps in Guided Reading (Next Steps)	

Note: Provided by HISD's Curriculum and Development Department

The sample is restricted to students in kindergarten through fifth grade who appear in the 2018-19 Renaissance 360 and BRR files at the EOY. Given these parameters, 53,727 students who participated in English assessments and 14,329 students who participated in Spanish assessments are included in this

analysis.[1] The sample is further limited to students enrolled on the 50 campuses where the type of BRR utilized is known, resulting in a final sample of 14,969 students who took Renaissance 360 and BRR in English and 3,559 students who took the assessments in Spanish.

Measures

Student proficiency: On Renaissance 360, students who received a result of "at/above benchmark" were coded as proficient, and students who received a result of "on watch," "intervention" or "urgent intervention" were coded as not proficient. On their campus' respective literacy BRRs, students who received a result of "advanced development" or "meeting expectations" were coded as proficient, and those who received a result of "more development needed" were coded as not proficient.

Benchmark Running Record: Administrators at schools with kindergarten through fifth grade were asked to report which English and Spanish BRR are used on their campus.

Analytic Plan

At each time point, student records from the Renaissance 360 and BRR files were linked in an effort to determine if students' performance (proficient or not proficient) was the same on Renaissance 360 and BRR. If a student was proficient (not proficient) on Renaissance 360 and proficient (not proficient) on their campus's BRR, they were considered to match. If the result of the Renaissance 360 did not match the result of the campus' BRR, they were considered a non-match (see Figure 1). Using the data from all students at a campus, we calculated a match rate for each BRR relative to the Renaissance 360, and compared across BRRs.

Figure 1. Conceptual model of literacy assessment match.[1]

		Renaissance 360 Result	
		Proficient	Not Proficient
BRR Result	**Proficient**	MATCH	MISMATCH
	Not Proficient	MISMATCH	MATCH

Comparing the match rates of BRRs with Renaissance 360 provides HISD with insight on how well BRRs from around the district are at predicting performance on Renaissance 360 and informs the degree to which BRRs from different campuses should be trusted to show they are measuring the same thing. If, for example, one BRR appears to have substantially higher rates of match with Renaissance 360 in identifying proficient/not proficient students than another BRR, this may suggest certain BRRs are measuring different, underlying constructs, and are not comparable.[2,3]

---

[1] The sample represents approximately two-thirds of kindergarten-fifth graders enrolled in HISD as of October 2018 (N=99,700). A majority of students are excluded from the analysis because they only have a score for either Renaissance 360 or BRR.

[2] Note: The purpose of this exercise is not to suggest that Renaissance 360 is better or worse than a BRR at assessing student proficiency. Instead, Renaissance 360 is used as a comparison tool because it is used universally across the district.

[3] These comparisons were also tested using a more formal statistical technique: the area under the ROC curve, a measure of diagnostic accuracy that show how well the BRR distinguishes among students who receive a proficient and not proficient result on Renaissance 360. Further detail about this analysis can be found in the accompanying online technical appendix.

1

When looking at all English BRRs and Renaissance 360, the matching proficient and not proficient rates were similar.

When we look at the English BRRs as a whole, 76% of students who were proficient on Renaissance 360 were also proficient on their campus' BRR assessment (see dark gray bar in Figure 2). In contrast, 80% of students who were not proficient on Renaissance 360 were also not proficient on their campus's BRR assessment (see light gray bar in Figure 2).

