Equating Words-Correct-Per-Minute (WCPM) Scores Across Passages of MAP Reading Fluency

March 2020

Jing Chen, Psychometric Solutions
Mary Ann Simpson, Psychometric Solutions

© 2020 NWEA. NWEA and MAP Growth are registered trademarks of NWEA in the U.S. and in other countries. All rights reserved. No part of this document may be modified or further distributed without written permission from NWEA.

Suggested citation: Chen, J., & Simpson, M. A. (2020). *Equating words-correct-per-minute (WCPM) scores across passages of MAP Reading Fluency*. Portland, OR: NWEA.

Table of Contents

List of Tables

List of Figures

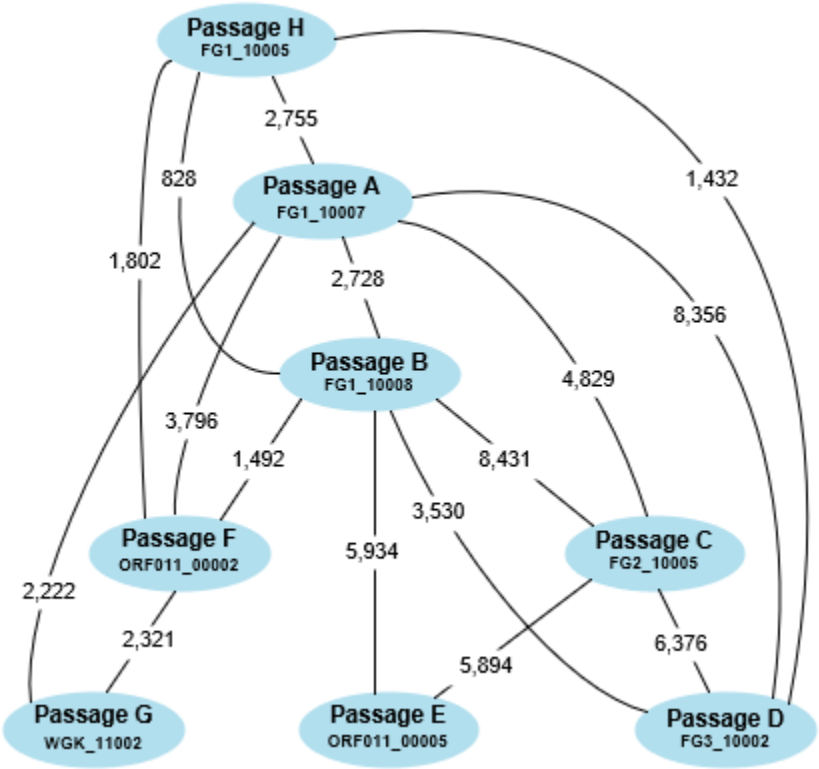
Executive Summary

NWEA® equated words-correct-per-minute (WCPM) scores from oral reading passages included in the NWEA MAP® Reading Fluency™ assessment. Equipercentile equating with loglinear pre-smoothing was applied to convert raw WCPM scores from a non-reference passage to those from a reference passage. The goals of this study were to (1) develop a method to choose a reference passage and passage pairs for equating WCPM scores from a large number of passages and (2) determine if equated WCPM scores provide a more accurate indication of students' oral reading fluency ability compared to the raw WCPM scores from passages that vary in difficulty.

Data for this study were from students in Grades K–3 who took the adaptive oral reading forms of MAP Reading Fluency in Fall 2018, Winter 2019, and Spring 2019. The number of student records available for analysis ranged from 29,028 for fall and 59,773 for spring. There was considerable overlap across terms, with 84,262 unique students represented.

A single-group equating design with equipercentile equating was used to equate the 60 passages included in this study. Equating such a large number of passages would be daunting without a shortcut, so graph theory was used to identify the reference passage and passages requiring the least amount of chaining to "reach" the reference passage. For example, Figure E.1 shows that Passages A and B are the best connected to the other passages. Each passage has six paths to other passages. Passage A was chosen as the Fall 2018 reference passage because it was of medium difficulty and had more respondents than Passage B. The most direct path from a passage to the reference passage suggests which other passage to pair it with. For example, the shortest path from Passage E to Passage A is through Passage C.

Figure E.1. Fall 2018 Passage Pairs



Results suggest that equating greatly reduced or eliminated score fluctuations due to passage difficulty. Equated WCPM scores were more highly correlated across testing occasions than their raw counterparts and showed much less within-student variance, as shown in Table E.1 and Table E.2. Finally, Table E.3 suggests that equated WCPM scores are not affected by the changes in passage difficulty that occur within or across testing terms. This finding suggests that equated WCPM scores are especially suited to progress monitoring.

