



Center for Research in Educational
Policy

The University of Memphis
325 Browning Hall
Memphis, Tennessee 38152
Toll Free: 1-866-670-6147

Words Their Way Spelling Inventories:

Reliability and Validity Analyses





Center for Research in Educational
Policy

The University of Memphis
325 Browning Hall
Memphis, Tennessee 38152
Toll Free: 1-866-670-6147

Words Their Way Spelling Inventories:

Reliability and Validity Analyses

February 2007

Allan Sterbinsky, PhD
Center for Research in Educational Policy

Words Their Way Spelling Inventories

Reliability and Validation Study

Introduction

Words Their Way (WTW) is an approach to spelling and word knowledge that is based on extensive research literature and includes stages of development and instructional levels that are critical to the way students learn to read. It compliments the use of phonics, spelling, and vocabulary instruction that are often used in schools. Included in the WTW approach is a set of three inventories that assess student ability in key areas. These three inventories include the Primary Spelling Inventory, the Elementary Spelling Inventory, and the Upper Level Spelling Inventory.

As with all educational instruments, it is essential to evaluate the reliability and validity of the instruments to ensure that educators and policymakers base their instructional and policy decisions on instruments that measure what they purport to measure. Additionally, if the instruments are to be used to gauge changes in student knowledge, then it is critical that the instrument be reliable. With reliable instruments, educators can be sure that changes in test scores reflect changes in student knowledge rather than any instability in the instrument itself. For these reasons, the Center for Research in Educational Policy (CREP) at The University of Memphis was asked to conduct a reliability and validity study of all three inventories, using data from students in a variety of grades and backgrounds. Results of the reliability (measured in two different ways) and validity (both in predictive and concurrent contexts) analyses are discussed in the next sections.

Method

Sample

The school district that agreed to participate in this reliability and validity study included a total of 10,902 students. Of these students, 49% were female and 51% were male. Across the district, approximately 8% ($N = 847$) were identified as eligible for special education services. The sample of participating students came from seven schools within the district. These schools served a total of 4290 students.

Of the seven schools that agreed to participate in the research study, two were middle schools and five were elementary schools. As seen in Table 1, the size of the schools ranged from a low of 426 students to a high of 1,098 students. The two middle schools and three of the elementary schools served primarily Hispanic students (55% to 76% Hispanic). The two remaining elementary schools served primarily Caucasian students (51% to 71% Caucasian). All schools were located in a suburban environment and the percentage of students eligible for free or reduced-priced lunches ranged from a low of 35% to a high of 74% across the schools.

Table 1
Participating Schools

	School A	School B	School C	School D	School E	School F	School G
	Middle	Elem.	Elem.	Elem.	Elem.	Elem.	Middle
# Teachers Grades 1-2	0	9	1	5	4	5	0
# Teachers Grades 3-4	0	6	1	6	2	4	0
# Teachers Grades 5-6	3	3	1	3	0	0	1
# Teachers Grades 7-8	0	0	0	0	0	0	2
Total # students in school	1,098	481	632	531	628	426	494
Percent Native American	1.2	0.6	2.5	1.5	1.4	2.1	1.0
Percent Asian	3.6	2.7	1.3	1.7	2.2	2.8	2.4
Percent African American	5.7	9.4	6.5	4.7	5.7	4.5	6.7
Percent Hispanic	54.8	11.6	75.6	68.2	67.2	37.8	67.6
Percent White/non Hispanic	32.8	71.3	13.6	21.7	21.7	50.9	20.9
Free/Reduced Lunch	56%	36%	74%	62%	67%	35%	73%
Urbanicity	Suburban	Suburban	Suburban	Suburban	Suburban	Suburban	Suburban

Instrumentation

Words Their Way Instruments. The research study included three separate spelling inventories, the Primary Spelling Inventory, the Elementary Spelling Inventory, and the Upper Level Spelling Inventory. The Primary Spelling Inventory is comprised of 26 spelling words, ranging from “fan” to “riding.” The Elementary Spelling Inventory includes 25 words, ranging from “bed” to “opposition.” Finally, the Upper Level Spelling Inventory is comprised of 31 words, including “switch” and “succession.”

In proctoring these tests, administrators called out the words to the students, used the words in sentences, and then repeated the words. Students then spelled the words on a sheet of paper.

