Non-linear Relations Improve Clinical Interpretations

Aidan Campagnolio and W. Joel Schneider

Psychological Studies in Education, Temple University

Author Note

Aidan Campagnolio https://orcid.org/0009-0003-1210-0877

W. Joel Schneider https://orcid.org/0000-0002-8393-5316

This document may be found at https://twu-ir.tdl.org/home. This research was supported by Texas Woman's University Woodcock Institute Research Grant.

Correspondence concerning this article should be addressed to Aidan Campagnolio, Psychological Studies in Education, Temple University, 1301 Cecil B. Moore Ave., Philadelphia, PA 19122, USA, Email: acamp@temple.edu

Abstract

School psychologists often rely on traditional statistical models to interpret the relationship between academic abilities. Ordinary least squares (OLS) regression and structural equation modeling (SEM) are particularly susceptible to overestimating a predictor's effect at the extremes of the academic ability distribution (e.g., students with specific learning disabilities or giftedness). This research aims to illustrate the importance of using non-linear models when interpreting and making predictions about students' academic abilities; especially students who are evaluated for an SLD or giftedness. Using the WJ V, generalized additive models (GAMs) will allow for the investigation of non-linear relations across all academic abilities. Preliminary analyses of non-linear relationships among basic and complex reading abilities illustrate that linear models significantly overestimate the predictor's effect for students with scores at the lower- and higher-extremes of the distribution (Campagnolio et al., 2025); the proposed research aims to build upon the previous work. By using large standardization samples from popular academic assessments, this research seeks to provide clinicians with a more nuanced understanding of the relationships among academic skills and increasing their ability to support youth across the entire ability spectrum.

Keywords: non-linear relationships, academic achievement, generalized additive model (GAM)

Non-linear Relations Improve Clinical Interpretations

Because there is a single word *reading*, it is deceptively parsimonious to conceptualize reading and its components (e.g., decoding, fluency, comprehension) as correlates of a singular latent construct. We argue that reading is not a single latent variable but is best thought of as a causal network of interrelated skills. This network frames components like basic decoding skills and fluency as significant predictors of complex reading comprehension skills (Hajovsky et al., 2025; Schneider & McGrew, 2018). In such models, it is common to assume that the relations among variables is linear until evidence suggests otherwise. If non-linear relationships among relevant variables are never explored, then school psychologists are in danger of making predictions and inferences about a student's performance that are suboptimal or worse, simply inaccurate.

The same problem exists with other academic abilities because of the methodological limitations of OLS regression or structural equation modeling (SEM). These statistical methods are not inherently faulty, but they only produce one estimated slope for each predictor on the desired outcome (Hajovsky et al., 2020). As a result, the research is limited and can only report on conditional means of an outcome and its predictors. This might seem like an unnecessary nitpick, but school psychologists mostly work with individuals at either side of the distribution's extremes (e.g., students with a specific learning disability, intellectual disability, giftedness, etc.). Most statistical models derive primarily from data with students performing in the average range (i.e., not the students school psychologists tend to serve and evaluate). This research used Generalized Additive Models (GAMs; Hastie & Tibshirani, 1986) which highlight relationships at all points in the ability continuum so that they are relevant to all students.

We sought to investigate whether there are non-linear relationships among multiple predictors in the Woodcock Johnson–Fifth Edition (WJ V; McGrew et al. (2025)). We previously tested this question using pairs of reading abilities: one predictor and one outcome for decoding, fluency, or comprehension. The results of this smaller project informed us that linear models overestimate the outcome variable's score at the lower and higher ends of the distribution. In other words, OLS regression significantly overestimated lower- and higher-performing students' abilities across multiple reading subtests. Because the WJ family of tests provides both age-adjusted standard scores and age-invariant *W* scores, it is possible to investigate three kinds of non-linear relationships: non linear growth rates with age, non-linear relations among scores across the entire spectrum of age and ability, and non-linear relations after adjusting for age. For example, 6-year-old and an 18-year-old both with the same low standard score are similarly behind their same-age peers, but they are not struggling with the same kinds of reading difficulties. For example, 6-year-old and an 18-year-old both with the same low *W* score find the same tasks to be difficult (e.g., sounding out words). It is possible that some non-linear relationships are masked by standard scores because the non-linear curve bends at certain difficulty thresholds rather than at points of relative standing with one's peers.

Previous analyses of this research suggest significantly different predictions of reading abilities when using a GAM compared to a linear model (Campagnolio et al., 2025). A GAM was used to accommodate the hypothesized non linearity among the reading abilities. The GAM illustrated that the strongest slopes were present in the lower end of the distribution, thus providing more accurate interpretations of basic and complex reading skills. This subtle yet important scientific change allows for those at the extremes to be better understood by school psychologists. For example, decoding words and reading words fluently have a significantly different relationship for a student with standard scores in the 60s-70s than a student with average scores (i.e., the previous research found that the slopes on the low end are 50% greater than slopes on the higher end).

Method

This research explored non-linear relationships of subtests within three academic subjects: math, reading, and writing. GAMs were chosen for this research for their capacity to handle large data sets (Wood, 2025) and analyze nonlinear effects in regression models (Hastie & Tibshirani, 1986). The data originated from various academic test batteries' normative samples, meaning this research exclusively used secondary data to inform its analyses. The authors of this paper contacted testing companies for their normative sample, provided the grant funding information and informed them of the purpose of their research.

All three subjects were broken into three components. Math is made up of calculation, fluency, and applied math; reading is made up of decoding, fluency, and comprehension; and writing is made up of spelling, writing fluency, and written expression. Every subtest used in this analysis was grouped by subject and their respective component (e.g., the subtest "Essay Composition" on the WIAT-III was given the 'writing' and 'written expression' designations). Subtests that touch on more than one of these components were given the designation which best describes the task. For example, a decoding fluency subtest requires both decoding and reading fluency skills; however, the task is ultimately a fluency measure (i.e., it measures how quickly a person reads), so it was labeled as such. Last, duplicates, or scores that predict themselves (e.g., math computation on the WIAT-III predicting math computation on the WIAT-III), were removed from the analysis. This is why the calculation predicting calculation quadrant is empty in Figure 2 and Figure 6.

The academic batteries included in this research are the Woodcock Johnson, Fifth Edition (WJ V; McGrew et al., 2025) , Woodcock Johnson, Fourth Edition [WJ IV; Schrank et al. (2014)], Woodcock Johnson, Third Edition (WJ III; Woodcock et al., 2007), Woodcock Johnson-Revised [WJ-R; Woodcock and Johnson (1989)], Woodcock Johnson Psychoeducational Battery (WJ 77; Woodcock & Johnson, 1977), Wechsler Individual Achievement Test, Third Edition (WIAT-III; Wechsler, 2009), Kaufman Test of Educational Achievement, Third Edition (KTEA-3; Kaufman & Kaufman, 2014), Kaufman Test of Educational Achievement, Second Edition (KTEA-II; Kaufman & Kaufman, 2004), Tests of Dyslexia, Comprehensive (TOD-C; Mather et al., 2024), & Tests of Dyslexia, Early (TOD-E; Mather et al., 2024).

GAMs were deployed to detect non-linearity among academic abilities by using more basic abilities to predict more complex abilities (e.g., decoding → reading fluency → reading comprehension). The same approach was used to analyze non-linearity among the other academic abilities. For example, basic writing abilities predicted more complex abilities (e.g., spelling → sentence fluency → essay composition) and basic math abilities predicted more complex abilities (e.g., math concepts → math fluency → math applications). These relations were observed across the academic distribution at one time point (i.e., cross-sectional analysis).

Further, the analysis was two-fold: Standard scores and *W* scores were used in separate GAMs to investigate whether a) non-linearity existed in either type of score distribution and b) the pattern of non-linearity, if present, was similar across score types. Although there was not a formal hypothesis concerning the potential differences between the non-linearity, it was assumed that *W* scores would provide an alternative picture of these relationships compared to standard scores due to their inherent inclusion of one's developmental age and ability. In other words, standard scores explicitly inform one's abilities in relation to similar-aged peers whereas *W* scores can be analyzed across all ages.

The initial exploratory analyses of the *W* scores displayed non-linear relations that were inexplicably non-linear. Upon further investigation, each subtest had its own mean and standard deviation (SD), which explained the exaggerated non-linear distributions in the initial analyses. Standard scores, on the other hand, were much less complex: they share the same mean (100) and SD (15). Therefore, the *W* scores underwent *z*-standardization by each predictor-outcome pairing in order to reduce unnecessary influences on the academic relationships. Since the analyses compared subtests within a given academic battery, the *W* scores were *z*-standardized according to the pairing and battery (e.g., Reading Fluency predicting Reading Comprehension on the WJ-V).

The following include the primary hypotheses for this research:

1. There will be non-linear relations among all academic domains (e.g., math, reading, and writing).
2. The rate of growth (i.e., the slope) will be stronger at the lower end of the academic distribution compared to the higher end.
3. If these hypothesized results are present in the WJ V, are they generalizable to other academic batteries (e.g., KTEA-3, TOD)?

Participants

This research exclusively used archival data, which originated from normative samples for each academic battery. The sample characteristics are described in their respective manuals, but in each case, the normative samples were intended to match the population characteristics of the United States population at the time of standardization. Sample sizes can be found in Table 1.

Table 1

Sample sizes for each data set

Battery	*n*
WJ V	5,837
WJ IV	7,416
WJ III	8,782
WIAT-III (WISC-IV)[a]	117
WIAT-III (WISC-V)[b]	211
KTEA-3	2,050
KTEA-II (KABC-2 NU)[c]	2,520
KTEA-3 (KABC-2 NU)[c]	99
WJ 77	3,957
WJ-R	6,359
TOD-C	1,748
TOD-E	342
Total	39,438

[a]The data from this row originate from a linked sample with the WISC-IV

[b]The data from this row originate from a linked sample with the WISC-V

[c]The data from this row originate from a linked sample with the KABC-2 NU

Measures

The test classifications for each test battery can be seen in Table 2.

Table 2

Test Classifications

Test	Battery
Calculation	
Calculation	WJ77, WJ R, WJ III, WJ IV, WJ 5
Math Computation	KTEA-II, KTEA-III
Numerical Operations	WIAT-III
Math Fluency	
Magnitude Comparison	WJ 5
Math Facts Fluency	WJ IV, WJ 5
Math Fluency	KTEA-III, WJ III
Math Fluency–Addition	WIAT-III
Math Fluency–Multiplication	WIAT-III
Math Fluency–Subtraction	WIAT-III
Applied Math	
Applied Problem Solving	WJ77
Applied Problems	WJ R, WJ III, WJ IV, WJ 5
Math Concepts and Applications	KTEA-II, KTEA-III
Math Problem Identification	WJ 5
Math Problem Solving	WIAT-III
Decoding	
Blending	TOD-C
Decoding Fluency	KTEA-III
Early Reading Skills	WIAT-III
Early segmenting	TOD-E
Irregular Word Reading	TOD-C
Letter and Word Recognition	KTEA-II, KTEA-III
Letter-Word Identification	WJ77, WJ R, WJ III, WJ IV, WJ 5
Nonsense Word Decoding	KTEA-II, KTEA-III
Nonsense Word Reading	TOD-C
Oral Reading	WJ IV, WJ 5
Oral Reading Accuracy	WIAT-III
Phonological Awareness	KTEA-II

Test	Battery
Phonological Awrareness–Short Set	KTEA-II
Phonological Processing	KTEA-III
Pseudoword Decoding	WIAT-III
Segmenting	TOD-C
Sounds and Pseudoword	TOD-E
Word Attack	WJ R, WJ III, WJ IV, WJ 5
Word Reading	WIAT-III
Reading Fluency	
Decoding Fluency	KTEA-II, KTEA-III
Oral Reading Fluency	WIAT-III, TOD-C
Oral Reading Rate	WIAT-III
Oral Word Fluency	WIAT-III
Passage Reading Fluency	TOD-C
Question Reading Fluency	TOD-C, TOD-E
Rapid Irregular Word Reading	TOD-C
Rapid Nonsense Word Reading	TOD-C
Reading Fluency	WJ III
Sentence Reading Fluency	WJ IV
Silent Reading Fluency	KTEA-III
Word Reading Fluency	WJ IV, WJ 5, TOD-C, TOD-E
Word Recognition Fluency	KTEA-III
Letter and Sight Word Recognition	TOD-E
Letter and Sound Knowledge	TOD-E
Paragraph Reading Comprehension	WJ 5
Passage Comprehension	WJ R, WJ III, WJ IV, WJ 5
Reading Comprehension	KTEA-II, KTEA-III, WIAT-III
Reading Recall	WJ 5
Reading Vocabulary	KTEA-III, WJ R, WJ IV
Spelling	
Irregular Word Spelling	TOD-C
Regular Word Spelling	TOD-C
Spelling	KTEA-II, KTEA-III, WIAT-III, WJ77, WJ R, WJ III, WJ IV, WJ 5

Test	Battery
Spelling of Sounds	WJ III, WJ IV, WJ 5
Writing Fluency	
Alphabet Writing Fluency	WIAT-III
Letter Writing Fluency	WJ 5
Sentence Writing Fluency	WJ IV, WJ 5
Writing Fluency	KTEA-III, WJ R, WJ III
Written Expression	
Essay Composition	WIAT-III
Sentence Building	WIAT-III
Sentence Combining	WIAT-III
Sentence Composition	WIAT-III
Sentence Writing Accuracy	WJ 5
Theme Development and Text Organization	WIAT-III
Writing Samples	WJ III, WJ IV
Written Expression	KTEA-II, KTEA-III, WIAT-III, WJ 5

Procedure

The first author reached out to each test publishing company to gain access to the normative sample data. The data were then organized, cleaned, and analyzed in R 4.5.2 (R Core Team, 2025b) via RStudio 2026.01 (Posit team, 2025), prioritizing functions from tidyverse collection of packages (Wickham et al., 2019b). The comprehensive list of packages used in our analyses can be found in Table 3.