Table E.1. Correlations Between Students' Raw and Equated WCPM Scores

Passage Pair	N	Correlations	
		Raw WCPM	Equated WCPM
Fall 2018			
p1_p2	22,073	0.865	0.894
p1_p3	20,688	0.845	0.882
p2_p3	19,955	0.872	0.902
Winter 2019			
p1_p2	43,249	0.896	0.911
p1_p3	41,166	0.844	0.890
p2_p3	40,709	0.851	0.907
Spring 2019			
p1_p2	50,853	0.878	0.913
p1_p3	48,309	0.851	0.896
p2_p3	47,899	0.841	0.906

Table E.2. SD of Raw and Equated WCPM Scores

Term	Average Within-Student SD	
	Raw WCPM	Equated WCPM
Fall 2018	12.19	7.59
Winter 2019	9.52	6.94
Spring 2019	9.92	7.69

Table E.3. Change of Students' WCPM Scores Across Terms by Passage Position

Passage	N	Raw WCPM		Equated WCPM	
		Mean	SD	Mean	SD
Spring 2019 – Winter 2019					
$P1_{s19}$-$P1_{w19}$	33,215	7.19	17.28	6.70	16.84
$P2_{s19}$-$P2_{w19}$	31,591	7.06	18.26	6.36	16.97
$P3_{s19}$-$P3_{w19}$	29,615	8.32	23.07	6.44	17.30
Winter 2019 – Fall 2018					
$P1_{w19}$-$P1_{f18}$	16,882	0.31	19.37	12.63	16.81
$P2_{w19}$-$P2_{f18}$	15,275	9.77	19.52	11.95	17.34
$P3_{w19}$-$P3_{f18}$	14,282	11.21	21.93	12.38	17.78
Spring 2019 – Fall 2018					
$P1_{s19}$-$P1_{f18}$	14,397	7.23	21.31	18.17	19.50
$P2_{s19}$-$P2_{f18}$	13,083	14.99	22.98	17.71	20.16
$P3_{s19}$-$P3_{f18}$	12,180	18.58	23.71	18.52	20.33

1. Introduction

In 2019, NWEA® evaluated the feasibility of equating words-correct-per-minute (WCPM) scores from the oral reading portions of the MAP® Reading Fluency™ assessment, an early reading assessment developed by NWEA. The test adapts to accommodate pre-readers, early readers, and independent readers in Grades K–3. Equipercentile equating with loglinear pre-smoothing was applied to convert raw WCPM scores from non-reference passages to those from a reference passage. The goals of this equating study were as follows:

1. Develop a method to quickly choose a reference passage and passage pairs for equating WCPM scores from a large number of passages
2. Evaluate equating results to see if equated scores provide a more accurate indication of students' oral reading fluency ability compared to raw WCPM scores from passages that vary in difficulty

Reading fluency is an important skill in reading development and connects tightly with reading comprehension (Young & Bowers, 1995; Nathan & Stanovich, 1991). Many researchers have suggested using fluency as an indicator of reading progress because of the relative ease of assessing fluency over comprehension (Burns et al., 2002; Fuchs, Fuchs, Hosp, & Jenkins, 2001). Oral reading fluency measures are designed to assess students' oral reading speed and accuracy and monitor their reading progress. A student reads one or more brief passages aloud for several minutes, and the resulting WCPM score serves as a formative indicator of a student's oral reading fluency.

However, raw WCPM scores are problematic for progress monitoring because passages vary in difficulty. Even with careful selection and development procedures, it is difficult to control passage properties to ensure similar difficulty. Readability formulas have been heavily relied on for selecting equivalently difficult passages (Ardoin, Suldo, Witt, Aldrich, & McDonald, 2005), but readability scores such as Lexile® scores are imperfect for equating passages because the rank ordering of passages will vary depending on the readability index used. More importantly, readability formulas have limited utility for predicting actual reading performance across passages (Francis et al., 2008).

Statistical equating methods provide ways to account for text difficulty differences to ensure that WCPM scores from different passages can be used interchangeably. Large numbers of passages are often used for monitoring progress throughout the year and across grades, and it is challenging to equate scores from a large number of passages and put them all on a common scale. Thus, NWEA developed a method based on graph theory to identify the pairwise relationships between passages for equating. Such equated scores provide a more accurate indication of students' oral reading fluency compared to raw WCPM scores by accounting for differences in passage difficulty.

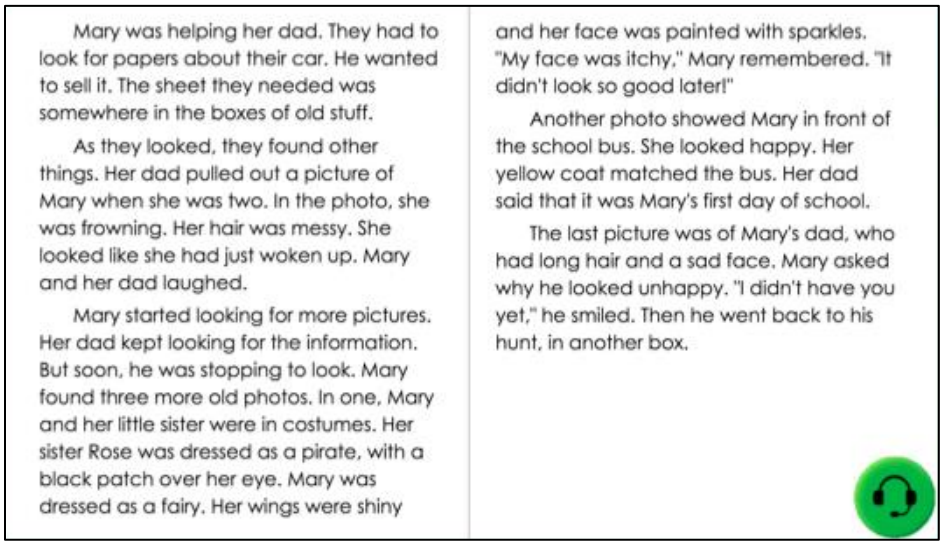
1.1. Assessment Overview

MAP Reading Fluency measures oral reading fluency (i.e., the ability to read text aloud quickly, accurately, and with good inflection); decoding accuracy (i.e., the ability to translate a printed letter or word into a sound); and literal comprehension (i.e., the ability to understand the meaning of a passage). These areas are evaluated based on oral reading of up to three passages. From these results, a reader profile and recommended next steps are generated.