Figure 2. The rates of matching proficient and not proficient students are similar between English Renaissance 360 and BRRs.
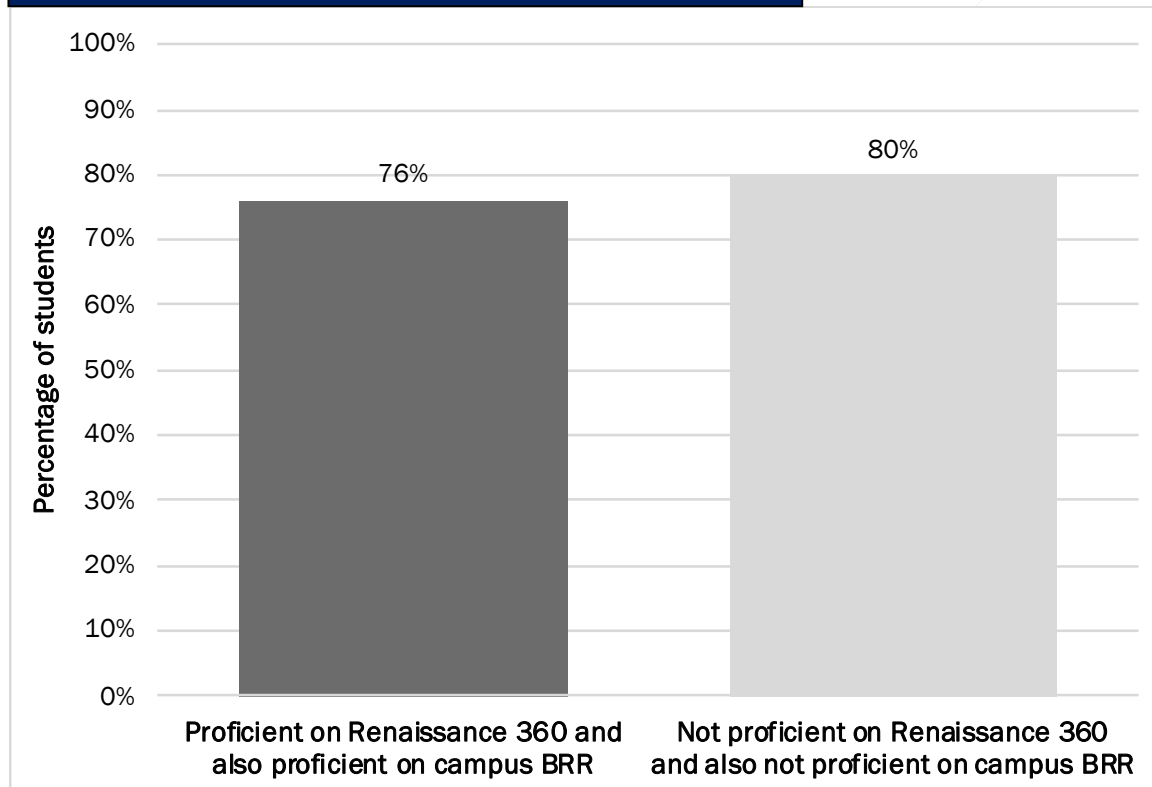
Sources: Administrative student-level data, 2018-19 and literacy survey results, fall 2018

2

For English assessments, the match rate between Renaissance 360 and BRRs varied based on the BRR used.

When results were analyzed for each BRR, results showed students who took the Scholastic and DRA2 BRRs were significantly more likely to have matching proficient results (see dark gray bars in Figure 3). Seventy-nine percent of students who were proficient on Renaissance 360 were also proficient on the Scholastic BRR, and 78% of students who were proficient on Renaissance 360 were also proficient on the DRA2 BRR. There was no significant difference found between the Scholastic and DRA2 BRRs. In contrast, only 70% of students who were proficient on Renaissance 360 were also proficient on the F&P and the Next Steps BRRs.

Students who took the Scholastic BRR were significantly less likely to have matching not proficient results with the Renaissance 360 than students who score not proficient on the other BRRs (see light gray bars in Figure 3). Seventy-seven percent of students who were not proficient on Renaissance 360 were not proficient on the Scholastic BRR, while in contrast about 87% of students who were not proficient on the Renaissance 360 were also not proficient on the Next Steps BRR, and that percent was slightly lower (though statistically significantly so) with regards to the F&P (83%) and DRA2 (82%) BRRs.