Table 3

R packages used to create this document.

Package	Version	Citation
apa7	0.1.0	Schneider (2025a)
base	4.5.2	R Core Team (2025)
flextable	0.9.11	Gohel and Skintzos (2026)
foreign	0.8.91	R Core Team (2026)
ftExtra	0.6.4	Yasumoto (2024)
ggdiagram	0.1.1	Schneider (2025b)
ggtext	0.1.2	Wilke and Wiernik (2022)
janitor	2.2.1	Firke (2024)
lavaan	0.6.21	Rosseel (2012); Rosseel, Jorgensen, and De Wilde (2025)
lme4	1.1.38	Bates et al. (2015)
Matrix	1.7.4	Bates, Maechler, and Jagan (2025)
mgcv	1.9.4	Wood (2003); Wood (2004); Wood (2011); Wood, Pya, and Säfken (2016); Wood (2017)
modelbased	0.14.0	Makowski et al. (2025)
nlme	3.1.168	J. C. Pinheiro and Bates (2000); J. Pinheiro, Bates, and R Core Team (2025)
psych	2.6.1	William Revelle (2026)
tidyverse	2.0.0	Wickham et al. (2019)
WJSmisc	0.3	Schneider (2021)
writexl	1.5.4	Ooms (2025)

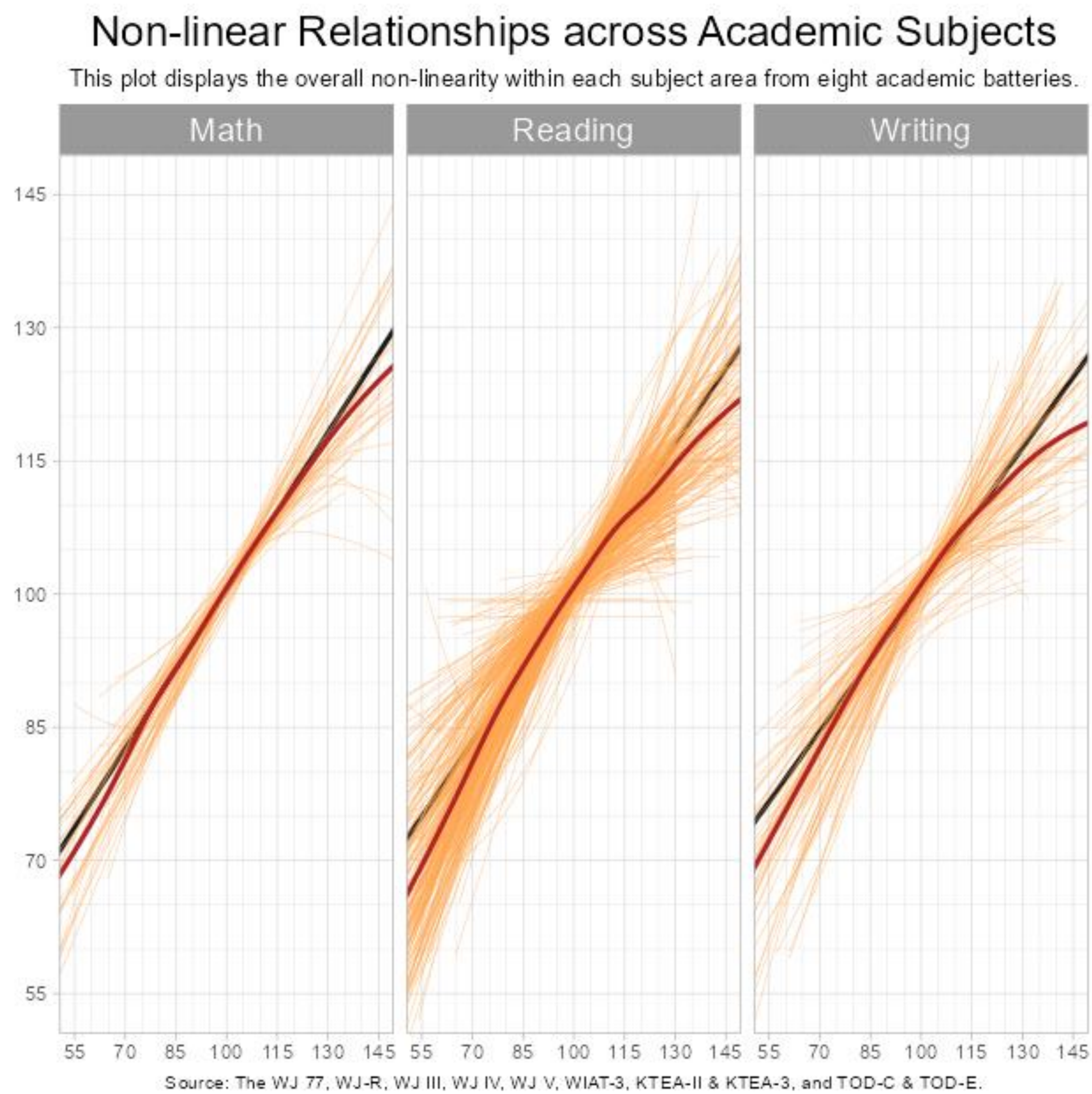
Results

Standard Scores

The first hypothesis stated there will be non-linear relations among all subjects, and although that is technically true (i.e., there is at least one instance of non-linearity in the three subjects) it does not convey the whole truth. Reading was evidently non-linear at the lower- and higher-ends, once again, and writing was similarly non-linear. However, math was quite linear, overall.

Figure 1

Within-Subject Relationships of Math, Reading, and Writing (Standard scores)

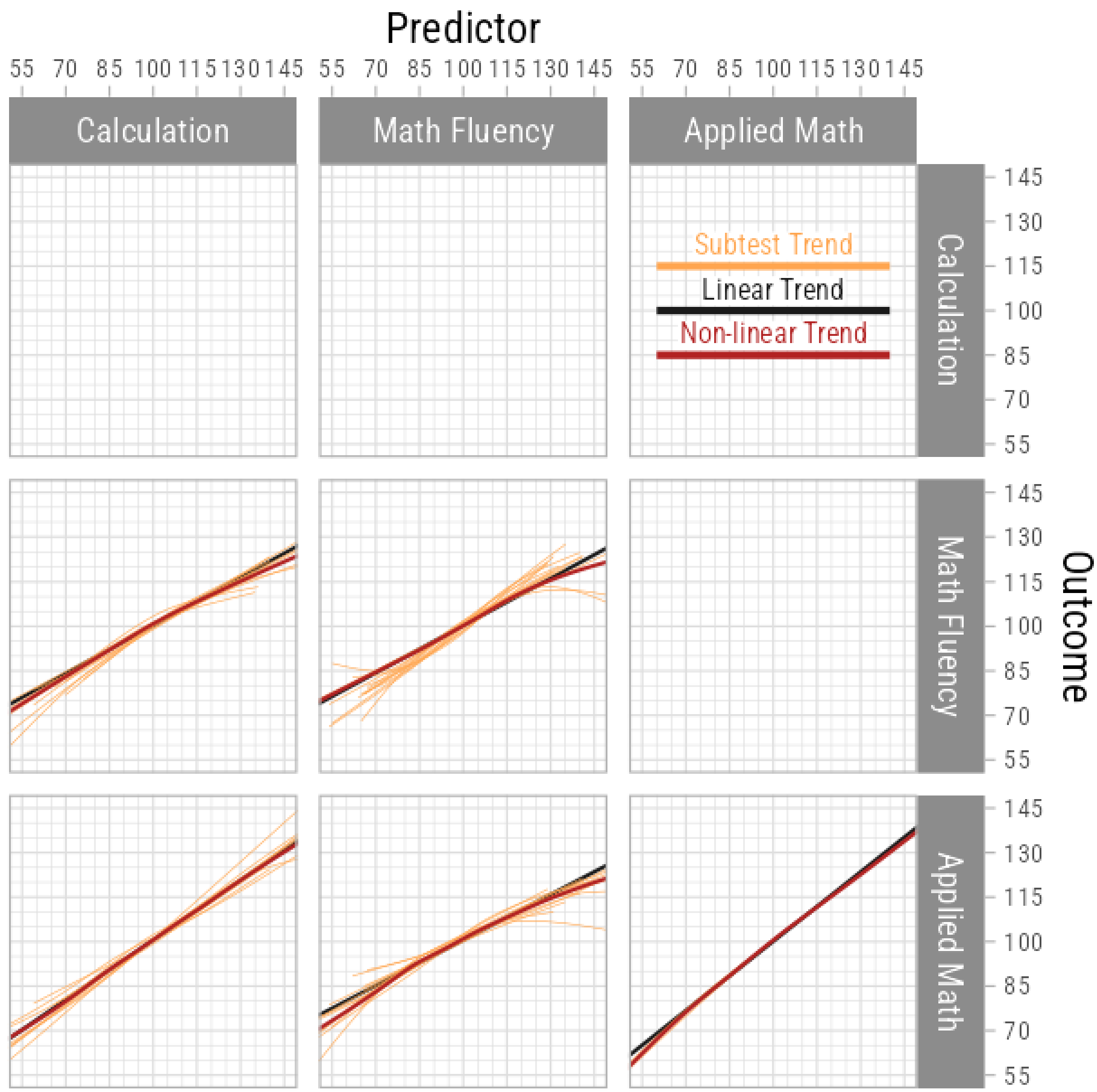


Math

Starting with the most linear domain, math calculation predicting itself, math fluency predicting itself, and applied math predicting itself were all linear. Calculation predicting applied math was also linear, indicating the two's relationship is appropriately described with a linear model. However, there were two instances of non-linear relations in math. The first instance appeared when calculation predicted math fluency. Although minimal, there were apparent discrepancies between the linear model and the GAM: individuals at the extremes of the math achievement distribution (greater than 2 SD from the mean) were predicted to have higher scores with a linear model compared to a GAM. In other words, students who exhibit considerable challenges with math and those who are extremely talented in math are actually performing worse than what is shown with a linear model. A similar pattern occurred when math fluency predicted applied math. This time, the discrepancy grew larger: students at both extremes of the distribution had their abilities overestimated with a linear model. See Figure 2 for a visualization of the results.