MAP Reading Fluency can be administered in a group setting rather than one-on-one. Students wear headsets with microphones and read the test content out loud. The audio is recorded, scored by a speech recognition-based scoring engine, and saved for future playback. Each student reads a short picture book to get started. Then they read sentences silently and identify a matching picture, which gauges if the student is ready to read passages out loud.

If the student is ready to read passages out loud, they read up to three passages, each with 200 words. The full text of the passage is presented onscreen without the need for scrolling or page turning. Students use a button to indicate when they are finished and are given up to five minutes to read each passage. After reading a passage, students answer six selected-response literal or low-inference items (i.e., simple reading comprehension items) to demonstrate their comprehension. Figure 1.1 presents a sample passage, and Figure 1.2 presents sample comprehension items. Passages are presented to students adaptively according to their grade level and performance on these comprehension items. In general, higher grade levels and better performance on the comprehension items bring harder passages, and vice versa.

Figure 1.1. Sample Passage



Mary was helping her dad. They had to look for papers about their car. He wanted to sell it. The sheet they needed was somewhere in the boxes of old stuff.

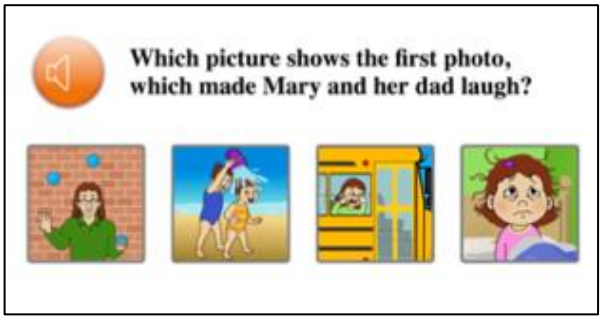
As they looked, they found other things. Her dad pulled out a picture of Mary when she was two. In the photo, she was frowning. Her hair was messy. She looked like she had just woken up. Mary and her dad laughed.

Mary started looking for more pictures. Her dad kept looking for the information. But soon, he was stopping to look. Mary found three more old photos. In one, Mary and her little sister were in costumes. Her sister Rose was dressed as a pirate, with a black patch over her eye. Mary was dressed as a fairy. Her wings were shiny and her face was painted with sparkles. "My face was itchy," Mary remembered. "It didn't look so good later!"

Another photo showed Mary in front of the school bus. She looked happy. Her yellow coat matched the bus. Her dad said that it was Mary's first day of school.

The last picture was of Mary's dad, who had long hair and a sad face. Mary asked why he looked unhappy. "I didn't have you yet," he smiled. Then he went back to his hunt, in another box.

Figure 1.2. Sample Comprehension Items



For these students who read the passages, the report shows the raw WCPM, decoding accuracy, and reading comprehension scores. Automated scores for the oral reading include the raw WCPM score and the decoding accuracy (i.e., words correct ratio that is independent of the time taken to read the passage and indicates how much of the passage the student read out loud correctly, such as 75%). The reading comprehension score is the percent correct out of the six items.

If the student is not ready to read passages, a series of measures are presented that assess foundational reading skills, including phonological awareness, early phonics and word recognition skills, listening comprehension, and picture vocabulary. For these students who were not ready for passages, the foundational skills report shows their proficiency in decoding and oral language comprehension.

1.2. Passages Included in this Study

MAP Reading Fluency passages were developed at varying levels of text complexity, as gauged by their Lexile® measures. Passage length could vary by grade but was constrained by screen real estate; no passages requiring scrolling or page turning were included. Passages were reviewed by experts in primary grades literacy assessment for quality and age-appropriate content, form, and tone. They were reviewed separately for any issues with bias or sensitivity. The Lexile® measures of passages were used to guide the selection of passages to be administered at each grade level.

Table 1.1 presents the 60 passages that were included in this study, including their Lexile® measures. Each passage was administered in at least one term in 2018–2019.

Table 1.1. Descriptive Statistics for 60 Oral Reading Passages

Passage ID	Passage Title	Lexile®
Fall 2018		
FG1_10007	Be a Teacher	380
FG1_10005	Losing Teeth	400
FG1_10008	Airplanes	470
FG3_10002	Drinking Fountain	470
FG2_10005	Butterflies and Moths	500
ORF011_00005	Emperor Penguins	560
Fall 2018, Spring 2019		
WGK_11002	Pink the Pig	200
ORF011_00002	Birds and Nests	220
Winter 2019		
ORF011_00016	That's No Bug	190
ORF011_00018	Little Cat	190
ORF011_00001	Ann's Bear	210
ORF011_00012	Bus Stop	210
ORF011_00020	Art on a Plate	210
WGK_11001	Sal Gets Wet	210
ORF011_00013	Ben's Flag	270
ORF011_00008	Tell Time	290
ORF011_00014	Snack Time	360
ORF011_00011	Lost Coat	370

