BASI Manual Addendum: Growth Scale Value (GSV) Scales and College Report

Growth Scale Value (GSV) Scales

Overview

The BASI GSV scales were created to help users measure the progress of an individual student or a group of students on BASI subtests and composites. GSV scores were chosen for this purpose over other types of scores for several reasons. Although normative scores such as standard scores and percentiles are valuable as indicators of growth, using them for assessing change is complex because the normative reference group also changes over time. Comparing raw scores from different BASI levels or forms would be invalid, and grade equivalents lack the basic scale properties required for making meaningful comparisons.

For a given subtest or composite, a particular GSV score indicates a particular level of skill or ability, regardless of which BASI form or level the student took. It is this feature that makes the GSV scale valuable for measuring growth. GSV scores from any form or level of the subtest or composite may be compared. The numerical values of the GSV scale are arbitrary; they do not have any meaning in themselves. Therefore, GSV scores may only be compared within a subtest or composite. GSV scores from different subtests or composites have no meaningful relationship to each other.

GSV scores have a second benefit in addition to providing a common scale across forms and levels. They are closer than raw scores to an interval scale of measurement. An interval scale is one in which a particular difference in scores designates the same size of difference in the underlying skill or ability regardless of where on the scale it occurs. In other words, a five-point difference in GSV scores at the low end of the scale represents the same difference in ability as a five-point difference at the high end of the scale. As a general rule, raw score scales do not have this feature. Interval scaling is obviously beneficial for measuring growth because it allows the user to interpret change in the same way wherever it occurs on the scale.

Interpretation of GSV Scores

The BASI GSV scales are similar to the GSV (or W) scales on many other tests, such as the Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001), the Kaufman Test of Educational Achievement, Second Edition (K-TEA; Kaufman & Kaufman, 2004), and the Woodcock Reading Mastery Tests (WRMT; Woodcock, 1998). (For a full description of

the W scale, see Woodcock and Dahl, 1971.) In each case, the GSV scale is based on the "ability" scale created by a calibration of a subtest or composite using the Rasch version of item response theory (IRT). When a test is calibrated using IRT, all of the items and all of the examinees are placed on a single scale of difficulty/ability. At the low end of the scale are the easiest items and the people with the lowest levels of ability, and at the high end are the most difficult items and the most able examinees.

Although the numbers on the difficulty/ability scale are arbitrary, the size of the difference between an item's difficulty value and a person's ability score indicates the probability that the person will answer the item correctly. If a person's ability score and an item's difficulty value are numerically the same, the person has a 50% probability of answering the item correctly. On the GSV or W scale, which is a linear transformation of the Rasch scale, a person whose ability is 10 points higher than an item's difficulty has a 75% probability of answering the item correctly. When the difference is 20 points, the probability of answering the item correctly is 90%. (These results are obtained by using a factor of 9.1024 in the linear transformation of Rasch scale values to GSV values.) Thus, the GSV scale can be used to meaningfully interpret differences between people's abilities or items' difficulties in terms of the probability of success on certain tasks.

Construction of the GSV Scales

The GSV scales for the subtests and composites of the BASI Comprehensive Form were constructed using several data sets:

- the 3,345 students who took Comprehensive Form A and the Survey Form in the fall of 2002 (this group includes the Form A norm sample and part of the Survey Form norm sample)
- the 2,130 students who took Comprehensive Form B in the spring of 2003 and who constitute the Form B norm sample
- the alternate-forms sample consisting of 277 students who took Level 1, 2, or 3 of Comprehensive Form A followed by Form B within the same week

These data sets were merged into a single data file for the purpose of performing the Rasch calibrations.

Linkages between forms and levels of the Comprehensive Form were established by both common persons and common items. The alternate-forms sample provided a common-persons linkage between Forms A and B for Levels 1 through 3. There were several sources of common items, including:

- items that appeared on both Form A and Form B
- items that appeared on two or more levels of the same or different forms
- items that appeared on the Survey Form as well as on Form A and/or Form B

Because the Survey Form has a single level that is administered to all grades, it was an important source of across-level linkages for the Comprehensive Form.