California Standards Tests. As part of the state-mandated STAR (Standardized Testing and Reporting) program in California, the English-Language Arts (ELA) tests are administered to students in the second through 11th grades during the spring of each academic year. The tests are composed of subtests including the Word Analysis and Vocabulary Development, Reading Comprehension, Literary Response and Analysis, Written Conventions, and Writing Strategies. Additionally, the English Language Arts Cluster 6 Writing Applications is administered in grades 4 and 7. The California Standards Tests also include an ELA scale score and performance levels (CST ss ELA and the CST pl ELA, respectively), which are included as separate scales in this analysis.

Procedure

A total of 1,944 Words Their Way tests were completed at the participating schools during the fall of 2005 as seen in Table 2. Some students in the sample completed multiple inventories (e.g. primary and elementary). A total of 647 students completed the Primary Spelling Inventory, 862 completed the Elementary Spelling Inventory, and 442 completed the Upper Level Inventory. Data from these students were analyzed using indices of item difficulty, item discrimination, and Cronbach's alpha (internal consistency).

During the spring of 2006, 901 students from these schools participated in the test-retest reliability portion of this study. Students were asked to complete a primary, elementary, or upper form of the instruments, and one week later, the same students were asked to complete the same instrument again. The first spring 2006 test was treated as the pretest, while the second Spring 2006 test was treated as a posttest. Together, the scores from the test administrations provided an estimate of the test-retest reliability of the instruments.

During the spring of 2006, students in California were required to participate in standardized tests, which included the students in this district. These tests were proctored via the state-mandated protocols and test results for each student were made available to the researchers. Student names and other identifiers were withheld from researchers to protect the confidentiality of data, but a unique identifier was used to link WTW test results with those from the standardized tests. Results from a total of 685 students (52% female, 48% male) were linked in the database, including 133 in the second grade, 181 in the third grade, 165 in the fourth grade, and 192 in the fifth grade. Results from approximately 14 students in higher grades were available, but were not used in the validation study due to insufficient sample size for those grades.

Validity estimates were calculated using both a predictive and concurrent design. The Fall 2005 (first administration) of the Words Their Way instrument was used as a predictor and the Spring 2006 administration of the California Standards Test as the criterion. The next validity estimate (concurrent validity) used the Spring 2006 (second administration) as the predictor and the Spring 2006 administration of the California Standards Test as the criterion. Additionally, validity estimates were derived from all students in the study, and an additional validity estimate was derived from a subsample of these original students who were not identified as ELL, SPED, or Gifted. Reporting predictive and concurrent results for both samples provides educators and policymakers with clear evidence of the validity of the instruments within both a predictive and concurrent context.

Table 2
Number of Students Participating by School, Grade, and WTW Version

School	WTW Version	Grade	Grade	Grade	Grade	Grade	Grade	Grade	Grade
		1	2	3	4	5	6	7	8
A	Primary	0	0	0	0	0	0	0	0
A	Elementary	0	0	0	0	0	149	0	0
A	Upper	0	0	0	0	0	149	0	0
B	Primary	94	73	59	0	0	0	0	0
B	Elementary	0	32	81	59	77	0	0	0
B	Upper	0	0	0	0	79	0	0	0
C	Primary	0	19	0	0	0	0	0	0
C	Elementary	0	18	0	28	23	0	0	0
C	Upper	0	0	0	0	25	0	0	0
D	Primary	53	47	54	0	0	0	0	0
D	Elementary	0	31	54	75	80	0	0	0
D	Upper	0	0	0	0	80	0	0	0
E	Primary	67	9	34	0	0	0	0	0
E	Elementary	0	0	32	0	0	0	0	0
E	Upper	0	0	0	0	0	0	0	0
F	Primary	57	31	50	0	0	0	0	0
F	Elementary	0	29	48	0	0	0	0	0
F	Upper	0	0	0	0	0	0	0	0
G	Primary	0	0	0	0	0	0	0	0
G	Elementary	0	0	0	0	0	46	0	0
G	Upper	0	0	0	0	0	47	46	9

Results

Data from the first administration of the three inventories was analyzed using estimates of item difficulty, item discrimination, and internal consistency (Cronbach's alpha). Item difficulty was calculated by identifying the percent of students who correctly spelled each word. Higher numbers in the item difficulty index identify items that are easier for students to spell. Lower numbers identify items that are harder for students to spell.