Passage ID	Passage Title	Lexile®
ORF011_00021	Crickets	390
ORF011_00023	Skate	390
FG1_10006	Bears	410
FG1_10004	Jay and Gus	460
ORF011_00026	Cleaning Our Room	460
ORF011_00025	Bike Ride	470
FG3_10008	Playground Alien	500
ORF011_00028	Tree House	580
FG3_10005	Dad Versus Socks	590
ORF011_00006	Grandma Babysits	610
ORF011_00007	Movie Magic	620
ORF011_00019	Birdwatchers	620
ORF011_00022	Rubber Bands	720
ORF011_00009	Lava Monster	730
ORF011_00010	Owls	760
Winter 2019, Spring 2019		
ORF011_00027	Casey's Walk	440
Spring 2019		
MRF011_1774_00001	The Box	180
ORF011_00015	Jump Rope	180
ORF011_00003	Zack in the Rain	220
MRF011_1769_00001	Paper Jet	260
MRF011_1801_00001	Fancy Pie	270
MRF011_1757_00001	John Loved the Moon	350
MRF011_1773_00001	Sore Throat	370
MRF011_1753_00001	Alex's Collection	380
MRF011_1775_00001	The Class Garden	390
ORF011_00024	Parker the Peacock	460
FG2_10007	Game Inventor	480
FG1_10003	Bus Love	490
FG2_10003	Old Photos	490
FG2_10009	Pam and the Toy Chest	490
FG2_10006	Class Trip	500
FG2_10008	Spell Pizza	500
FG3_10003	Hamster on the Loose	500
FG3_10009	Popcorn Science	520
ORF011_00004	Blue Whales	520
FG3_10007	Training a Puppy	540
MRF011_1814_00001	Kangaroo	600
FG3_10001	Field Mice	610
MRF011_1806_00001	Monster Baby	630
MRF011_1813_00001	Bricks	710
MRF011_1771_00001	Shopping for Food	740
MRF011_1812_00001	What's in the Mirror	770

2. Method

2.1. Data

Data for this study are from students who took the adaptive oral reading forms of MAP Reading Fluency in Fall 2018, Winter 2019, and Spring 2019, as shown in Table 2.1. “Overall” indicates the total number of unique student IDs across all three terms. In each term, thousands of students across over 40 states took the adaptive oral reading forms. Table 2.2 presents the demographic distribution of the student sample by gender and ethnicity. The samples are close to the demographic characteristics of the U.S. national K–3 student population based on the U.S. Census Bureau’s 2017 demographic data (U.S. Census Bureau, 2017).

Table 2.1. Student Sample

		%Students by Grade									
		Grade K		Grade 1		Grade 2		Grade 3		N/A or other	
Term	Total N	N	%	N	%	N	%	N	%	N	%
Fall 2018	29,028	121	0.4	3,017	10.4	12,051	41.5	13,839	47.7	–	–
Winter 2019	56,343	42	0.1	12,092	21.5	26,405	46.9	15,564	27.6	2,240	4.0
Spring 2019	59,773	2,573	4.3	15,335	25.7	25,705	43.0	11,601	19.4	2,596	7.6
Overall	82,462	2,736	3.2	30,444	23.9	64,161	42.8	41,004	26.4	4,836	3.6

Table 2.2. Student Sample Demographics

	%Students by Term			
Demographic Subgroup	Fall 2018	Winter 2019	Spring 2019	Overall
Gender				
Female	49.8	49.6	50.8	49.7
Male	50.2	50.4	49.2	50.3
Ethnicity				
American Indian or Alaskan	1.9	1.5	3.0	1.7
Asian	2.8	7.1	3.5	6.3
Black	16.7	16.9	14.8	16.7
Hispanic	13.7	15.7	13.3	15.4
Multiethnic	4.0	4.7	4.1	4.5
Native Hawaiian/Other Pacific Islander	0.2	0.3	0.2	0.2
Not Specified or Other	10.6	7.1	11.4	8.7
White	50.0	46.7	49.6	46.5

2.2. Equating Procedure

To equate all passages, equipercentile equating with loglinear pre-smoothing was applied to convert raw WCPM scores from a non-reference passage to those from a reference passage following the steps below:

1. Choose equating design
2. Identify the reference passage in each term
3. Define the shortest path to the reference passage
4. Choose equating method
5. Equate passages across terms through “chain” passage pairs

All equating processes were conducted using data from students across grades because the relationship between the WCPM scores of passages is assumed to remain the same across grades. Otherwise grade-specific equating relationships using data from a specific grade would have been needed, which is difficult to implement and even more difficult to justify.

Outliers were excluded from the data when building the equating relationship between a pair of passages to build more reasonable relationships. They were identified by the Mahalanobis distance of < -10 or > 10, a statistic that helps find observations that are outlying on all variables involved in an analysis. Linear interpolation was used to identify integer score points in chained equatings, and linear extrapolation was used to build the equating relationship beyond the range of the scores in the data to produce plausible results. Conversion tables for reported scaled WCPM (SWCPM) scores were capped at 20 SWCPM at the low end and 170 SWCPM at the high end because very low SWCPM scores are likely unreliable, and the high-end caps were introduced to prevent over-interpretation of SWCPM scores. Oral reading fluency manifests a "good enough" quality beyond which extra speed offers little further benefit to the reader.