The Reading Comprehension subtest presented a special problem with regard to linking across levels because there are no common items across Comprehensive Form levels and few items common to the Comprehensive and Survey Forms. The calibration of this subtest was achieved through the calibration of the Reading Total, which is made up of the Reading Comprehension and Vocabulary subtests. The latter subtest has ample across-level linking items.

If a student took two forms with common items (i.e., Form A and either Form B or the Survey Form), the Form A score for each common item was retained for analysis, and the item score from the other form was blanked out. Examinees who responded to fewer than 10 items on a subtest were excluded from analysis of that subtest. All missing (unanswered) items were scored as incorrect.

Rasch calibration was carried out using the Winsteps computer program (Linacre, 2003). A total of eight

calibrations were conducted, one for each subtest (except Reading Comprehension) and one for each of the three composites (Reading Total, Written Language Total, and Math Total). Each composite calibration included all the items from the two subtests in that composite.

Using the item difficulties from the calibration as anchored values, Winsteps was run again, this time separately for each form and level of each subtest and composite. These analyses generated the raw-score-to-ability conversion tables. Ability scores were then multiplied by 9.1024, and a constant of 100 was added to eliminate negative values. This linear transformation produced the GSV scores corresponding to raw scores on each form and level of each subtest and composite.

Table 1 presents means and standard deviations of GSV scores on each subtest and composite for the Comprehensive Form norm sample by grade and season. Figures 1 through 3 show the shape of the GSV growth curve for each composite, in this case based on the median GSV score at each grade level. Interestingly, the composite score growth curves are very similar not only in shape but also in the span of GSV scores from grade 3 to grade 12. In each instance, performance increases most rapidly at the youngest grades, and the rate of increase declines gradually until it reaches a plateau at about grade 10.

TABLE 1. Means and Standard Deviations of GSV Scores for the BASI Norm Sample by Grade and Season