Figure 2

Non-linear Relationships among Math Skills (Standard scores)

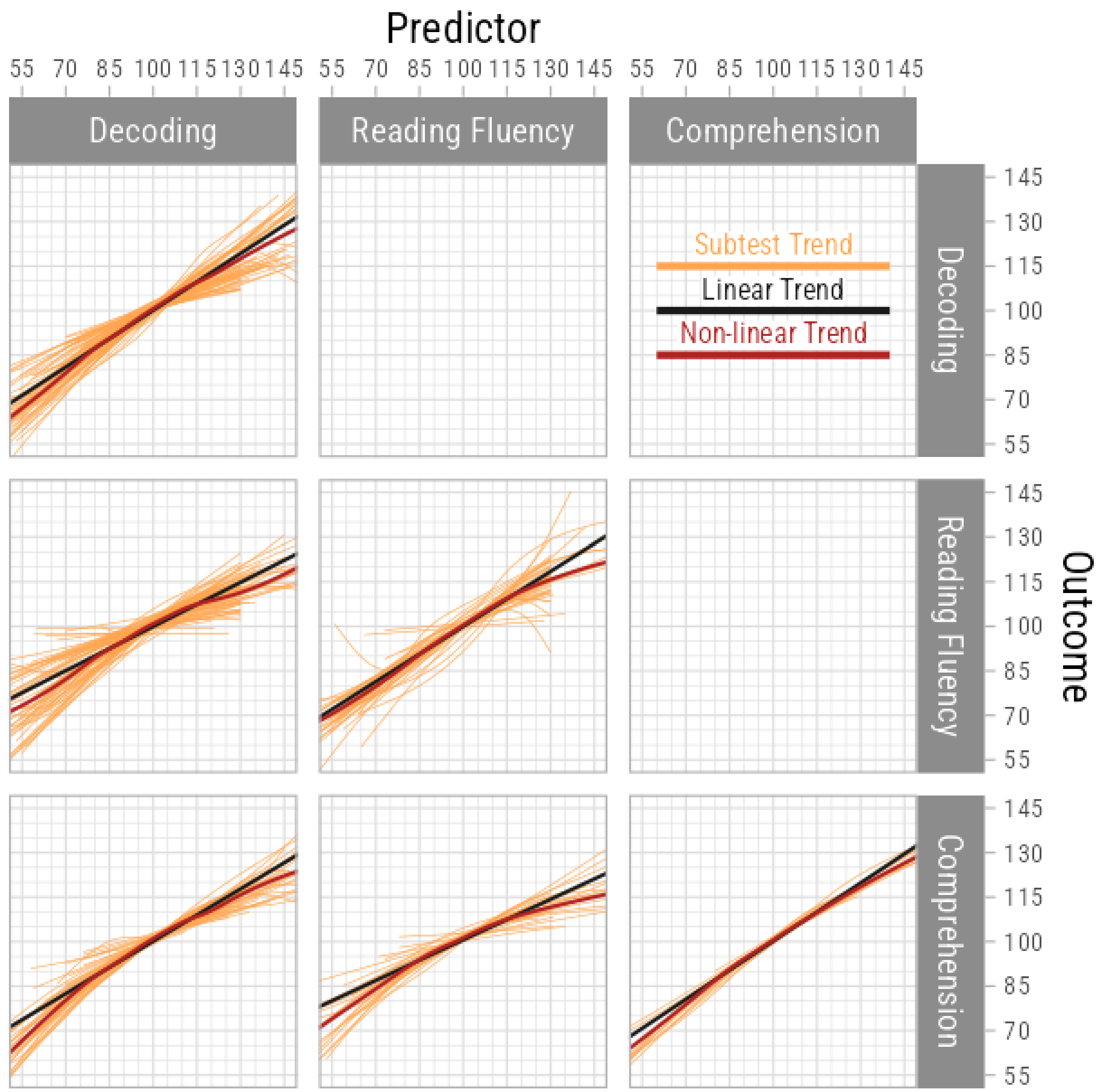


Reading

Reading was the most non-linear academic subject. Similar to math, reading contained patterns where the GAM was considerably different than the linear model at both ends of the achievement distribution (see decoding predicting comprehension and reading fluency predicting comprehension in Figure 3). A new pattern of non-linearity occurred in reading, specifically when decoding predicted reading fluency. While the other non-linear patterns looked like a concave lens, this pair has two u-like curves at both extremes and a linear regression around the first standard deviation. This new pattern highlights that not all non-linear predictor-outcome pairs display the same manner of non-linearity.

Figure 3

Non-linear Relationships among Reading Skills (Standard scores)



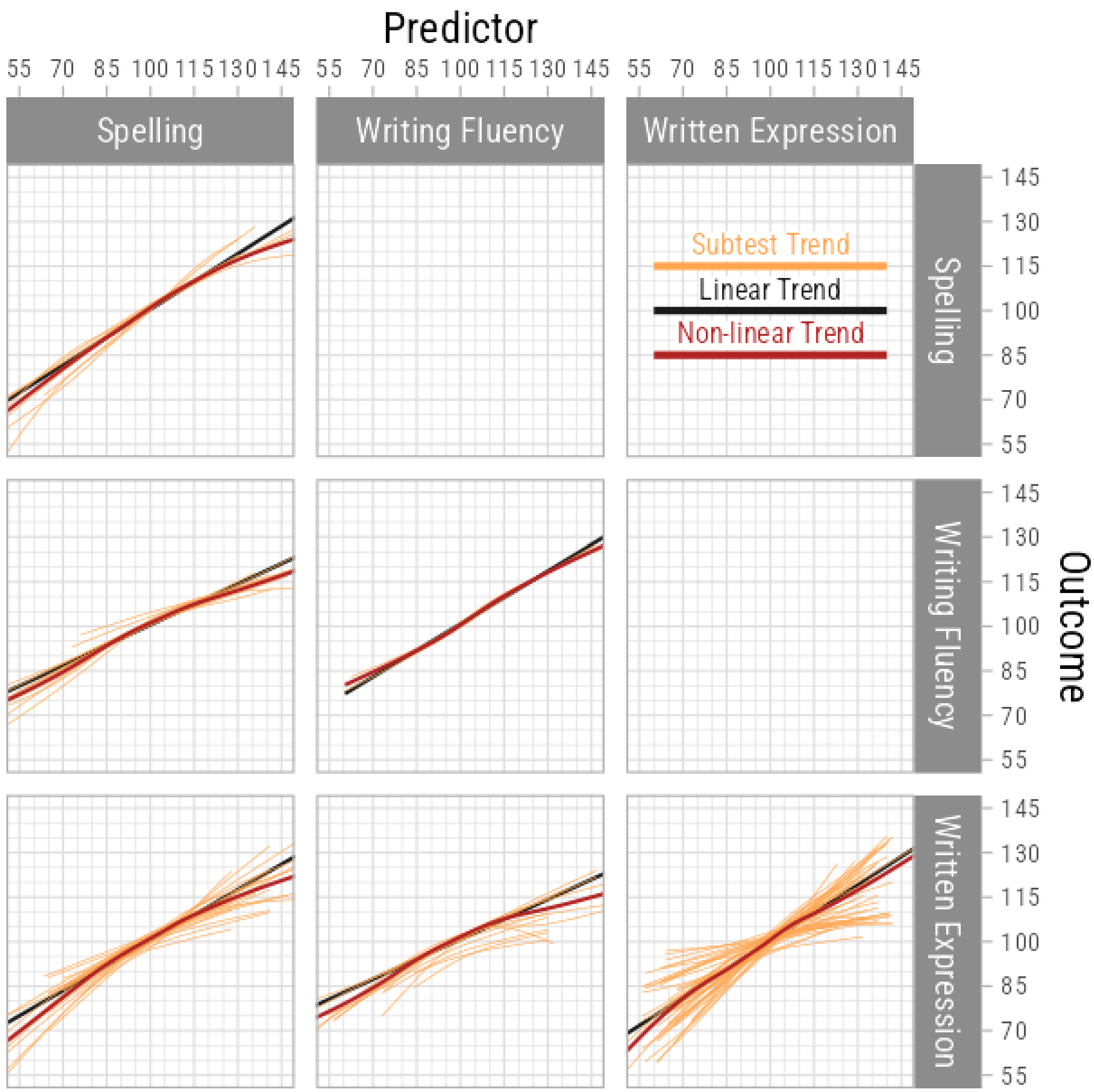
Writing

Writing subtest pairs also displayed non-linear regression trends. Spelling predicting writing fluency resulted in similar trajectory that decoding predicting reading fluency contained.

Spelling predicting written expression was the most discrepant pair among all writing subtests. Again, those at both ends of the achievement distribution were best captured using a GAM, as their abilities were overestimated using a linear model. Writing fluency predicting written expression contained the u-like curves which were previously mentioned. This finding revealed that multiple academic subjects contain multiple types of non-linear relationships (see in Figure 4).

Figure 4

Non-linear Relationships among Writing Skills (Standard scores)

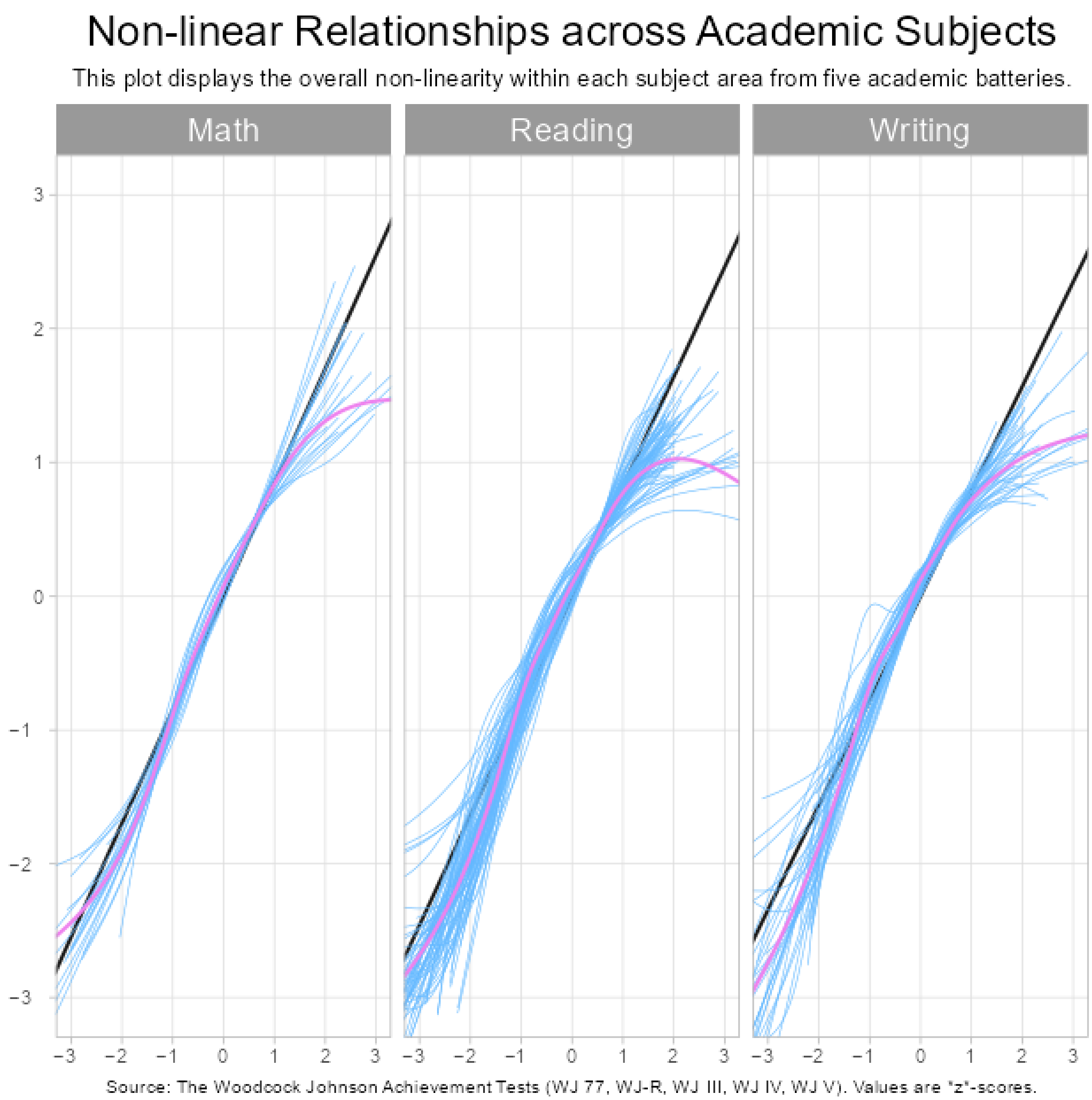


W Scores

The *z*-standardized *W* score plots displayed new non-linear patterns and revealed more instances of non-linearity. Math, reading, and writing all contained instances of non-linearity, however these scores mostly contained non-linear curves at the higher half of the achievement distribution (see Figure 5). The exception to this rule occurred in reading, where a curvilinear line stretched from the negative third standard deviation (−3 SD) and flattens out after the negative first standard deviation (−1 SD).