2.2.1. Choose Equating Design

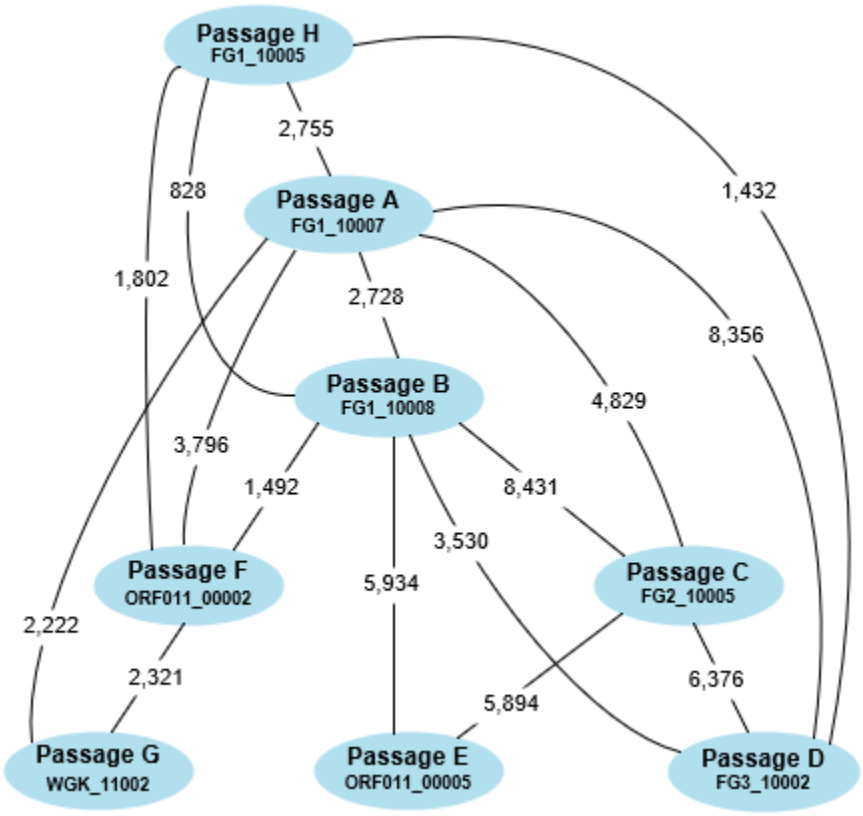
A reference passage (i.e., anchor passage) was needed to place all 60 passages on a common scale. Given the adaptive nature of MAP Reading Fluency, a single group design was adopted where the equating relationship is built based on the scores from a pair of passages read by the same student during one test event. Although no official "reference passage" had been worked into the original test design, NWEA developed a method to identify a representative reference passage to which all the other passages could be equated. With this design, a new passage can be equated if enough students read the new passage and an existing passage has already been equated. This design does not require administering the exact same test forms or the exact same reference passages.

2.2.2. Identify the Reference Passage in Each Term

Graph theory is a mathematically structured way to visualize networks of objects (exSTEMsions, 2019). In graph theory, objects in a graph comprise nodes and edges. In this analysis, nodes are passages, and the edges identify pairs of passages read by the same students. To identify the reference passage in each term, all possible passage pairs were first identified. Figure 2.1 presents a graph representation of passage pairs from Fall 2018. The number along each line is the number of students that took each passage pair and received valid WCPM scores on both passages (i.e., WCPM score >0). For example, 2,755 students took Passage A and Passage H. A small proportion of students with invalid WCPM scores from either passage in each pair were excluded from all data analyses. All the passages administered in the term and all students across grades that took each passage pair were included.

A reference passage was then selected based on having a greater number of edges in the graph, adequate sample sizes for most pairs, and medium text difficulty. A larger number of edges for a passage indicates more frequent pairing with other passages. For example, Passages A and B both have six edges in Figure 2.1. They connect to most other passages and are good candidates for the reference passage. However, Passage A has more data than Passage B, and the Lexile® measure of Passage A is closer to the mean of the Lexile® measures of all the passages. The Lexile® measures of all the eight passages administered in Fall 2018 were from 200L to 560L. The Lexile® measure of Passage A was 380L, while that of Passage B was 470L. Therefore, Passage A (FG1_10007) was selected as the reference passage for Fall 2018. All other passages administered in this term were equated to this reference passage.

Figure 2.1. Fall 2018 Passage Pairs



The reference passages for Winter 2019 and Spring 2019 were selected following this same process. The reference passages selected for each term are as follows:

- Fall 2018: FG1_10007
- Winter 2019: FG1_10006
- Spring 2019: FG2_10006

2.2.3. Define the Shortest Path to the Reference Passage

As shown in Figure 2.1, two passages can be connected through more than one path, which makes it complicated to equate scores from one passage to the scores of the reference passage. Graph theory provides a way to model this type of pairwise relationship by defining the shortest path between two objects. Based on graph theory, NWEA defined the shortest path from a passage to the reference passage as the path with the most data points and fewest connecting nodes. This shortest path results in the least amount of equating error because the equating relationship will be built based on more direct connections and larger sample sizes, making it desirable. For example, in Figure 2.1, the shortest path from Passage E to Passage A is through Passage C rather than other paths. The WCPM scores of Passage E can be equated to the WCPM scores of Passage C and then to the scores of Passage A, the reference passage, through the shortest path.

2.2.4. Choose Equating Method

The equipercentile equating method with loglinear pre-smoothing was used to build the equating relationships by identifying the WCPM scores from each passage with the same percentile ranks as the WCPM scores from a reference passage. The equipercentile equating method allows for greater flexibility for differences in difficulty between passages than other equating methods such as mean or linear equating. Previous research (Stoolmiller, Biancarosa, & Fien, 2012) also suggested equipercentile equating would be the best method compared to mean and linear equating for equating the raw WCPM scores.