Figure 3. The type of English BRR used impacted the match rate with Renaissance 360.
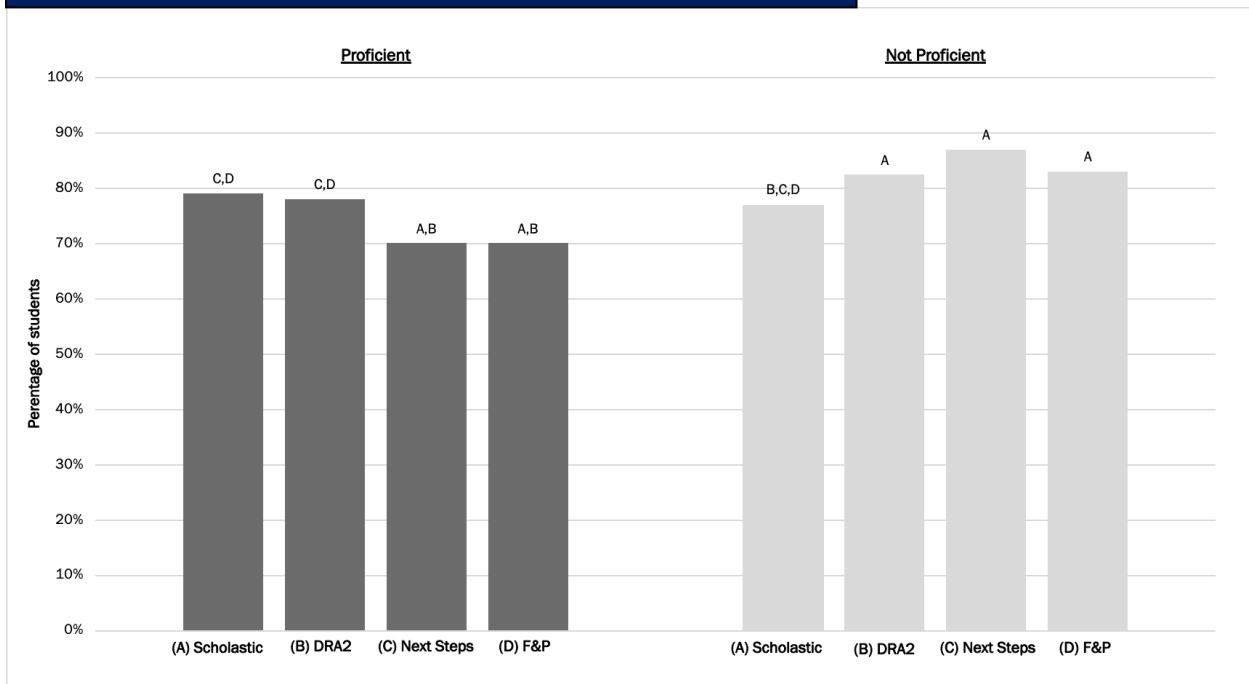
Notes: Significant at $p < 0.05$
[A] indicates significantly different results from Scholastic BRR; [B] indicates significantly different results from DRA2 BRR; [C] indicates significantly different results from Next Steps BRR; [D] indicates significantly different results from Fountas & Pinnell BRR
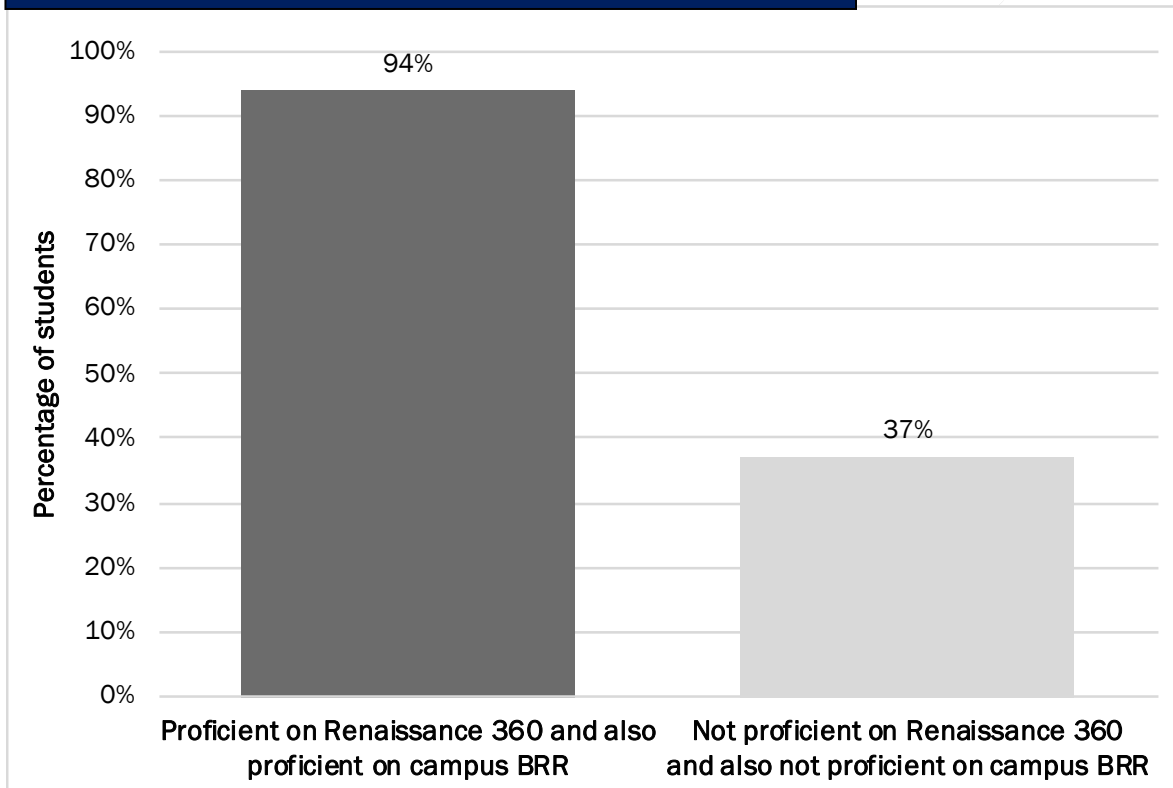**Sources:** Administrative student-level data, 2018-19 and literacy survey results, fall 2018