Item discrimination was also analyzed. The purpose of item discrimination is to provide an index of how well an individual item discriminates between students who scored relatively high on the total test score versus those who scored relatively low. For this index, lower numbers indicate that an item does not differentiate between higher and lower performers on the overall test. Higher numbers, however, indicate that the item has substantive ability to differentiate between higher and lower overall performers on the test.

Finally, the data from the first administration was analyzed to estimate the internal consistency of the overall instrument using Cronbach's alpha. This statistic is appropriate for use with dichotomous variables (e.g. word correct or incorrect).

Primary Spelling Inventory

As can be seen in Table 3, the item difficulty ranged from a high of 96.3 (fan) to a low of 16.1 (clapping). The index of discrimination ranged from a low of 6.3 (fan) to a high of 77.7 (shine). Analysis of the reliability of the Primary Spelling Inventory using Cronbach's alpha procedure indicated an overall reliability coefficient of .9341. Individual items were then examined to determine if deletion of any items would substantively improve the overall reliability index. No item was recommended for deletion on the basis of the impact on coefficient alpha. It is, however, recommended that results from the item difficulty and item discrimination indices be used to consider continued inclusion or exclusion of individual items from the instrument.

Table 3
Primary Spelling Inventory Item Difficulty and Index of Discrimination (N = 647)

Item	Item Difficulty	Index of Discrimination	Mean	St. Dev.	Alpha if Item Deleted
blade	44.8	66.70	.448	.498	.930
camped	35.4	63.90	.354	.479	.930
clapping	16.1	29.30	.178	.563	.935
coach	32.9	58.40	.329	.470	.929
crawl	19.8	37.80	.198	.399	.931
dig	90.3	18.20	.903	.297	.934
dream	41.9	75.80	.419	.494	.928
fan	96.3	6.30	.963	.189	.935
fright	29.5	56.80	.295	.457	.929
growl	26.6	45.40	.266	.442	.931
gum	87.8	22.10	.878	.328	.934
hope	61.2	65.10	.612	.488	.931
pet	92.3	12.00	.923	.267	.934
riding	40.8	43.40	.408	.492	.933
rob	82.8	13.70	.828	.377	.935
chewed	24.3	46.50	.243	.429	.930
shine	52.7	77.70	.527	.500	.929
shouted	33.7	64.30	.337	.473	.930
sled	79.0	24.10	.790	.408	.935
spoil	24.6	47.10	.263	.604	.933
stick	62.9	57.30	.629	.483	.932
third	28.3	51.90	.283	.451	.930
thorn	55.6	54.30	.556	.497	.932
tries	23.6	41.60	.236	.425	.931
wait	33.7	59.30	.337	.473	.930
wishes	41.0	69.70	.410	.492	.929

Elementary Spelling Inventory

For the Elementary Spelling Inventory, the item difficulty indices ranged from a high of 98.9 (bed) to a low of 15.0 (opposition) (see Table 4). The index of discrimination ranged from a low of 2.2 (bed) to a high of 65.3 (carries). Examination of the internal consistency of the instrument yielded an overall reliability coefficient of .915 (Cronbach's alpha). Examination of the alpha levels if an item was deleted indicated that no items should be removed solely on the basis of the change in alpha levels. Results from the item difficulty and item discrimination indices should,

however, be used to consider continued inclusion or exclusion of individual items from the instrument.

Table 4
Elementary Spelling Inventory Item Difficulty and Index of Discrimination (N = 856)

Item	Item Difficulty	Index of Discrimination	Mean	St. Dev.	Alpha if Item Deleted
carries	49.4	65.30	.494	.500	.910
favor	55.8	64.20	.558	.497	.909
chewed	66.8	59.10	.668	.471	.908
throat	44.6	59.10	.446	.497	.911
pleasure	32.8	57.70	.328	.470	.911
bottle	68.2	57.50	.682	.466	.909
spoil	67.5	56.20	.675	.469	.909
confident	32.7	52.20	.327	.469	.911
serving	58.4	51.50	.584	.493	.911
float	73.0	50.60	.730	.444	.909
marched	79.6	38.20	.796	.403	.910
bright	80.8	35.50	.808	.394	.910
civilize	19.3	35.30	.193	.395	.913
fortunate	17.4	33.10	.174	.379	.913
shower	83.5	32.80	.835	.371	.910
ripen	59.6	31.80	.596	.491	.916
cellar	16.4	29.20	.176	.524	.917
train	85.0	27.80	.850	.357	.911
opposition	15.0	27.30	.149	.357	.914
place	89.0	22.50	.890	.313	.911
lump	88.7	21.80	.887	.317	.913
when	92.8	14.40	.928	.259	.913
drive	92.8	13.40	.928	.259	.913
ship	96.7	6.30	.967	.178	.915
bed	98.9	2.20	.989	.102	.916