A large sample size is required for equipercentile equating to achieve small equating errors. Kolen and Brennan (1995) suggested a sample size of 1,500. Therefore, the sample size of students with valid WCPM scores from both passages was around 1,500 or more for each passage pair. Conversion tables were then created to place scores from each passage onto the scale of the reference passage. All data preparation, smoothing, and equating was accomplished with the equate R package (Albano, 2016).

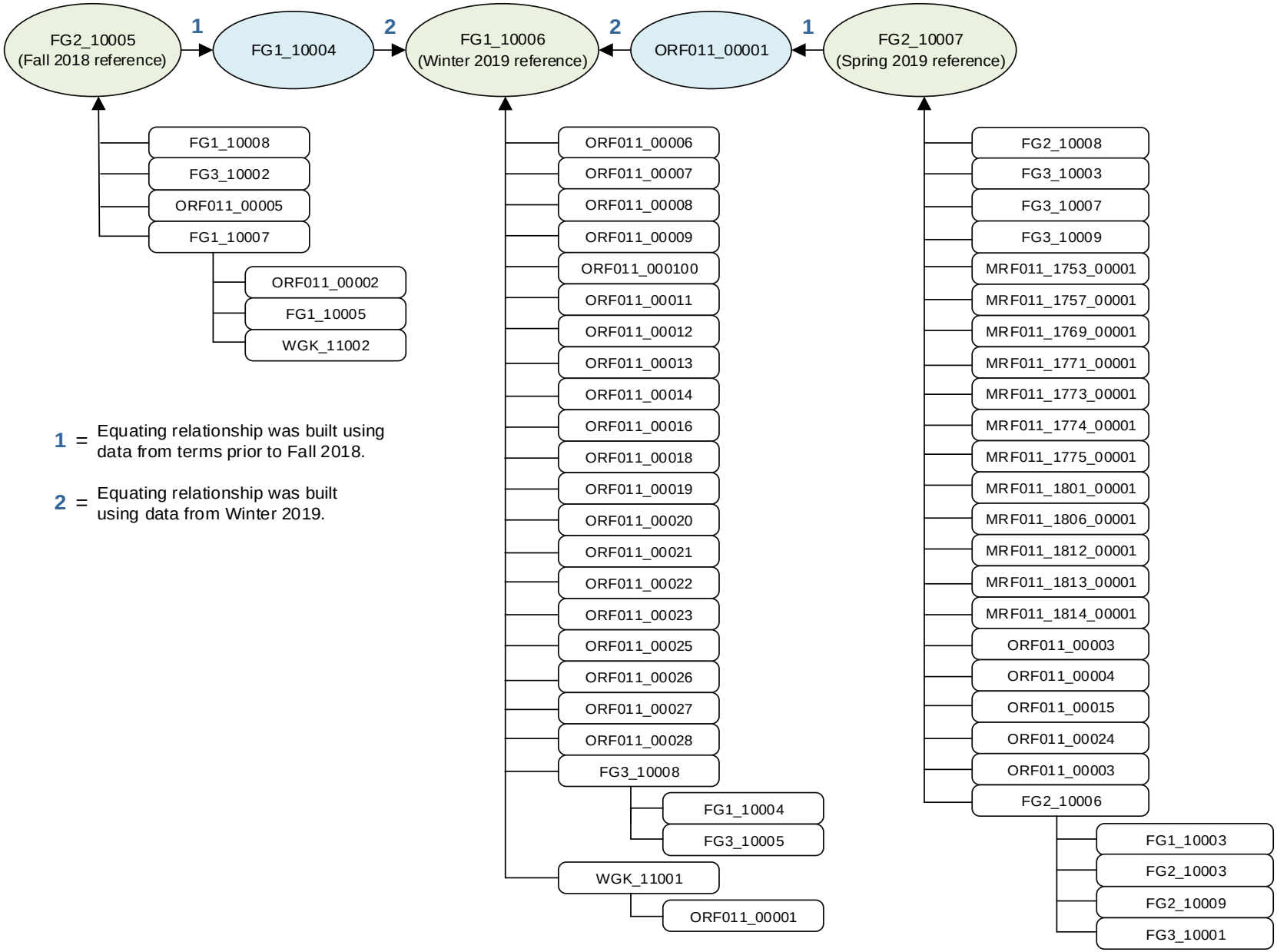
2.2.5. Scale Passages Across Terms Through "Chain" Passage Pairs

After passages within a term are equated to a common scale, the scales across all three terms must be aligned to place all the passages across terms on one final common scale. To accomplish this, NWEA identified the reference passage of Winter 2019 (FG1_10006) as the final reference passage for all three terms because it connects to a large number of passages, has adequate sample sizes for most pairs, and has medium text difficulty. The Fall 2018 and Spring 2019 passages were then equated to the final reference passage using passage pairs that worked as "chains" to connect the scales across terms.

Practice effect occurs when a student reads the same passage in all three terms, thereby increasing their score artificially. To reduce any practice effect, each passage was only administered in one of the three terms so that each passage only had pairwise connections with other passages of the same term. However, in previous terms (i.e., before Fall 2018), many of the passages had been administered and two passage pairs were identified as the "chains" to connect the scales across terms.