Spanish were more likely to have matching proficient rates than matching not proficient rates.

3

When we look at Spanish BRRs as a whole, 94% of students who were proficient on Renaissance 360 were also proficient on their campus's BRR assessment (see dark gray bar in Figure 4). In contrast, only 37% of students who were not proficient on Renaissance 360 were not proficient on their campus' BRR assessment (see light gray bar in Figure 4).

Figure 4. Spanish BRRs were more likely to have matching proficient rates than matching not proficient rates with Renaissance 360.

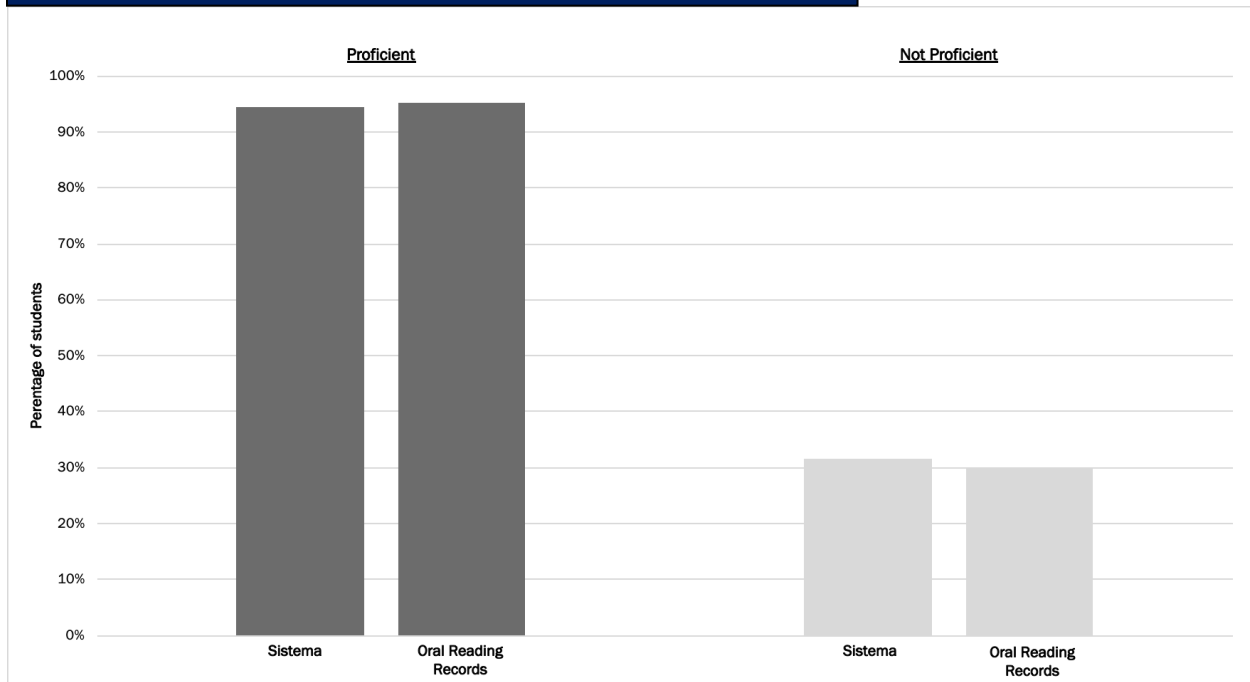**Sources:** Administrative student-level data, 2018-19 and literacy survey results, fall 2018

4 For Spanish assessments, the match rate between Renaissance 360 and BRRs did not vary based on the BRR used.

After analyzing results for each BRR, we found 95% of students who were proficient on Renaissance 360 were also proficient on the Sistema and Oral Reading Records BRRs (see dark gray bars in Figure 5). No significant difference in matching proficient results was found between those students who took the Sistema and Oral Reading Records BRRs.