Figure 5

Within-Subject Relationships of Math, Reading, and Writing (W scores)



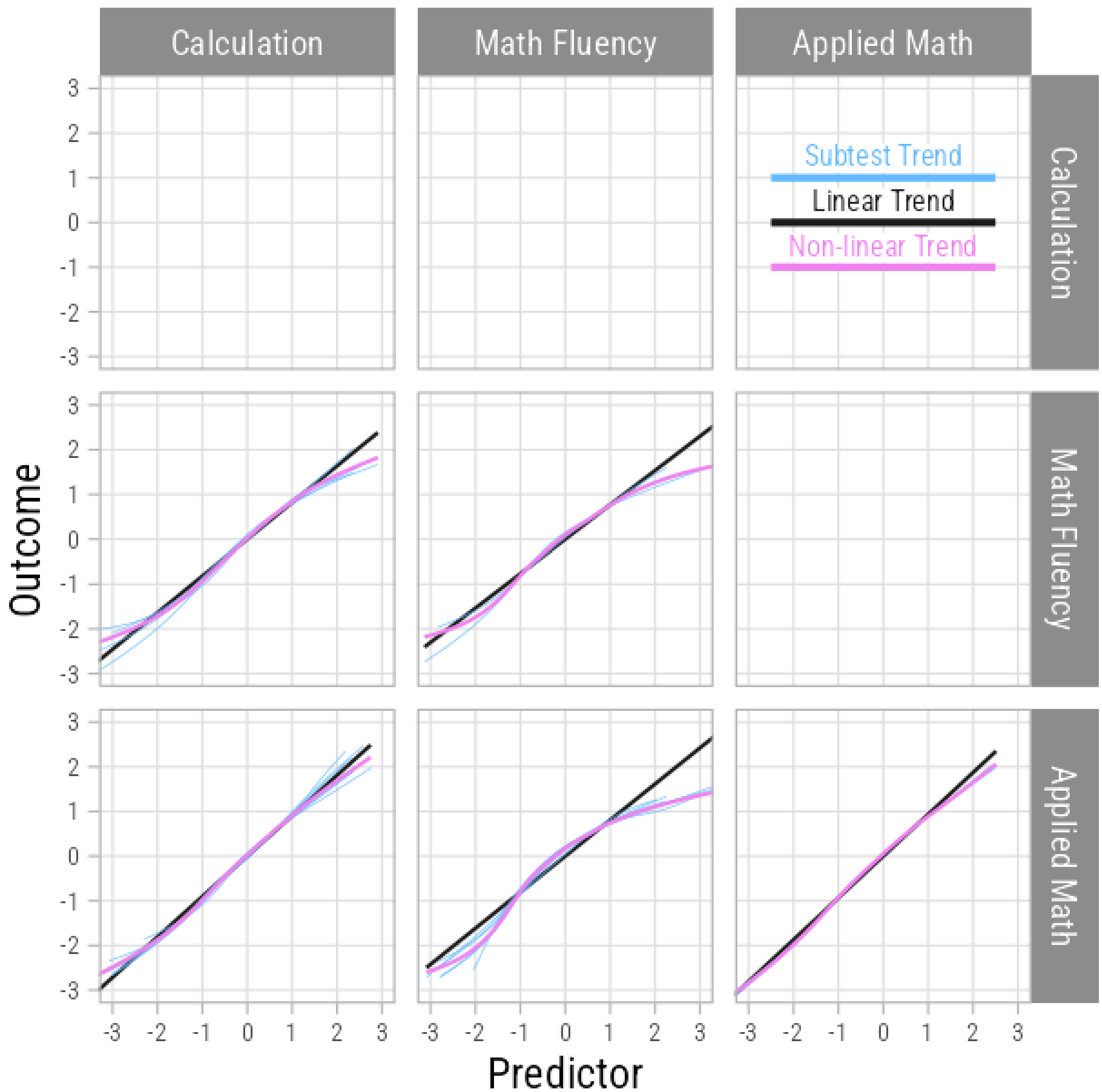
Math

When calculation predicted math fluency, individuals' abilities on the low end were underestimated with a linear model. Meaning, their scores were estimated to be greater with a

GAM as opposed to the linear model. Individuals' scores were underestimated at the higher end of the distribution, also. The most interesting plot was math fluency predicting applied math. The GAM crossed above and below the linear trend at the −1 SD and the +1 SD region. Both low and high performers' abilities were overestimated with a linear model. Also, average performers' abilities were partially underestimated with a linear model (see in Figure 6).

Figure 6

Non-linear Relationships among Math Skills (W scores)



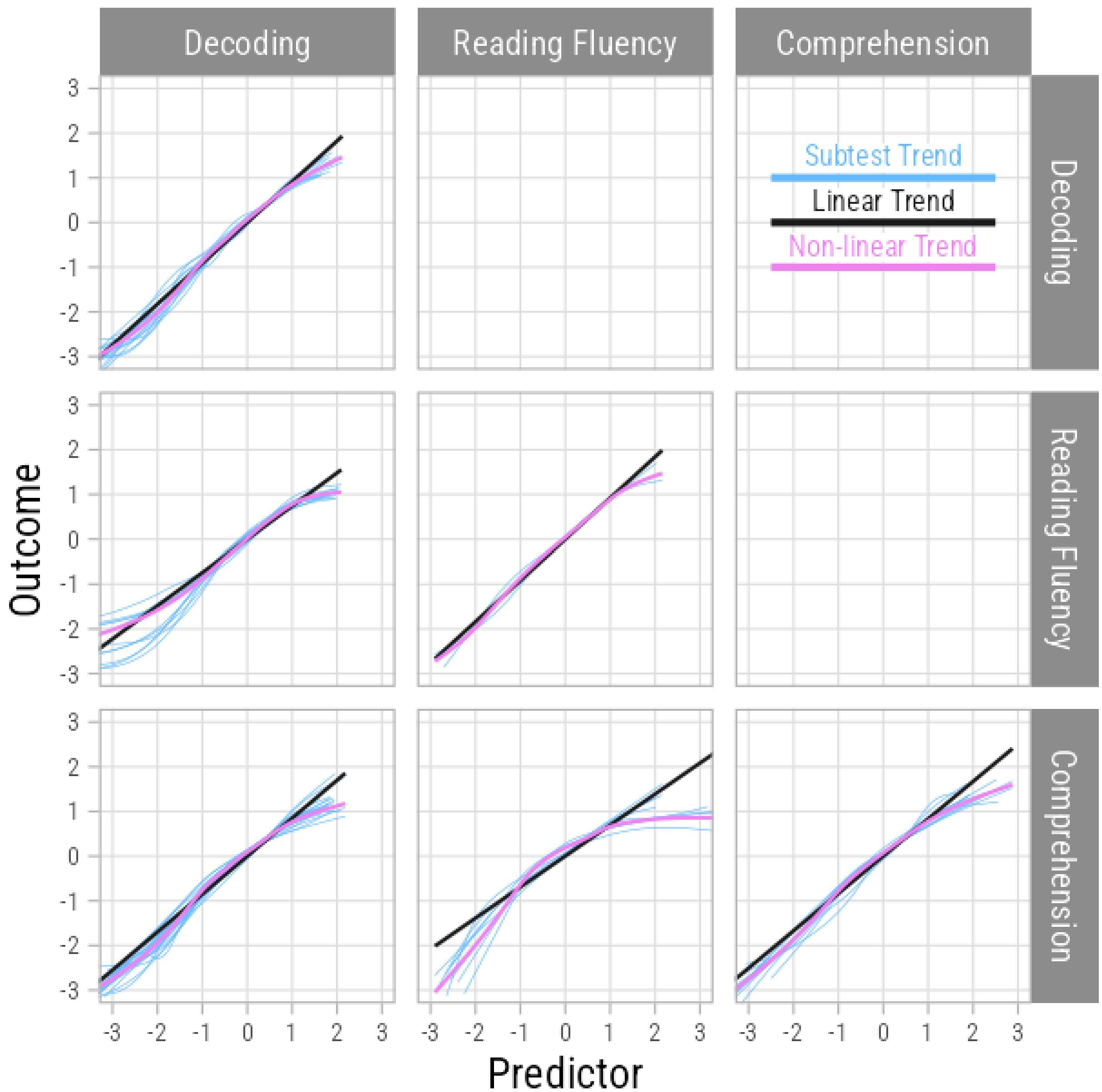
Reading

Another fascinating instance of non-linearity occurred when reading fluency predicted reading comprehension. A linear model significantly overestimated people's scores on both ends of the distribution, while underestimating average readers' abilities. A linear trend drastically overestimates readers' abilities on the low end of the spectrum, specifically between the negative third and first standard deviations. On the other end, the GAM revealed that reading fluency seems to peak in terms of its relationship with reading comprehension at the first standard deviation. After this peak, the abilities almost appeared unrelated. This story contrasts significantly with the linear model, which portrays reading fluency as a significant predictor of reading comprehension for skilled readers.

There were additional non-linear trends in other reading subtest pairs. Decoding predicting reading fluency and decoding predicting reading comprehension showed that the relationship trails off past starting at the first standard deviation. When decoding predicted comprehension, readers at the high end of the distribution had their abilities significantly overestimated with a linear model, according to the GAM (see Figure 7). Interestingly, reading comprehension predicting reading comprehension contained a discrepancy in the estimates of high scoring individuals' abilities, where the linear model overestimates people's abilities.

Figure 7

Non-linear Relationships among Reading Skills (W scores)

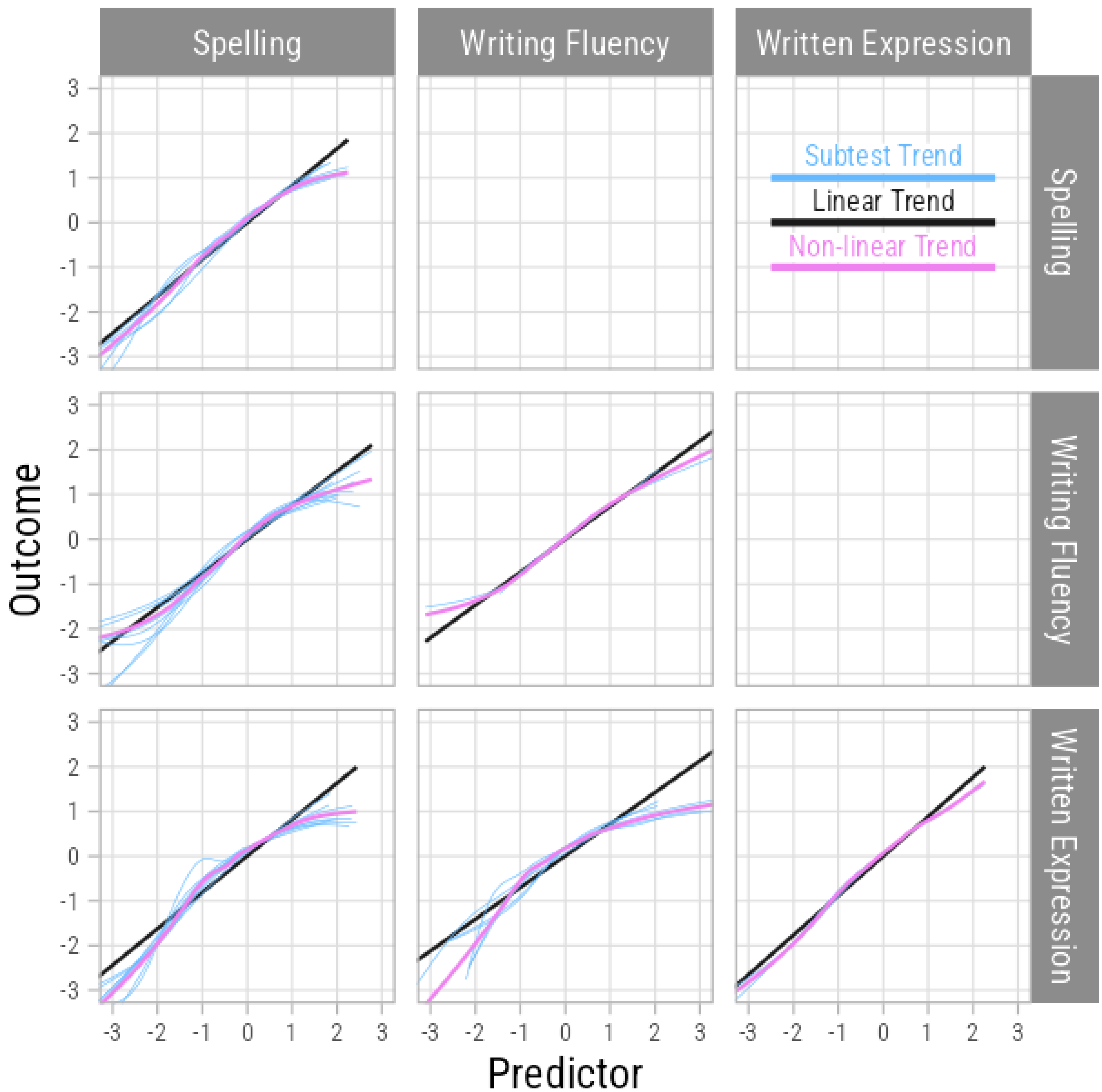


Writing

Spelling as a predictor of writing fluency appeared to adopt the same look as a cubic function (see Figure 8). At the low end, the GAM highlighted how lower performers' abilities crossed above and below the linear model twice. At the high end, the linear model considerably overestimated spelling's relationship with writing fluency. A similar overestimation occurred when spelling predicted written expression: individuals who scored at or greater than one standard deviation above the mean had their abilities significantly overestimated by a linear model. The same overestimation occurred on the low end, too. Also, average writers' abilities were slightly underestimated by a linear model.