Figure 2.2 presents how the passages across three terms are connected through the chains, represented by the blue ovals. Passages FG1_10004 and ORF011_00001 worked as the chains to equate the Fall 2018 and Spring 2019 passages to the final reference passage, respectively. Passages FG2_10005 and FG1_10004 were administered together in previous terms. There were 1,102 students who took these two passages as a pair and received valid WCPM scores for both passages. Using these data, scores of FG2_10005 could be equated to scores of FG1_10004. Also, because Passages FG1_10004 and FG1_10006 were administered together in Winter 2019, scores of FG1_10004 could be equated to scores of FG1_10006. With these two equatings, scores from FG2_10005 could then be equated to scores on FG1_10006. In this way, Fall 2018 and Spring 2019 WCPM scores were equated to those from Winter 2019.

Figure 2.2. Equating all Passages across Terms through Passage Bridges



3. Results

3.1. Reliability Evidence

Table 3.1 presents the correlations between the WCPM scores from the three passages that the same student read. “Passage Pair” shows the pair of passages of the correlation. For example, p1_p2 is the first and the second passage that a student read. The correlations between students’ equated WCPM scores are higher than the correlations between students’ raw WCPM scores, suggesting that the equated scores are more reliable.

Table 3.1. Correlations Between Students’ Raw and Equated WCPM Scores

		Correlations	
Passage Pair	N	Raw WCPM	Equated WCPM
Fall 2018			
p1_p2	22,073	0.865	0.894
p1_p3	20,688	0.845	0.882
p2_p3	19,955	0.872	0.902
Winter 2019			
p1_p2	43,249	0.896	0.911
p1_p3	41,166	0.844	0.890
p2_p3	40,709	0.851	0.907
Spring 2019			
p1_p2	50,853	0.878	0.913
p1_p3	48,309	0.851	0.896
p2_p3	47,899	0.841	0.906

Table 3.2 presents the standard deviation (SD) of each student’s three WCPM scores (i.e., one score per passage) averaged across students. Compared with raw scores, students’ equated WCPM scores were much less variable. This suggests that the equating process has eliminated some fluctuations driven by text difficulty.

Table 3.2. SD of Raw and Equated WCPM Scores

	Average Within-Student SD	
Term	Raw WCPM	Equated WCPM
Fall 2018	12.19	7.59
Winter 2019	9.52	6.94
Spring 2019	9.92	7.69

3.2. Residuals

The final reference passage from Winter 2019 (FG1_10006) was administered together with 22 other passages as a pair on test forms. This means some students took the final reference passage and a non-reference passage together, resulting in their raw WCPM scores from the reference passage and their equated WCPM scores from a non-reference passage (i.e., equated to the scores of the reference passage). Residuals were calculated to find the difference between the actual WCPM scores from the reference passage and both the raw and equated WCPM scores from the non-reference passage to determine which score from the non-reference passage more closely resembles the raw score from the reference passage. Smaller residuals indicate greater success in equating.

Table 3.3 presents the residual sum of squares between students' scores on the reference passage and their scores from a non-reference passage averaged across students for both raw and equated WCPM scores. Only 22 passages were administered together with the final reference passage (FG1_10006), so the results in Table 3.3 are from those passages only.

Differences between students' equated scores from a non-reference passage and their actual scores on the reference passage are much smaller than the differences between students' raw scores on the non-reference passage and their actual scores on the reference passage. Compared to the raw WCPM scores, the equated WCPM scores had smaller mean residual sum of squares except for Passage WGK_11001, indicating that the equated scores are much closer to students' actual scores on the reference passage. This suggests that the equating process worked well to equate students' WCPM scores from a non-reference passage to scores from the reference passage.

Table 3.3. Mean Residual Sum of Squares between Raw and Equated WCPM Scores from the Reference Passage and a Non-Reference Passage

Passage	N	Raw WCPM	Equated WCPM
FG3_10008	16,990	189.98	132.26
ORF011_00006	1,306	467.37	160.95
ORF011_00007	1,381	311.97	138.60
ORF011_00008	1,404	217.65	169.18
ORF011_00009	1,225	555.06	168.70
ORF011_00010	1,283	599.44	151.87
ORF011_00011	1,429	176.07	148.53
ORF011_00012	1,486	219.44	180.39
ORF011_00013	1,495	208.11	149.21
ORF011_00014	1,479	188.83	153.39
ORF011_00016	1,462	216.61	175.20
ORF011_00018	1,439	209.13	187.39
ORF011_00019	1,287	496.44	170.90
ORF011_00020	1,420	211.27	146.74
ORF011_00021	1,421	194.66	146.04
ORF011_00022	1,301	550.29	160.62
ORF011_00023	1,471	219.91	140.20
ORF011_00025	1,461	226.33	165.44
ORF011_00026	1,454	271.77	150.32
ORF011_00027	1,414	176.21	145.71
ORF011_00028	1,354	254.98	137.84
WGK_11001	12,082	157.89	161.50

3.3. Validity Evidence

Table 3.4 presents the correlation between the raw and equated WCPM scores and scores from an external reading measure, the NWEA MAP® Growth™ Reading assessment, based on the order of passages presented to students. In each term, a certain proportion of the students took both MAP Reading Fluency and MAP Growth Reading. The WCPM scores measure reading fluency, whereas the RIT scores from MAP Growth Reading measure reading achievement. Only data from students who took both MAP Reading Fluency and MAP Growth in the same term were used in this analysis.

Compared with the raw WCPM scores, equated WCPM scores generally showed higher correlations with students' RIT scores from MAP Growth Reading, with only one exception (Fall 2018, p1). This suggests that equated scores may reflect students' true reading ability better by eliminating fluctuations driven by text difficulty.