Upper Level Spelling Inventory

The item difficulty for the Upper Level Spelling Inventory ranged from a high of 88.2 (shaving) to a low of 7.2 (correspond). The index of discrimination ranged from a low of 12.3 (circumference) to a high of 62.5 (smudge). Cronbach's Alpha yielded an overall reliability estimate of .9086. By examining how the removal of an item would impact alpha, it is not recommended that any item be deleted. Results from the item difficulty and index of discrimination

should be examined to determine if an individual item (or items) should be removed from the instrument. See Table 5 for summary statistics.

Table 5
Upper Level Spelling Inventory Item Difficulty and Index of Discrimination (N = 442)

Item	Item Difficulty	Index of Discrimination	Mean	St. Dev.	Alpha if Item Deleted
chlorine	11.1	20.40	.111	.314	.907
circumference	10.6	12.30	.106	.309	.908
civilization	36.2	53.90	.362	.481	.904
commotion	18.1	30.60	.181	.386	.906
confidence	48.6	57.70	.486	.500	.904
correspond	7.2	12.40	.072	.259	.908
crater	68.6	22.40	.685	.465	.910
disloyal	65.6	55.20	.656	.475	.904
dominance	14.9	23.80	.149	.357	.906
emphasize	11.5	17.60	.115	.320	.908
fortunate	35.5	55.40	.355	.479	.904
humor	63.1	52.40	.631	.483	.904
illiterate	16.7	27.20	.167	.374	.906
irresponsible	17.6	28.80	.176	.382	.906
knotted	30.1	39.00	.301	.459	.906
medicinal	17.2	25.30	.172	.378	.907
monarchy	18.8	27.30	.188	.391	.906
opposition	27.8	37.60	.278	.449	.906
pounce	76.9	40.40	.769	.422	.905
sailor	55.9	47.30	.559	.497	.906
scrape	73.5	39.70	.735	.442	.905
scratches	58.6	61.50	.586	.493	.903
shaving	88.2	18.50	.882	.323	.908
smudge	52.3	62.50	.523	.500	.903
squirt	72.6	50.80	.726	.446	.903
succession	7.5	12.80	.075	.263	.908
switch	77.6	33.50	.776	.417	.906
trapped	56.3	50.90	.563	.496	.905
tunnel	59.7	61.70	.597	.491	.903
village	82.8	27.60	.828	.378	.906
visible	43.2	30.40	.432	.496	.908

Norms

Based on the results of the Fall 2005 administration, the following norms were derived for each instrument (see Tables 6 – 8 for primary, elementary, and upper norms, respectively). These norms should be considered within the context of the geographic location of the study data, as well

as the demographic data for the schools in the sample before being applied to other geographic locations and/or demographically similar/dissimilar populations. Given the small sample size for these norms, generalization of these results should be viewed cautiously.

Table 6
Primary Spelling Inventory – Norms

Grade	N	Mean	Minimum	Maximum	St. Dev.
1	271	7.04	0	19	3.34
2	167	13.83	1	25	5.55
3	209	18.76	2	26	6.22
All Grades	647	12.58	0	26	7.12

Table 7
Elementary Spelling Inventory – Norms

Grade	N	Mean	Minimum	Maximum	St. Dev.
2	114	8.88	1	20	4.49
3	191	13.49	1	24	5.40
4	174	15.56	2	24	4.67
5	182	18.36	1	25	5.11
6	195	19.27	6	25	3.79
All Grades	856	15.65	1	25	5.84

Table 8
Upper Level Spelling Inventory – Norms

Grade	N	Mean	Minimum	Maximum	St. Dev.
4	8	16.75	10	27	5.70
5	183	13.36	0	30	7.35
6	196	13.25	1	28	6.44
7	46	12.22	4	26	5.35
8	9	13.11	1	29	9.55
All Grades	442	13.25	0	30	6.79

Test-Retest Reliability Estimates

Test-retest reliability was estimated for each of the Words Their Way instrument versions. Two forms of test-retest reliability were calculated. The first estimates the reliability of the instrument using the fall 2005 test as the pretest and the spring 2006 test (third administration) as the posttest. This estimates the reliability of the instrument with an interval of four months between

test administrations. The second form of test-retest reliability was calculated based solely on the spring 2006 administrations of the instruments. These calculations included the Spring 2006 (second overall administration) and the Spring 2006 (third overall administration), with a one week interval between tests.