Figure 8

Non-linear Relationships among Writing Skills (W scores)



Slopes

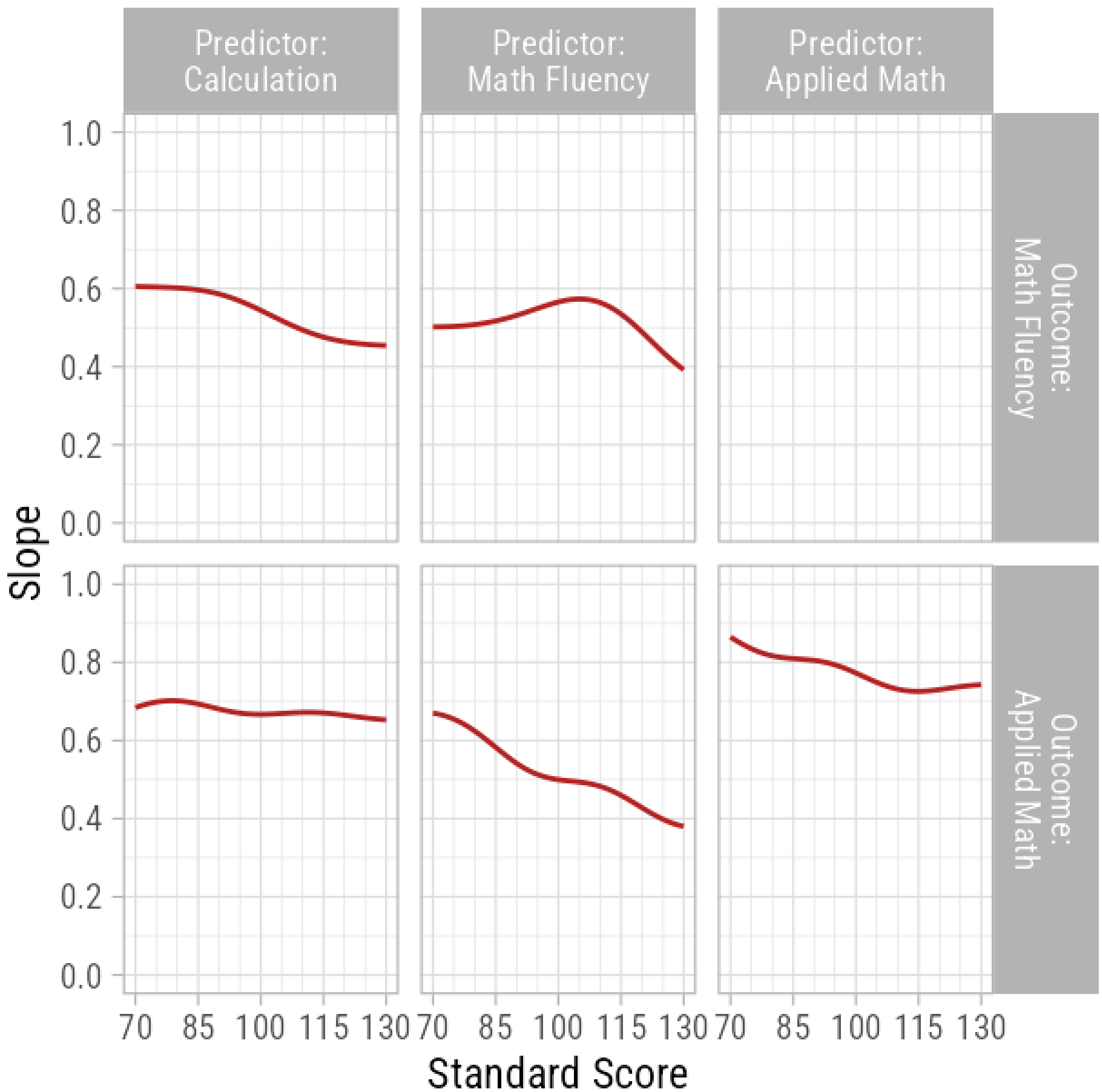
Standard Scores

Math

The variation of slopes within math subtest pairings were unique. Calculation predicting applied math portrayed a simple picture: the two abilities are strongly correlated across the entire achievement spectrum. On the other hand, math fluency predicting applied math highlighted a strong relationship at the low end (SS = 70) with a slope slightly below 0.7 (r = .68) but a weaker relationship at the high end (SS = 130) with a slope slightly below 0.4 (r = .38). The other pairings contained less-variable slopes: Calculation predicting math fluency slope hovered between 0.6 at the low end and 0.45 at the high end; math fluency predicting math fluency slope was mostly consistent throughout the academic distribution (r = .5; r = .39); and applied math predicting applied math slope started strong on the low end (r = .86) and slightly decreased as it approached the high end (r = .74). As a reminder, calculation predicting calculation was not included as it did not contain sufficient unique pairings for the analyses.

Figure 9

Slopes of Math Predictor-Outcome Pairs (Standard scores)

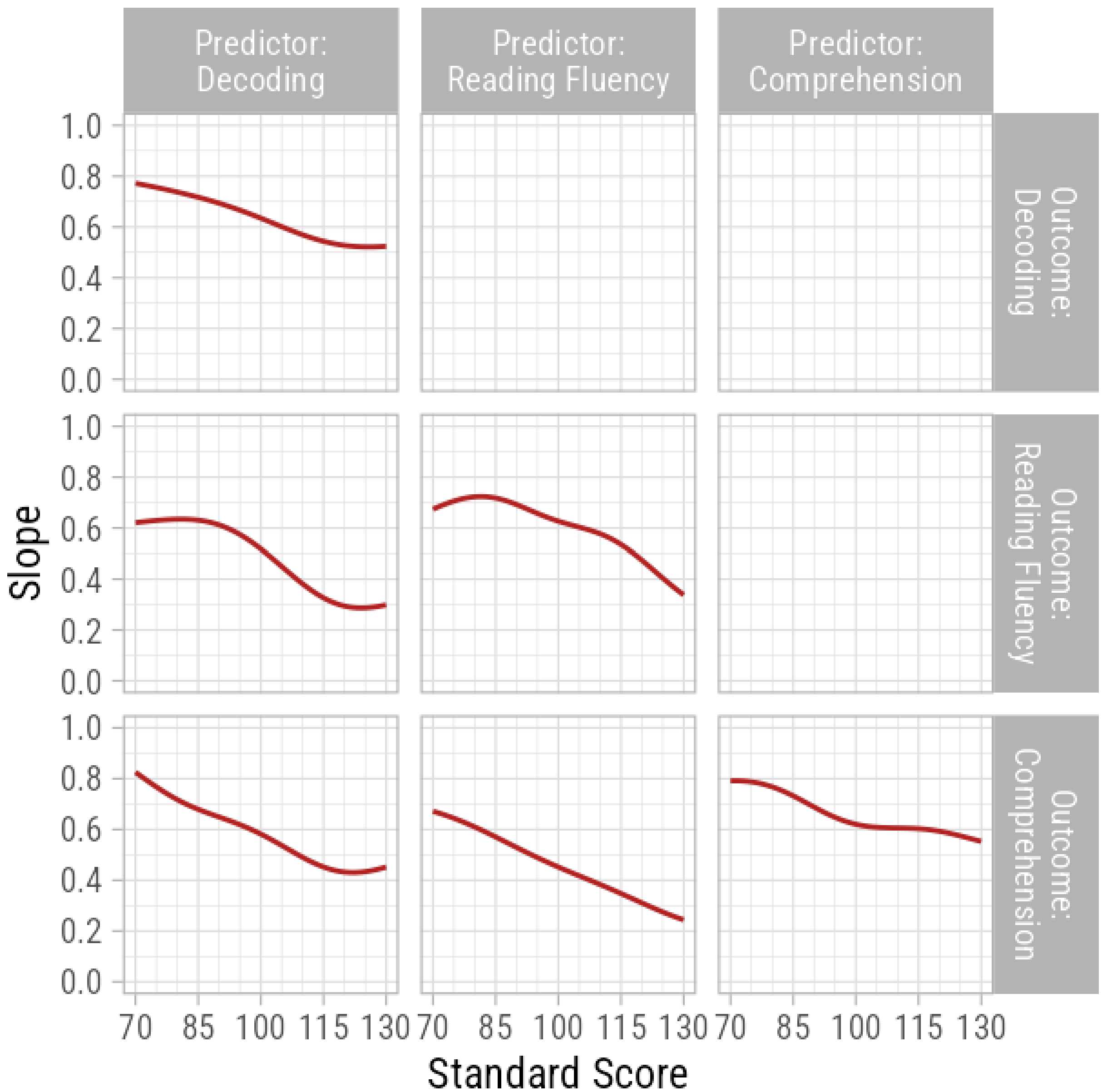


Reading

Reading was the most non-linear subject in our previous work (Campagnolio et al., 2025), and it remained true here. There are various subtest pairings which highlighted the rate of skill development being stronger at the low end and significantly weakening at the high end. For example, when decoding predicted reading fluency, the relationship started strong at 70 SS (r = .62) then halved at 130 SS (r = .3). Reading fluency predicting itself and predicting reading comprehension revealed interesting patterns in the slope over the achievement distribution. First, reading fluency predicting itself displayed a strong slope (r = .72) at 80 SS , yet plummeted at 130 SS (r = .34). Second, reading fluency predicting comprehension displayed a rather linear decrease in slope across the achievement distribution. It peaked at the lower end (r = .67 at 70 SS) and steadily decreased (r = .24 at 130 SS). Decoding predicting comprehension included a similar trend, where the slope was strong at the low end (r = .83 at 70 SS) but drastically dropped at the positive second standard deviation (r = .45).