Table 3.4. Correlations Between Raw/Equated WCPM Scores and MAP Growth Reading RIT Scores

Passage Position	N	Correlations between MAP Growth Reading	
		Raw WCPM	Equated WCPM
Fall 2018			
p1	22,053	0.596	0.588
p2	19,945	0.556	0.605
p3	18,802	0.529	0.595
Winter 2019			
p1	29,272	0.594	0.602
p2	27,842	0.556	0.608
p3	26,735	0.544	0.576
Spring 2019			
p1	34,354	0.585	0.586
p2	32,756	0.538	0.586
p3	31,026	0.535	0.569

3.4. Progress Monitoring

Students' oral reading fluency is expected to improve from term-to-term across a school year. However, score differences when raw WCPMs are involved are greatly affected by differences in passage difficulty, and the size and variability of these differences in raw scores tend to increase as students progress from the first-to-second and second-to-third passages because of the adaptive nature of MAP Reading Fluency. However, after equating the passages, there should be consistent size and variability in score differences across terms, meaning it should not matter whether a difference score is calculated based on the first, second, or third passage.

Table 3.5 presents descriptive statistics for the differences between the same student's WCPM scores across terms by passage position. Students with valid WCPM scores for the respective passage from both terms were included in each analysis. Overall, the mean and SD of the differences in equated WCPM scores of each of the three passages across two terms were much more consistent than those of the differences in raw WCPM scores. For example, the mean of the differences of the equated WCPM scores between Winter 2019 and Fall 2018 ranged from 11.948 to 12.627 for all three passages, whereas the mean of the differences of the raw WCPM scores ranged from 0.307 to 11.206. The SDs of the differences of the equated scores are also smaller than those of the raw scores.

Table 3.5. Change of Students' WCPM Scores Across Terms by Passage Position

Passage	N	Raw WCPM		Equated WCPM	
		Mean	SD	Mean	SD
Spring 2019 – Winter 2019					
$P1_{s19}$-$P1_{w19}$	33,215	7.19	17.28	6.70	16.84
$P2_{s19}$-$P2_{w19}$	31,591	7.06	18.26	6.36	16.97
$P3_{s19}$-$P3_{w19}$	29,615	8.32	23.07	6.44	17.30
Winter 2019 – Fall 2018					
$P1_{w19}$-$P1_{f18}$	16,882	0.31	19.37	12.63	16.81
$P2_{w19}$-$P2_{f18}$	15,275	9.77	19.52	11.95	17.34
$P3_{w19}$-$P3_{f18}$	14,282	11.21	21.93	12.38	17.78
Spring 2019 – Fall 2018					
$P1_{s19}$-$P1_{f18}$	14,397	7.23	21.31	18.17	19.50
$P2_{s19}$-$P2_{f18}$	13,083	14.99	22.98	17.71	20.16
$P3_{s19}$-$P3_{f18}$	12,180	18.58	23.71	18.52	20.33

3.5. Conclusion

NWEA introduced a new method based on graph theory to select reference passages for equating WCPM scores. This method makes equating a large number of passages practical. Results from this study suggest that, compared to the raw WCPM scores, equated WCPM scores have higher reliability, better reflection of students' true reading ability as indicated by higher correlations with scores from external reading measures, and better monitoring of students' reading fluency progress over time indicated by more consistent changes across passages over time. These results provide evidence of the advantages of equated scores compared to the raw scores. Passage equating can potentially reduce the assessment burden for students and teachers by reducing the number of passages students need to read to get a reliable and valid oral reading fluency score.

References

Albano, A. D. (2016). Equate: An R package for observed-score linking and equating. *Journal of Statistical Software, 74*(8), 1–36.

Ardoin, S. P., Suldo, S. M., Witt, J. C., Aldrich, S., & McDonald, E. (2005). Accuracy of readability estimates' predictions of CBM performance. *School Psychology Quarterly, 20*(1)*,* 1–22.

Burns, M. K., Tucker, J. A., Hauser, A., Thelen, R., Holmes, K., & White, K. (2002). Minimum reading fluency rate necessary for comprehension: A potential criterion for curriculum-based assessments. *Assessment for Effective Intervention, 28*(1)*,* 1–7.

exSTEMsions. (2019, May 2). What is graph theory, and why does it matter so much? [Blog post]. Retrieved from, https://exstemsions.com/blog/graphtheory.

Francis, D. J., Santi, K. L., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. R. (2008). Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology, 46*(3)*,* 315–342.

Fuchs, L. S., Fuchs. D. F., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*(3)*,* 239–256.

Kolen, M. J., & Brennan, R. J. (1995). *Test equating: methods and practices.* New York: Springer-Verlag.

Nathan, R. G., Stanovich, K. E. (1991). The causes and consequences of differences in reading fluency. *Theory into Practice, 30*(3)*,* 76–184.

Stoolmiller, M., Biancarosa, G., & Fien, H. (2012). Measurement properties of DIBELS oral reading fluency in Grade 2: Implications for equating studies. *Assessment for Effective Intervention, 38*(2), 76–90. doi:10.1177/1534508412456729.

U.S. Census Bureau. (2017). *U.S. Census Bureau QuickFacts: Population estimates, July 1, 2017 (V2017).* Retrieved from https://www.census.gov/quickfacts/fact/table/US/PST045217.

Young, A., & Bowers, P. G. (1995). Individual difference and text difficulty determinants of reading fluency and expressiveness. *Journal of Experimental Child Psychology, 60*(3), 428–454.