Reliability estimates were also separated into two samples. The first sample included all students, including any students identified as ELL, SPED, or Gifted. The second reliability estimate included only those students that were not identified as ELL, SPED, or Gifted. Reporting reliability estimates for both samples provides educators and policymakers a clearer picture of the reliability for the instruments with differing populations of students.

Primary Spelling Inventory. For the Primary Inventory, as seen in Table 9, the test-retest reliability estimates for the second grade students ranged from a low of .82 when using the Fall 2005 administration as the pretest, to a high of .931 when using the Spring 2006 administration as the pretest. For the third grade, the estimates ranged from .764 to .946. With both samples (including and excluding ELL, SPED, and Gifted students), the reliability estimates using the Spring 2006 pretest were at least .90, which is an acceptable level of reliability. The strength of the reliability across four months is clearly evident from these data. All coefficients were significant at the $p < .001$ level.

Table 9
Test-retest Reliability Estimates using Spring 2006 (Third Administration) as the Retest - Primary Spelling Inventory

	Includes All Students	Excludes ELL, SPED and Gifted Students
<u>Second Grade</u>		
Fall 05 Pretest	0.824	0.729
Spring 06 Pretest	0.931	0.898
<u>Third Grade</u>		
Fall 05 Pretest	0.764	0.719
Spring 06 Pretest	0.946	0.949

Elementary Spelling Inventory. For the Elementary Inventory, the reliability estimates for all students ranged from .931 to .974 using the Spring 2006 (second administration) as the pretest and the Spring 2006 (third administration) as the posttest (as seen in Table 10). The coefficients using the Fall 2005 (first administration) as the pretest were a bit lower, ranging from .700 to .898. All coefficients were significant at the $p < .001$ level.

Table 10
Test-retest Reliability Estimates using Spring 2006 (Third Administration) as the Retest – Elementary Spelling Inventory

	Includes All Students	Excludes ELL, SPED and Gifted Students
<u>Second Grade</u>		
Fall 05 Pretest	0.781	0.740
Spring 06 Pretest	0.974	0.967
<u>Third Grade</u>		
Fall 05 Pretest	0.700	0.743
Spring 06 Pretest	0.950	0.936
<u>Fourth Grade</u>		
Fall 05 Pretest	0.898	0.873
Spring 06 Pretest	0.943	0.927
<u>Fifth Grade</u>		
Fall 05 Pretest	0.799	0.848
Spring 06 Pretest	0.959	0.942
<u>Sixth Grade</u>		
Fall 05 Pretest	0.742	.765
Spring 06 Pretest	0.931	.860

Upper Spelling Inventory. Finally, for the Upper Inventory, the reliability estimates for all students ranged from .818 using the Fall 2005 as the pretest, to .890 using the Spring 2006 as the pretest (see Table 11). The estimates for the restricted sample ranged from .765 to .860. These coefficients were significant at the $p < .001$ level.

Table 11

Test-retest Reliability Estimates using Spring 2006 (Third Administration) as the Retest – Upper Spelling Inventory

	Includes All Students	Excludes ELL, SPED and Gifted Students
<u>Fifth Grade</u>		
Fall 05 Pretest	0.818	0.765
Spring 06 Pretest	0.890	0.860

Summary. These results show clear evidence for the test-retest reliability of all forms of the Words Their Way inventories. This holds true using all students in the study or using only the sample of students not identified as ELL, SPED, or Gifted.

Validity Estimates

Validity coefficients were calculated using two separate designs and two samples. The predictive design used the Fall 2005 (first administration) as the predictor and the Spring 2006 California Standards Tests (CST) results as the criterion. The concurrent validity design used the Spring 2006 (second administration) as the predictor and the 2006 CST test results as the criterion. These estimates were calculated using the sample of all students as well as the subsample excluding students identified as ELL, SPED, or Gifted.