Thirty-two percent of students who were not proficient on Renaissance 360 were not proficient on the Sistema BRR, and 30% of students who were not proficient on Renaissance 360 were not proficient on the Oral Reading Records BRR (see light gray bars in Figure 5). Similar to the results for proficient match rates, no significant difference was found between non-proficient match rates for students who took the Sistema and the Oral Reading Records BRRs.

Figure 5. Spanish BRRs had similar matching proficient and not proficient rates with Renaissance 360.

Note: Percentages for each BRR are not significantly different from one another at $p<0.05$
Sources: Administrative student-level data, 2018-19 and literacy survey results, fall 2018

The purpose of this methodological brief was to examine the comparability of various literacy BRRs utilized by HISD elementary campuses.

When looking at English assessments, the match rate (whether examining proficient match rates or not proficient match rates) between Renaissance 360 and BRRs significantly varied based on the BRR used at each campus. In contrast, when looking at Spanish assessments, the match rate between Renaissance 360 and BRRs did not vary based on the BRR used at each campus. However, Spanish BRRs had higher proficient match rates than not proficient match rates.

These results suggest the district should use caution in comparing BRR results across the district, particularly those given in English. The various BRRs used throughout the district may not be providing comparable results, making it inappropriate to compare a proficient score at one campus to a proficient score at another. This may partly reflect that the various BRRs may be assessing different domains of knowledge of students (see Appendix Table 5).

There is a limitation of this study to note. When comparing the results from different literacy BRRs, only those students who attended schools with administrators who responded to the survey were included. Therefore, our analysis of the comparability of different assessments is limited to only a portion of those students who took both the Renaissance 360 and BRR assessments.

Despite these limitations, this study provides an important examination of the comparability of various literacy BRRs. The findings in this report provide a starting point for HISD to consider whether to use BRRs as a means for making comparisons between campuses as it relates to student literacy proficiency. If the district would like to use BRR results for this purpose, a recommendation is for the district to start collecting information about which BRR is used at each campus. HISD might also consider collecting the scale score for these BRRs at each time period, rather than only an indicator of level of proficiency. This would allow for the district to account for the use of different BRRs at each campus and have a more nuanced understanding of the similarities and variations between campuses as it relates to literacy performance and growth.

Recommendations

The district should start collecting information about which BRR is used at each campus. Collecting information about which BRR is used at each campus would allow the district to account for different assessments being used at individual campuses when examining differences in BRR proficiency rates across campuses.

The district should collect more detailed achievement information (e.g., scaled scores). Currently, the district only knows whether a student is proficient or not through campus-reported BRR results. More detailed information, such as the actual score received on the assessment or a students' oral-reading fluency, may help the district better understand what "proficient" means for each assessment and understand where students are making the greatest gains in literacy.

About HERC. Focusing on the most pressing challenges facing the region, the Houston Education Research Consortium (HERC) is a research-practice partnership between Rice University and 11 Houston-area school districts. HERC research is developed directly alongside district leaders with findings shared with decision makers – culminating in long-term, equity-minded solutions, opportunities and growth for Houston and beyond.

Houston Education Research Consortium
a program of the Kinder Institute for Urban Research
MS-258 Rice University **|** Houston, Texas 77005
713-348-2532 **|** herc@rice.edu
Find us online:
kinder.rice.edu/houston-education-research-consortium

Table 1. Accuracy of benchmark running record scores for predicting 360 proficiency, English assessments

Measure	EOY
Percentage of students with matching results on Renaissance 360 and the benchmark running record	78%
Percentage of proficient Renaissance 360 students who are also found to be proficient on the benchmark running record	76%
Percentage of not proficient Renaissance 360 students who are also found to be not proficient on the benchmark running record	80%
Percentage of students who are proficient on Renaissance 360	48%
Percentage of students who are proficient on the benchmark running record	50%
Number of students	53,727

Sources: Administrative student-level data, 2018-19 and literacy survey results, fall 2018