Figure 10

Slopes of Reading Predictor-Outcome Pairs (Standard scores)



Writing

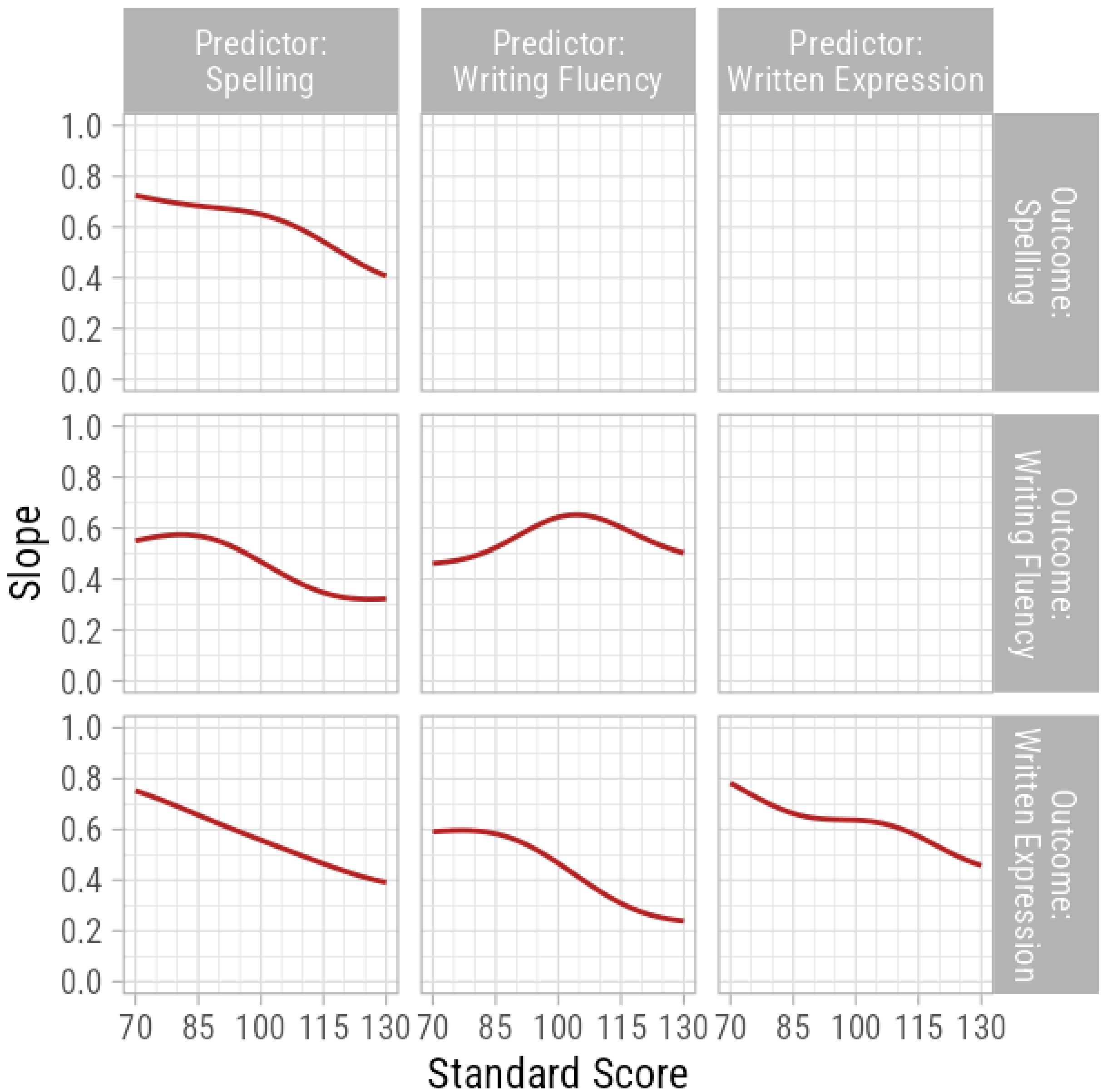
Typically, when one construct predicts itself, the relationship is strong (i.e., anxiety is known to be related to anxiety symptoms; depression is correlated to feelings of sadness).

However, spelling and written expression displayed uniquely different relationships than expected when they predicted themselves. Spelling had a slope of .72 at the low end (70 SS) which then fell to .41 at the high end (130 SS). A similar decline was observed in written expression: at the low end the relationship was strong (r = .78) but weaker at the high end (r = .46). The last within-subtest pairings' slope (writing fluency predicting writing fluency) was unlike the others; the slope mirrored that of a normal curve. It peaked around the mean (r = .65) and had mostly flat tails at both ends of the achievement distribution (r = .46 at 70 SS; r = .5 at 130 SS).

The other subtest pairings' (ones that did not predict themselves) slopes decreased as the abilities increased. Spelling initially was a strong correlate of written expression at the low end (r = .75), but the slope consistently weakened until bottoming at the high end (r = .39). Writing fluency was a surprisingly weak predictor of written expression, especially at the high end. Writing fluency boasted a respectable correlation coefficient of .59 at the low end, which deflated to a significantly weaker correlation coefficient of .24 at the high end. Last, spelling predicting writing fluency resulted in a moderately strong slope. The two were correlated most strongly between the negative second and negative first standard deviation (r = .55), but the slope decreased to its lowest just after the first positive standard deviation and plateaued (r = .32).

Figure 11

Slopes of Writing Predictor-Outcome Pairs (Standard scores)



W Scores

Unlike their SS counterparts, the academic subjects in *W* scores display a steep increase in skill development at the lower end at the academic distribution, then the slope flattens significantly. In the reading plot, there is a considerable reduction in the slope starting at the second deviation on the positive end. The writing plot boasts a similar idea: a steep increase at the lower end of the academic distribution and a considerable reduction in the slope at the higher end (although, less severe than the reduction seen in reading). The math scores in the *W* plot are mostly linear, barring a significant reduction in the slope at the higher end of the achievement spectrum. Overall, *W* scores argue that all academic subjects tend to become less related to one another (e.g., reading correlated with reading) once the subjects become more complex (i.e., when readers are advanced, there seems to be less of a relationship between decoding, fluency, and comprehension). It should be noted that the slopes for the *W* scores are often greater than the SS slopes, since there are considerably more non-linear relationships in the *W* score subjects.

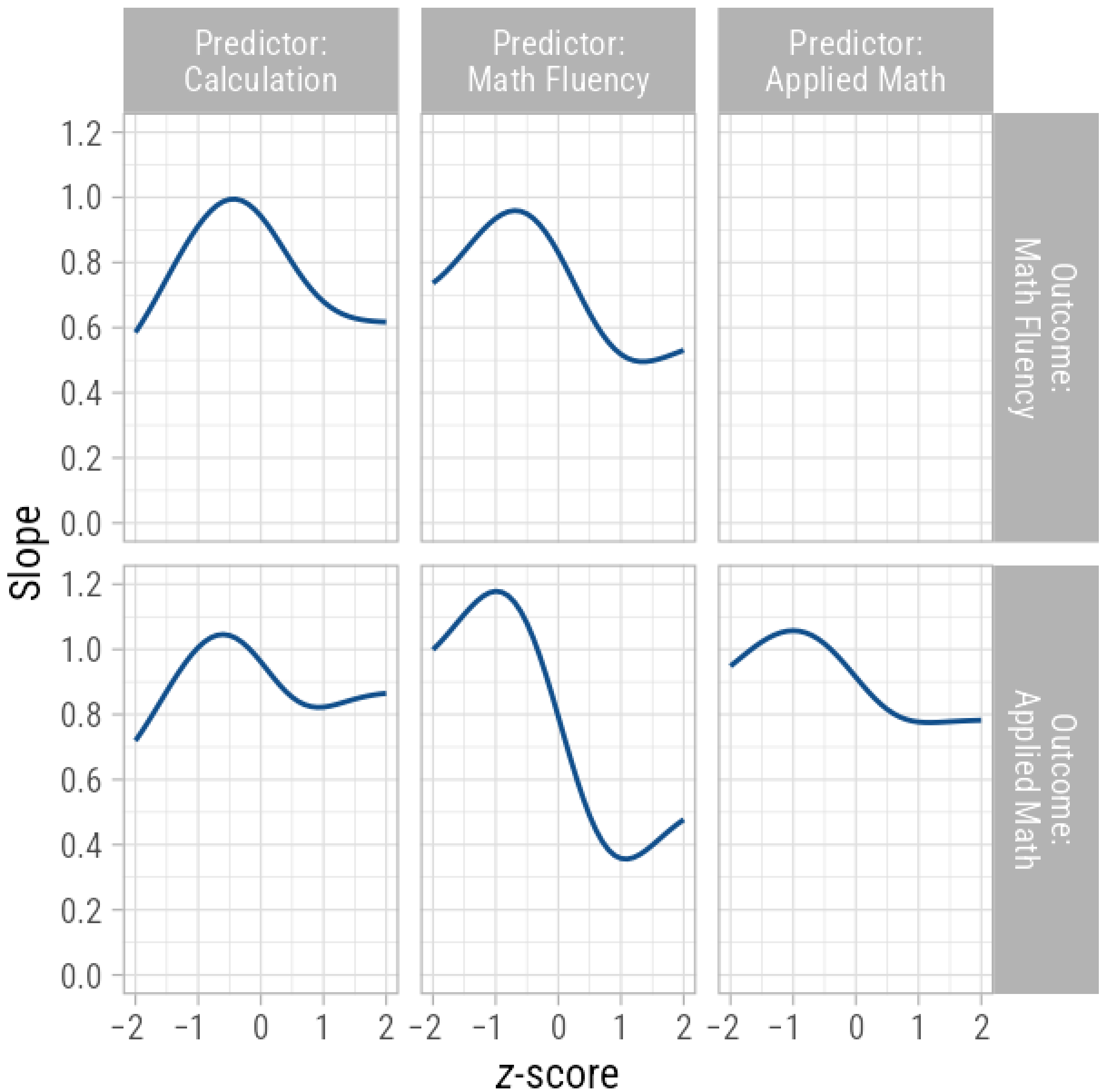
Math

The slopes of math subtests predicting each other mostly follow a similar pattern: they peak between the negative first standard deviation and the mean, and drop as they approach the positive first standard deviation. In other words, math abilities tend to be most related to one another below the mean. The reduction or drop in slopes was quite drastic, like when math fluency predicted applied math, or relatively small, like when calculation predicted applied math. First, math fluency had a correlation coefficient of 1.18 at the negative first standard deviation which then fell to 0.36 at the positive first standard deviation. Second, calculation had a slope of 1 at the negative first standard deviation and a slope of 0.82 at the positive first standard deviation.

Math fluency predicting itself revealed a clear non-linear relationship. The relationship began strong at −2 SD (r = .74), peaked at −0.7 SD (r = .96), dropped significantly at 1.5 SD (r = .50), then stabilized at the 2 SD (r = .53). When calculation predicted math fluency, the slope looked like a normal distribution: low and high-scoring (±2 SD) individuals' slopes were similar at 0.60, yet slightly below-average (−0.5 SD) scoring individuals' slope peaked at 1.0. When applied math predicted itself, the slope peaked at −1 SD (r = 1.05) and leveled off past the first positive SD (r = .78).

Figure 12

Slopes of Math Predictor-Outcome Pairs (z-standardized W scores)



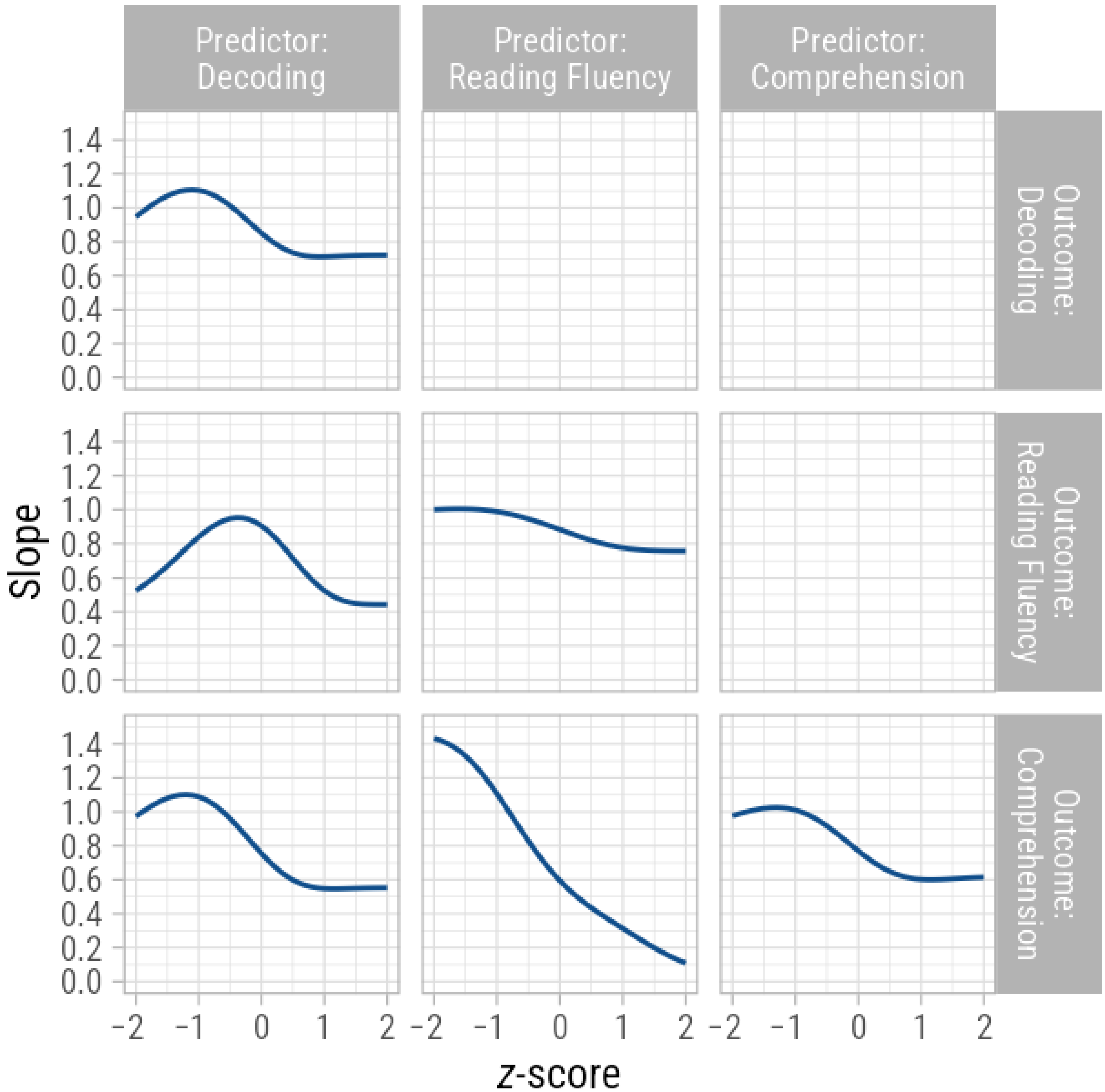
Reading

Overall, reading contained many non-linear slopes. Starting with the diagonal, decoding predicting itself had a strong relationship at the low end (−2 SD; $r = .95$) which peaked just before the −1 SD (r = 1.10) point, then fell at 0.5 SD (r = .74). Reading fluency predicting itself maintained the most linear slope out of all reading pairs, although not perfectly linear. Its slope was 1.0 at the lowest point of the academic distribution and reduced to .76 past the 1st SD. Last, reading comprehension predicting itself revealed a non-linear slope. The slope was steep (r = .98) on the low end (−2 SD), but reduced significantly at the 1st SD (r = .60) and plateaued from there on.