Primary Spelling Inventory. For the Primary Inventory, the predictive validity coefficients using the sample of all students ranged from a low of .540 (Reading Comprehension) to a high of .681 for (Word Analysis) for the second grade students as seen in Table 12. Concurrent validity coefficients for the second grade students ranged from a low of .484 (Reading Comprehension) to a high of .744 (Word Analysis). For the third grade students, the lowest predictive validation coefficient was .531 (Writing Strategies) while the highest coefficient was .726 (Word Analysis). These coefficients were all significant at the $p=.01$ level while some were significant at the $p<.001$ level. Calculation of the concurrent validity estimates for the third grade students ranged from a low of .474 (Reading Comprehension) to a high of .649 (Word Analysis). All coefficients were significant at least at the $p=.01$ level.

Table 12
Primary Spelling Inventory
Predictive and Concurrent Validity – Two Samples

	Includes All Students		Excludes ELL, SPED, Gifted students	
	Predictive Validity	Concurrent Validity	Predictive Validity	Concurrent Validity
<u>Second Grade</u>				
Reading – List	0.657	0.596	0.621	0.621
Word Analysis Vocabulary Development	0.681	0.744	0.609	0.743
Reading Comprehension	0.540	0.484	0.510	0.594
Literary Response and Analysis	0.556	0.646	0.571	0.624
Written Conventions	0.669	0.624	0.621	0.627
Writing Strategies	0.571	0.549	0.601	0.578
CST ss ELA	0.676	0.637	0.653	0.670
CST pl ELA	0.663	0.644	0.584	0.609
<u>Third Grade</u>				
Reading – List	0.720	0.614	0.536	0.468
Word Analysis Vocabulary Development	0.726	0.649	0.534	0.495
Reading Comprehension	0.623	0.474	0.400	0.287
Literary Response and Analysis	0.624	0.538	0.389	0.245
Written Conventions	0.691	0.544	0.579	0.377
Writing Strategies	0.531	0.558	0.408	0.576
CST ss ELA	0.710	0.603	0.530	0.470
CST pl ELA	0.711	0.611	0.531	0.463

Elementary Spelling Inventory. For the Elementary Inventory, the predictive validity coefficients from the second grade ranged from a high of .697 CST ELA proficiency level (CST pl ELA) to a low of .537 (Literacy Response and Analysis) (see Table 13). These coefficients were all significant at the $p=.05$ level. For the third grade, the predictive validity coefficients were substantively lower with a low of .135 (Writing Strategies) which was not significant, to a high of .553 (CST ss ELA). The only coefficients that attained significance for the third grade sample included Reading List, CST ss ELA, and CST pl ELA. For the fourth grade sample, the predictive coefficients were higher, with a low of .428 (ELA Cluster 6) to a high of .619 (Written Conventions). These were all significant at the $p<.001$ level. Lastly, for the fifth grade sample, the coefficients were substantively higher than for the third grade sample, and included a high of .706

(CST pl ELA) and a low of .501 (Reading Comprehension), which were all significant at the $p < .001$ level.

For the Elementary Inventory, the concurrent validity coefficients tended to be somewhat lower than the results from the predictive design. For the second grade sample, the lowest coefficient was .408 (Literary Response and Analysis) while the highest was .517 (CST pl ELA). All coefficients attained significance at least at the $p < .05$ level. For the third grade sample, the coefficients were substantively higher, with the lowest being .497 (Literary Response and Analysis) and the highest being .692 (Word Analysis). All coefficients for the third grade attained significance at the $p = .002$ level. In the fourth grade, the ELA Cluster 6 coefficient had the lowest validation coefficient (.384) but was still significant ($p < .001$). The highest coefficient was .656 (Written Conventions). All coefficients were significant at the $p < .001$ level. Finally, the fifth grade also showed evidence of concurrent validity. The lowest coefficient was .408 (Reading Comprehension) while the highest was .561 (CST ss ELA). All attained significance at the $p < .001$ level.