Appendix

Table 2. Accuracy of benchmark running record scores for predicting Renaissance 360 proficiency by assessment type, English assessments				
	(A) Scholastic	(B) DRA2	(C) Next Steps	(D) F&P
Measure	EOY	EOY	EOY	EOY
Percentage of students with matching results on Renaissance 360 and the benchmark running record	78%	80%	79%	77%
Percentage of proficient Renaissance 360 students who are also found to be proficient on the benchmark running record	79% [C,D]	78% [C,D]	70% [A,B]	70% [A,B]
Percentage of not proficient Renaissance 360 students who are also found to be not proficient on the benchmark running record	77% [B,C,D]	82% [A]	87% [A]	83% [A]
Percentage of students who are proficient on Renaissance 360	52% [C,D]	50% [C,D]	41% [A,B]	43% [A,B]
Percentage of students who are proficient on the benchmark running record	53% [C,D]	54% [C,D]	49% [D]	49% [A,B,C]
Number of students	9,802	1,932	1,524	1,711

Notes: Significant at p<0.05
[A] indicates significantly different results from Scholastic BRR; [B] indicates significantly different results from DRA2 BRR; [C] indicates significantly different results from Next Steps BRR; [D] indicates significantly different results from Fountas & Pinnell BRR
Sources: Administrative student-level data, 2018-19 and literacy survey results, fall 2018

Table 3. Accuracy of benchmark running record scores for predicting Renaissance 360 proficiency, Spanish assessments	
Measure	**EOY**
Percentage of students with matching results on Renaissance 360 and the benchmark running record	68%
Percentage of proficient Renaissance 360 students who are also found to be proficient on the benchmark running record	94%
Percentage of not proficient Renaissance 360 students who are also found to be not proficient on the benchmark running record	37%
Percentage of students who are proficient on Renaissance 360	80%
Percentage of students who are proficient on the benchmark running record	54%
Number of students	14,389

Sources: Administrative student-level data, 2018-19 and literacy survey results, fall 2018

Table 4. Accuracy of benchmark running record scores for predicting Renaissance 360 proficiency by assessment type, Spanish assessments

Measure	(A) Sistema EOY	(B) Oral Reading Records EOY
Percentage of students with matching results on Renaissance 360 and the benchmark running record	68%	70%
Percentage of proficient Renaissance 360 students who are also found to be proficient on the benchmark running record	95%	95%
Percentage of not proficient Renaissance 360 students who are also found to be not proficient on the benchmark running record	32%	30%
Percentage of students who are proficient on Renaissance 360	84%	86%
Percentage of students who are proficient on the benchmark running record	58%	62%
Number of students	3,041	518

Notes: Significant at $p<0.05$
[A] indicates significantly different results from Spanish Oral Reading Records BRR; [B] indicates significantly different results from Sistema de evaluación de la lectura BRR
Sources: Administrative student-level data, 2018-19 and literacy survey results, fall 2018

Table 5. Topics assessed by each benchmark running record, as described by vendor

Benchmark Running Record	Topics covered
English	
Scholastic Benchmark Book Running Record and Comprehension Assessment[4]	measures reading accuracy, fluency, and comprehension
Developmental Reading Assessment 2[5]	measures reading engagement, oral reading fluency, accuracy, and rate, and comprehension
Next Steps in Guided Reading[6]	measures reading interest, word knowledge, phonics, fluency, and comprehension
Fountas & Pinnell Benchmark Assessment[7]	measures reading accuracy, fluency, and comprehension through various means, including writing about reading, in-depth fluency assessments, phonemic awareness assessments, letter name assessments, early literacy behaviors assessments, phonics and word analysis assessments, high frequency word assessments, and vocabulary assessments
Spanish	
Sistema de evaluación de la lectura[8]	measures Spanish decoding, fluency, vocabulary, and comprehension skills
Spanish Oral Reading Records[9]	measures fluency, error rates, accuracy, and comprehension

---

[4] *District Pre-Approved Assessment Blueprint, 2015-16*. Houston Independent School District. https://www.houstonisd.org/cms/lib2/TX01001591/Centricity/Domain/16084//Running%20Records/Running%20Records%20Blueprint%20DPA.pdf

[5] *Developmental Reading Assessment | Second Edition Plus.* Pearson Assessments. https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Academic-Learning/Reading/Developmental-Reading-Assessment-%7C-Second-Edition-PLUS/p/100001222.html?tab=product-details

[6] *Next Step in Guided Reading Assessment.* Scholastic. http://teacher.scholastic.com/education/next-step-guided-reading-assessment/index.html

[7] *Benchmark Assessment System (BAS)*. Fountas and Pinnell. *https://www.fountasandpinnell.com/bas/*

[8] *Sistema de evaluación de la lectura (SEL)*. Fountas and Pinnell. *https://www.fountasandpinnell.com/sel/*

[9] *Oral Reading Records.* Benchmark Education. https://benchmarkeducation.com/teachers/series/oral-reading-records