The most drastic reduction in the relationship between two reading abilities occurred when reading fluency predicted reading comprehension. The relationship was strongest at the low end (−2 SD; r = 1.4) and weakest at the high end (2 SD; r = .11); the relationship was moderately strong at the mean (r = .59). Decoding as a predictor revealed more patterns of the slope reducing greatly past the mean. For example, when decoding predicted reading fluency, the slope was moderately strong on the low end (−2 SD; r = .52) and strongest just before the mean (−0.4 SD; r = .95); the slope fell quite a bit from this peak (1.5 SD, r = .45).

Figure 13

Slopes of Reading Predictor-Outcome Pairs (z-standardized W scores)



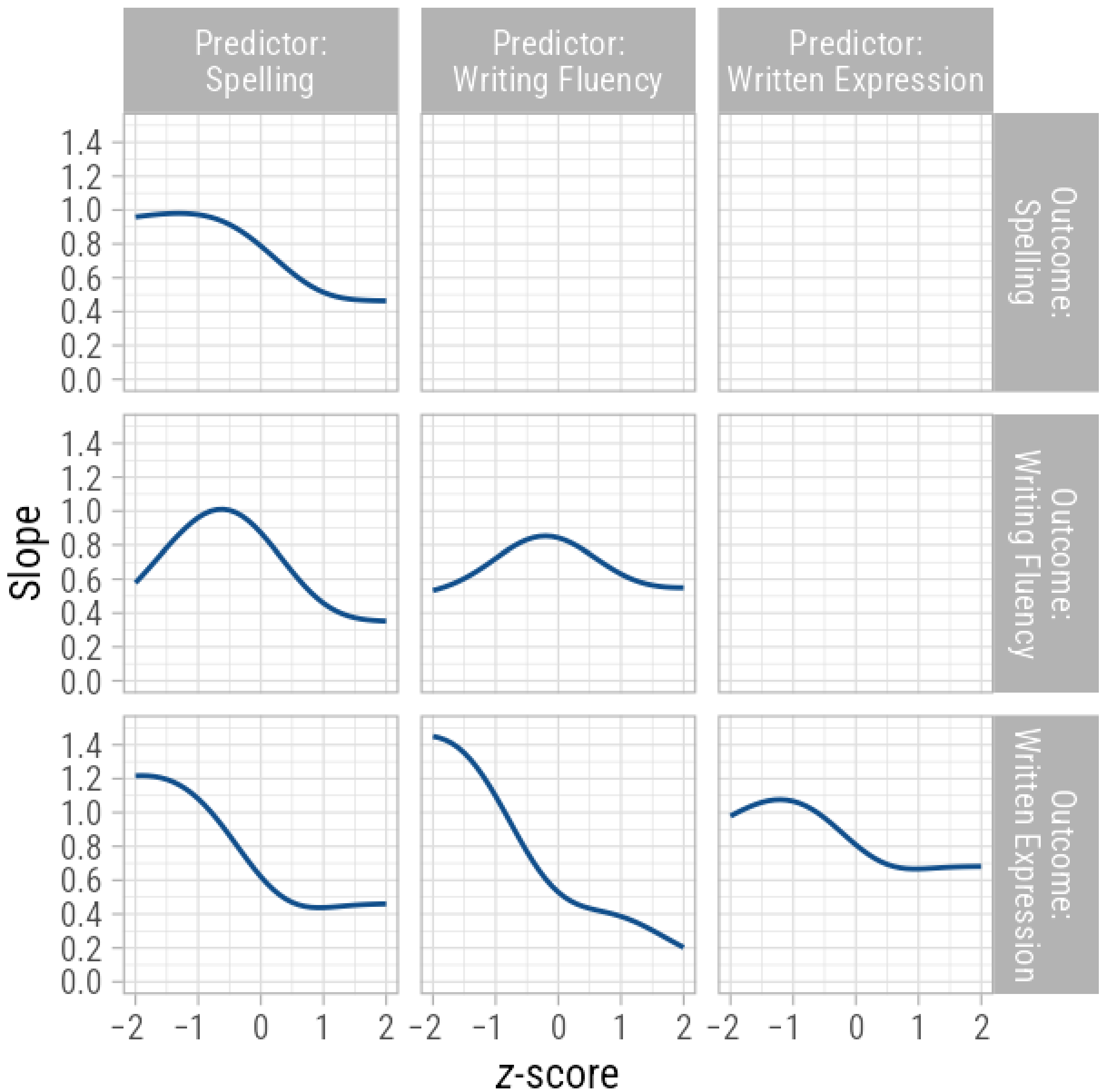
Writing

Writing contained many non-linear relationships. Spelling predicting writing fluency revealed a parabolic shape in the slope, where the relationship was strongest at -0.5 SD (r = 1) and weakened at the tails (−2 SD, $r = .58$; 2 SD, $r = .35$). A similar pattern occurred when writing fluency predicted itself, although the overall relationship had less variation. For instance, the slope never rose higher than $r = .9$ and never fell below $r = .5$. Writing fluency predicting written expression was responsible for the largest slope reduction across all writing pairs. Writing fluency had an incredibly strong relationship with written expression at the low end of the spectrum (−2 SD, $r = 1.45$), then the relationship weakened drastically at the mean and beyond (0 SD, $r = .53$; 2 SD, $r = .2$).

Interestingly, spelling as a predictor appeared to have a much stronger relationship with itself and with written expression at the lower end of the distribution. Starting with the spelling-spelling predictor-outcome duo, the slope was strong between the −2 SD and the mean ($r = .96$; $r = .79$) but significantly dropped past the mean (2 SD, $r = .46$). Spelling predicting written expression showed a similar pattern, where the relationship was strong at the low end (−2 SD, $r = 1.22$) but lowered considerably at the high end (2 SD, $r = .46$). Written expression predicting itself revealed a non-linear slope, with less variability than its counterparts. Before the mean, written expression was a strong predictor of written expression (−1 SD, $r = 1.1$). Past the mean, the slope decreased but still remained strong (2 SD, $r = .68$).

Figure 14

Slopes of Writing Predictor-Outcome Pairs (z-standardized W scores)



Generalizeability to Other Batteries

The relationships described above included many batteries, which all produced at least one significant non-linear pair within academic tests. In order to fully investigate whether the results are generalizable to other batteries, one would need to convert raw item scores to *W* scores, which was not possible for this experiment. *W* scores are rooted in item response theory (IRT) and use logit-based calculations to approximate an individual's ability. Thus, future research must acquire raw item scores in order to fully test whether these results are generalizable across batteries.

Discussion

The purpose of this research was to investigate whether math, reading, and writing subtests contained non-linear relationships across nine academic batteries (KTEA-II & KTEA-3, WIAT-III, TOD-C & TOD-E, and WJ 77, WJ-R, WJ III, WJ IV, and WJ V). Our previous findings solely focused on predictor-outcome pairs for reading tasks (e.g., decoding, fluency, and comprehension) across six academic batteries (KTEA 2 & 3, WIAT-III, and the WJIII, WJIV, & WJV), which revealed statistically significant differences in the slopes between linear and non-linear models at the lower- and higher-ends of the academic distribution (Campagnolio et al., 2025). Specifically, lower- and higher-scoring individuals' abilities were overestimated by a linear model, thus portraying their abilities to be higher than they actually were, according to the non-linear GAMs. Additionally, the rate of growth, or the slopes, of the predictor-outcome pairs were much stronger on the lower end of the distribution than the higher end of the distribution. Our previous findings were replicated with the current research, which included three more academic batteries, two additional subjects (math and writing), and *W* scores.

The findings of the current research are twofold: 1) overall, linear models tend to overestimate and thus inflate individuals' scores on both ends of the academic distribution when compared to a non-linear model, and 2) the correlation between the predictor-outcome pairs are much stronger at the lower end of the academic achievement distribution (i.e., below the mean) compared to the higher end of the academic achievement distribution (i.e., above the mean). These findings include both scores types: overall, non-linear models (like GAMs) provide a more accurate representation of all students' abilities and highlight the disparities in the strength of the slopes below and above the mean. The rationale behind these findings have not yet been investigated. Potential explanations for the overall difference in lower-scoring individuals' slopes and higher-scoring individuals' slopes include 1) the inherent 'skill-ceiling' within basic and moderately difficult academic tasks and 2) the role of an individual's age and its influence on the slopes.

Ceiling effects are well-known in psychology and measurement research, and it refers to the scenario where the sample's scores tend to congregate at or near a suspected "upper limit" (Garin, 2023). For example, researchers who wish to measure a random sample of adults' abilities to name all 26 letters of the alphabet would presumably get uniform results. A task that wished to measure whether children could correctly identify whether it is day or night outside would similarly run into ceiling effects. With this in mind, it would be reasonable to assume that an individual's ability to decode or spell can only get so good. In other words, once an individual is competent in basic spelling or decoding, there is only so much improvement to be made. Additionally, someone who scores 2 SD above the mean on a spelling task does not necessarily understand the underlying mechanisms of spelling more than their peer who scored in the average range. Instead, the higher scoring individual simply knows how to spell more unique and rarer words than their peer.

The glaring difference between standard scores and *W* scores is the inclusion of age. Standard scores compare one's performance to their similarly-aged peers, whereas *W* scores can be compared across the age continuum. This presents many differences in utility and interpretation. For example, a low standard score (e.g., 65) reflects a score that is more than two standard deviations below the mean, but it does not lend to a clear interpretation of what skill(s) the person cannot perform. Further, this low standard score means two separate things depending on the person's age: the individual cannot read basic words, if they are young child; or the individual can technically read, but they do not possess the sophisticated decoding skills as their peers, if they are an adult). On the contrary, *W* scores center around 500 and include an individual's developmental level, which allows for ability comparisons across age.

Limitations and Future Directions

Delineating the main component of academic subtests was an ambitious and sometimes arbitrary process. Our methodology required a neat and tidy organization of academic subtests into singular domains (e.g., writing fluency, decoding, applied math, etc.). However, many academic subtests are multifaceted and assess multiple skills simultaneously. For example, a reading comprehension subtest requires both decoding and reading fluency skills and a writing fluency task requires both spelling and decoding skills. Subtests were manually grouped into their main component (e.g., a writing fluency task was grouped into the 'fluency' category) as determined by the researchers, which is one limitation of this research.

To illustrate the previous point, the Tests of Dyslexia (TOD) blend multiple reading abilities into one subtest (e.g., fluency and reading comprehension are embedded in the 'reading comprehension' subtest). This combination blurs the line between the true main component of the task. Narrowing in one the 'main' component of a complex reading task has the risk to miss the empirical forest for the trees. In other words, a 1-to-1 predictor-outcome pair may overshadow the true nature of a complex reading task (e.g., reading comprehension), which are known to require multiple, smaller reading abilities to be used in conjunction.