Table 13
Elementary Spelling Inventory
Predictive and Concurrent Validity – Two Samples

	Includes all students		Excludes ELL, SPED, Gifted students	
	Predictive Validity	Concurrent Validity	Predictive Validity	Concurrent Validity
<u>Second Grade</u>				
Reading – List	0.691	0.470	0.617	0.426
Word Analysis Vocabulary Development	0.604	0.473	0.5	0.418
Reading Comprehension	0.598	0.428	0.512	0.433
Literary Response and Analysis	0.537	0.408	0.414	0.327
Written Conventions	0.631	0.456	0.537	0.428
Writing Strategies	0.595	0.426	0.554	0.42
CST ss ELA	0.697	0.517	0.615	0.486
CST pl ELA	0.625	0.424	0.546	0.404
<u>Third Grade</u>				
Reading – List	0.547	0.667	0.513	0.616
Word Analysis Vocabulary Development	0.225	0.692	0.18	0.748
Reading Comprehension	0.227	0.519	0.194	0.381
Literary Response and Analysis	0.138	0.497	0.128	0.453
Written Conventions	0.150	0.651	0.139	0.69
Writing Strategies	0.135	0.577	0.126	0.631
CST ss ELA	0.553	0.686	0.52	0.657
CST pl ELA	0.523	0.686	0.468	0.641
<u>Fourth Grade</u>				
Reading – List	0.601	0.613	0.597	0.516
Word Analysis Vocabulary Development	0.565	0.597	0.601	0.534
Reading Comprehension	0.510	0.499	0.444	0.326
Literary Response and Analysis	0.468	0.586	0.464	0.499
Written Conventions	0.619	0.656	0.621	0.593
Writing Strategies	0.541	0.566	0.54	0.462
ELA Cluster 6	0.428	0.384	0.409	0.283
CST ss ELA	0.607	0.628	0.608	0.525
CST pl ELA	0.609	0.621	0.583	0.516
<u>Fifth Grade</u>				
Reading – List	0.651	0.493	0.613	0.406
Word Analysis Vocabulary Development	0.651	0.493	0.658	0.414
Reading Comprehension	0.501	0.408	0.548	0.381
Literary Response and Analysis	0.654	0.520	0.58	0.51
Written Conventions	0.678	0.513	0.576	0.356
Writing Strategies	0.591	0.547	0.612	0.449
CST ss ELA	0.670	0.561	0.656	0.469
CST pl ELA	0.706	0.558	0.694	0.475

Upper Spelling Inventory. For the Upper Inventory, only one grade (fifth grade) had sufficient sample size to provide stable data for estimation purposes (see Table 14). The predictive validity coefficients for these students ranged from a low of .480 (Writing Strategies) to a high of .647 (Word Analysis). All coefficients attained significance at the $p < .001$ level. The concurrent validity coefficients were somewhat higher, with a low of .464 (Writing Strategies) and a high of .660 (Word Analysis). All coefficients were significant at the $p < .001$ level.

Table 14
Upper Form
Predictive and Concurrent Validity – Two Samples

	Includes all students		Excludes ELL, SPED, Gifted students	
	Predictive Validity	Concurrent Validity	Predictive Validity	Concurrent Validity
<u>Fifth Grade</u>				
Reading	0.638	0.611	0.559	0.544
Word Analysis Vocabulary	0.647	0.660	0.577	0.637
Reading Comprehension	0.617	0.583	0.535	0.467
Literacy Response and Analysis	0.543	0.534	0.463	0.459
Writing Conventions	0.522	0.533	0.5	0.539
Writing Strategies	0.480	0.464	0.355	0.376
CST ss ELA	0.633	0.589	0.519	0.492
CST pl ELA	0.585	0.601	0.525	0.568

Conclusion

The psychometric properties of the spelling inventories that accompany the Words Their Way: Word Study for Phonics, Vocabulary, and Spelling Instruction instructional guides were examined in this study. Based on the data in this study, the inventories are reliable instruments and valid predictors of student achievement. Analysis of the reliability by item discrimination, item difficulty, and internal consistency provide evidence that the instruments are able to differentiate between relatively higher and lower performing students, and do so reliably. Test-retest data shows clear evidence of robust reliability for samples that include students identified as ELL, SPED, and Gifted, and also for samples that do not include these students.

Validity evidence indicates that student scores on the instruments are able to significantly predict achievement on California Standardized Tests four months later. Additionally, student scores on the inventories administered approximately at the same time as the standardized tests show evidence of being significantly related to the standardized test scores. Based on the results from this study, these three instruments appear to be a valuable addition to any educator's toolbox.