Future research should investigate plausible explanations of the non-linearity in these plots. It is theorized that age is a key factor in the non-linear relationships observed, but the extent of its significance has yet to be tested. Future research should investigate whether an individual's age may be the best explanation of non-linear relationships between academic tests, especially for *W* scores. The subtests themselves warrant further investigation, as many subtests were gathered from nine academic batteries.

Conclusion

There is clear evidence of non-linear relations within reading and writing across both score types (standard scores and *W* scores). Academic subjects with standard scores appear to highlight how linear models overestimate individuals' abilities at both the lower and higher end of the academic achievement spectrum. Additionally, these plots show that the rate of skill development (i.e., the slope) for poor readers and writers is much stronger with a GAM compared to a linear model. On the other hand, the rate of improvement evidently decreases at the higher end of the achievement spectrum across both score types. *W* scores also convey a non-linear relationship, although more pronounced than the standard scores. Specifically, the slopes among the *W* scores were greater, overall, compared to the slopes of the standard scores.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, D., Maechler, M., & Jagan, M. (2025). *Matrix: Sparse and dense matrix classes and methods*. https://doi.org/10.32614/CRAN.package.Matrix

Campagnolio, A., Schneider, W. J., & An, C. (2025). *Virtual Poster: Non-linear Relations Among Reading Abilities* . National Association of School Psychologists (NASP). https://apps.nasponline.org/resources-and-publications/resources/virtual-posters/viewer.aspx?ID=7679

Firke, S. (2024). *janitor: Simple tools for examining and cleaning dirty data*. https://doi.org/10.32614/CRAN.package.janitor

Garin, O. (2023). Ceiling Effect. In *Encyclopedia of Quality of Life and Well-Being Research* (pp. 704–706). Springer, Cham. https://doi.org/10.1007/978-3-031-17299-1_296

Gohel, D., & Skintzos, P. (2026). *flextable: Functions for tabular reporting*. https://doi.org/10.32614/CRAN.package.flextable

Hajovsky, D. B., Chesnut, S. R., & Jensen, K. M. (2020). The role of teachers' self-efficacy beliefs in the development of teacher-student relationships. *Journal of School Psychology*, *82*, 141–158. https://doi.org/10.1016/j.jsp.2020.09.001

Hajovsky, D. B., Niileksela, C. R., Flanagan, D. P., Alfonso, V. C., Schneider, W. J., & Robbins, J. (2025). Toward a consensus model of cognitive–reading achievement relations using meta-structural equation modeling. *Journal of Intelligence*, *13*(8), 104. https://doi.org/10.3390/jintelligence13080104

Hastie, T., & Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science*, *1*(3), 297–310. https://doi.org/10.1214/ss/1177013604

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Test of Educational Achievement-Second Edition (KTEA-II)*. American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2014). *Kaufman Test of Educational Achievement - Third Edition (KTEA-3)*. NCS Pearson.

Makowski, D., Ben-Shachar, M. S., Wiernik, B. M., Patil, I., Thériault, R., & Lüdecke, D. (2025). modelbased: An R package to make the most out of your statistical models through marginal means, marginal effects, and model predictions. *Journal of Open Source Software*, *10*(109), 7969. https://doi.org/10.21105/joss.07969

Mather, N., McCallum, R. S., Bell, S. M., & Wendling, B. J. (2024). *Tests of dyslexia*. Western Psychological Services.

McGrew, K. S., Mather, N., LaForte, E. M., & Wendling, B. J. (2025). *Woodcock-Johnson V*. Riverside Assessments, LLC.

Ooms, J. (2025). *writexl: Export data frames to excel "xlsx" format*. https://doi.org/10.32614/CRAN.package.writexl

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-PLUS*. Springer. https://doi.org/10.1007/b98882

Pinheiro, J., Bates, D., & R Core Team. (2025). *nlme: Linear and nonlinear mixed effects models*. https://doi.org/10.32614/CRAN.package.nlme

Posit team. (2025). *RStudio: Integrated development environment for R* [Manual]. Posit Software, PBC. http://www.posit.co/

R Core Team. (2025b). *R: A language and environment for statistical computing* (Version 4.5.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2025a). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2026). *foreign: Read data stored by "Minitab," "S," "SAS," "SPSS," "Stata," "Systat," "Weka," "dBase," ...* https://doi.org/10.32614/CRAN.package.foreign

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rosseel, Y., Jorgensen, T. D., & De Wilde, L. (2025). *lavaan: Latent variable analysis*. https://doi.org/10.32614/CRAN.package.lavaan

Schneider, W. J. (2021). *WJSmisc: Miscellaneous functions from w. Joel schneider*. https://github.com/wjschne/WJSmisc

Schneider, W. J. (2025a). *apa7: Facilitate writing documents in american psychological association style, seventh edition*. https://doi.org/10.32614/CRAN.package.apa7

Schneider, W. J. (2025b). *ggdiagram: Object-oriented diagram plots with "ggplot2"*. https://doi.org/10.32614/CRAN.package.ggdiagram

Schneider, W. J., & McGrew, K. S. (2018). The Cattell–Horn–Carroll theory of cognitive abilities. In *Contemporary intellectual assessment: Theories, tests, and issues, 4th ed* (pp. 73–163). The Guilford Press.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV*. Riverside.

Wechsler, D. (2009). *Wechsler Individual Achievement Test, Third Edition (WIAT-III)*. Pearson.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019a). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019b). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilke, C. O., & Wiernik, B. M. (2022). *ggtext: Improved text rendering support for "ggplot2"*. https://doi.org/10.32614/CRAN.package.ggtext

William Revelle. (2026). *psych: Procedures for psychological, psychometric, and personality research*. Northwestern University. https://CRAN.R-project.org/package=psych

Wood, S. N. (2003). Thin-plate regression splines. *Journal of the Royal Statistical Society (B)*, *65*(1), 95–114. https://doi.org/10.1111/1467-9868.00374

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, *99*(467), 673–686. https://doi.org/10.1198/016214504000000980

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wood, S. N. (2017). *Generalized Additive Models: An introduction with R* (2nd ed.). Chapman; Hall/CRC.

Wood, S. N. (2025). Generalized Additive Models. *Annual Review of Statistics and Its Application*, *12*, 497–526. https://doi.org/10.1146/annurev-statistics-112723-034249

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models (with discussion). *Journal of the American Statistical Association*, *111*, 1548–1575. https://doi.org/10.1080/01621459.2016.1180986

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock-Johnson Psychoeducational Battery*. DLM.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery–Revised*. Riverside.

Woodcock, R. W., McGrew, K. S., Schrank, F. A., & Mather, N. (2007). *Woodcock-Johnson III Normative Update*. Riverside Publishing.

Yasumoto, A. (2024). *ftExtra: Extensions for "Flextable"*. https://doi.org/10.32614/CRAN.package.ftExtra

Appendix

Relationship Description Scale

Table A1

Correlation Coefficients (slope) for Math Predictor-Outcome Pairs at both ends of the Academic Distrbution (standard scores)

Outcome	Predictor	Slope at Predictor Value												
		70	75	80	85	90	95	100	105	110	115	120	125	130
Math Fluency	Calculation	.61	.60	.60	.60	.59	.57	.54	.52	.49	.48	.46	.46	.45
	Math Fluency	.50	.50	.51	.52	.53	.55	.57	.57	.56	.53	.49	.44	.39
Applied Math	Calculation	.68	.70	.70	.69	.68	.67	.67	.67	.67	.67	.66	.66	.65
	Math Fluency	.67	.65	.62	.58	.54	.51	.50	.49	.48	.46	.43	.40	.38
	Applied Math	.86	.83	.82	.81	.80	.79	.77	.75	.73	.73	.73	.74	.74

Table A2

All Correlation Coefficients (slope) for Reading Predictor-Outcome Pairs at both ends of the Academic Distrbution (standard scores)

Outcome	Predictor	Slope at Predictor Value												
		70	75	80	85	90	95	100	105	110	115	120	125	130
Decoding	Decoding	.77	.76	.74	.72	.69	.67	.63	.60	.57	.54	.53	.52	.52
Reading Fluency	Decoding	.62	.63	.63	.63	.61	.57	.52	.45	.38	.33	.29	.29	.30
	Reading Fluency	.67	.71	.72	.72	.69	.66	.63	.60	.58	.54	.47	.40	.34

Outcome	Predictor	Slope at Predictor Value												
		70	75	80	85	90	95	100	105	110	115	120	125	130
Comprehension	Decoding	.83	.77	.72	.68	.65	.62	.58	.54	.49	.45	.43	.43	.45
	Reading Fluency	.67	.64	.61	.57	.53	.49	.45	.42	.38	.35	.31	.27	.24
	Comprehension	.79	.79	.77	.73	.69	.65	.62	.61	.61	.60	.59	.57	.55

Table A3

All Correlation Coefficients (slope) for Writing Predictor-Outcome Pairs at both ends of the Academic Distrbution (standard scores)

Outcome	Predictor	Slope at Predictor Value												
		70	75	80	85	90	95	100	105	110	115	120	125	130
Spelling	Spelling	.72	.71	.69	.68	.67	.66	.65	.62	.59	.54	.49	.44	.41
Writing Fluency	Spelling	.55	.57	.57	.57	.55	.51	.47	.42	.38	.35	.33	.32	.32
	Writing Fluency	.46	.47	.49	.52	.57	.61	.64	.65	.64	.60	.56	.53	.50
Written Expression	Spelling	.75	.72	.69	.66	.62	.59	.56	.53	.50	.47	.44	.41	.39
	Writing Fluency	.59	.60	.59	.58	.56	.52	.47	.41	.36	.31	.27	.25	.24
	Written Expression	.78	.74	.70	.66	.64	.64	.64	.63	.61	.57	.53	.49	.46

Table A4

All Correlation Coefficients (slope) for Math Predictor-Outcome Pairs at both ends of the Academic Distrbution (z-standardized W scores)

Outcome	Predictor	Slope at Predictor Value								
		−2.0	−1.5	−1.0	−0.5	0.0	0.5	1.0	1.5	2.0
Math Fluency	Calculation	.59	.75	.91	.99	.94	.80	.68	.63	.62
	Math Fluency	.74	.83	.94	.95	.83	.64	.52	.50	.53
Applied Math	Calculation	.72	.87	1.01	1.04	.96	.85	.82	.85	.86
	Math Fluency	1.00	1.11	1.18	1.08	.79	.49	.36	.40	.48
	Applied Math	.95	1.02	1.06	1.02	.91	.82	.78	.78	.78

Table A5

All Correlation Coefficients (slope) for Reading Predictor-Outcome Pairs at both ends of the Academic Distrbution (z-standardized W scores)

Outcome	Predictor	Slope at Predictor Value								
		−2.0	−1.5	−1.0	−0.5	0.0	0.5	1.0	1.5	2.0
Decoding	Decoding	.95	1.07	1.10	1.01	.85	.74	.71	.72	.72
Reading Fluency	Decoding	.52	.67	.84	.95	.90	.71	.52	.45	.44
	Reading Fluency	1.00	1.00	.99	.95	.88	.82	.78	.76	.76
Comprehension	Decoding	.97	1.08	1.09	.96	.75	.60	.55	.55	.55
	Reading Fluency	1.43	1.33	1.11	.83	.59	.43	.31	.20	.11

Outcome	Predictor	Slope at Predictor Value								
		−2.0	−1.5	−1.0	−0.5	0.0	0.5	1.0	1.5	2.0
	Comprehension	.98	1.02	1.01	.92	.77	.65	.60	.61	.61

Table A6

All Correlation Coefficients (slope) for Writing Predictor-Outcome Pairs at both ends of the Academic Distrbution (z-standardized W scores)

Outcome	Predictor	Slope at Predictor Value								
		−2.0	−1.5	−1.0	−0.5	0.0	0.5	1.0	1.5	2.0
Spelling	Spelling	.96	.98	.97	.91	.79	.63	.51	.47	.46
Writing Fluency	Spelling	.58	.78	.96	1.01	.87	.64	.46	.37	.35
	Writing Fluency	.53	.60	.72	.83	.84	.75	.63	.56	.55
Written Expression	Spelling	1.22	1.20	1.08	.86	.62	.47	.44	.45	.46
	Writing Fluency	1.45	1.35	1.10	.77	.53	.43	.39	.30	.20
	Written Expression	.98	1.06	1.07	.97	.81	.69	.67	.68	.68