			Subtest												Composite					
			Vocabulary		Reading Comp.		Spelling		Language Mech.		Math Comp.		Math App.		Reading Total		Written Language Total		Math Total	
Grade	Season	*n*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
3	Fall	319	85.6	11.6	88.8	9.3	85.6	9.8	91.0	8.6	80.4	9.4	87.2	8.2	87.7	9.0	89.3	7.9	83.7	7.0
	Spring	273	88.2	13.2	93.2	13.5	89.6	12.2	94.3	11.1	89.0	12.7	89.1	12.0	91.5	11.6	91.7	10.7	89.1	11.7
4	Fall	301	97.3	13.9	96.4	11.4	93.8	12.1	98.7	11.4	93.5	14.3	95.6	11.1	96.7	10.8	96.9	9.9	94.0	12.0
	Spring	289	95.3	13.1	97.4	13.9	96.6	14.1	98.4	12.3	99.2	15.8	96.7	15.0	96.9	11.4	96.9	12.6	97.7	14.1
5	Fall	227	96.6	6.7	97.8	8.6	100.5	8.3	98.2	7.0	95.3	7.6	93.0	7.6	96.2	6.8	99.0	6.7	94.1	6.8
	Spring	234	101.8	8.1	99.0	8.6	102.2	9.1	99.9	8.8	98.6	9.3	95.9	10.7	99.1	7.2	100.8	8.2	97.4	8.7
6	Fall	224	102.5	8.8	102.2	9.3	106.3	11.6	103.1	10.1	99.8	9.6	96.3	9.1	101.0	8.8	104.0	10.1	97.8	8.4
	Spring	261	104.6	11.4	101.5	11.4	104.3	11.5	102.3	11.2	99.7	10.1	99.4	10.7	101.5	11.1	102.8	11.0	99.6	8.7
7	Fall	227	107.4	8.3	104.6	11.8	110.7	9.3	107.6	11.1	106.6	8.3	101.7	8.4	104.6	8.7	108.4	8.9	103.9	7.1
	Spring	209	106.7	9.9	104.6	11.4	112.3	9.3	106.1	11.1	106.1	8.8	100.6	9.2	104.0	9.6	108.7	8.3	103.0	8.5
8	Fall	253	109.0	8.7	108.1	10.7	112.2	10.9	108.5	11.0	108.7	10.6	102.9	9.8	106.9	8.8	109.6	9.3	105.4	10.0
	Spring	207	110.6	7.9	109.3	8.7	113.2	11.7	108.7	9.8	108.2	10.1	103.1	8.6	108.3	7.5	110.4	9.0	105.2	7.6
9	Fall	217	110.6	8.4	107.1	9.7	113.5	11.0	106.6	10.7	108.2	10.6	100.8	10.7	107.4	6.9	109.4	9.0	104.4	8.7
	Spring	154	110.9	10.9	109.2	12.5	111.0	9.5	106.7	12.6	107.8	13.6	101.8	11.1	108.5	10.4	108.6	10.0	104.4	11.6
10	Fall	225	112.0	10.7	108.7	9.1	117.0	10.4	108.0	11.7	110.8	12.1	102.6	8.2	108.6	8.5	111.7	9.2	106.0	9.2
	Spring	135	111.0	10.3	108.8	10.8	112.7	8.0	110.2	11.3	108.2	11.4	101.9	8.4	108.2	9.4	110.9	8.2	104.6	9.4
11	Fall	227	113.8	11.1	108.3	11.2	118.2	9.2	105.8	16.8	112.3	12.2	103.3	7.9	109.2	9.3	111.7	9.6	107.2	8.4
	Spring	204	112.5	13.3	109.7	13.8	113.7	12.0	108.9	14.5	112.5	11.8	105.1	8.7	109.2	13.1	110.6	13.0	108.3	9.1
12	Fall	219	112.9	10.8	107.7	10.1	117.1	10.5	104.9	15.9	109.4	12.5	103.0	8.0	108.6	9.0	110.7	9.8	105.8	8.4
	Spring	158	113.2	14.2	108.2	14.6	113.1	12.4	106.7	14.3	108.7	12.7	103.0	9.8	109.5	11.9	109.4	13.1	105.6	10.4

FIGURE 1. Median GSV by Grade: Reading Total

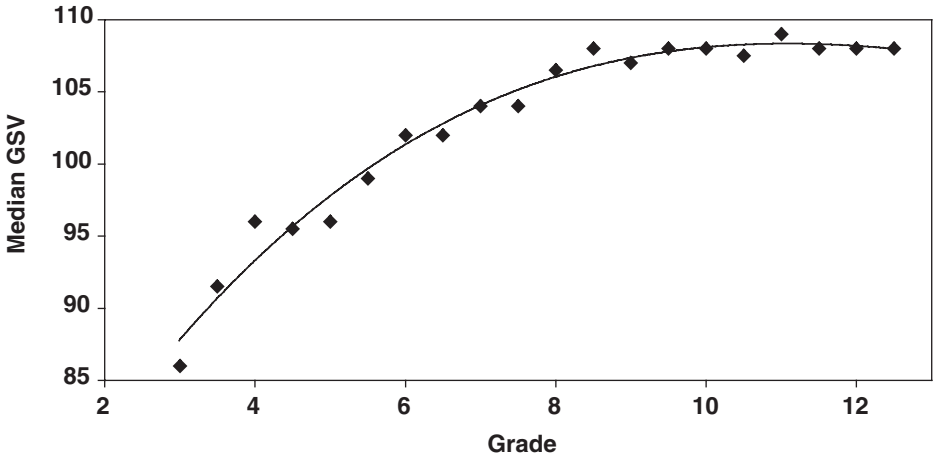


FIGURE 2. Median GSV by Grade: Written Language Total

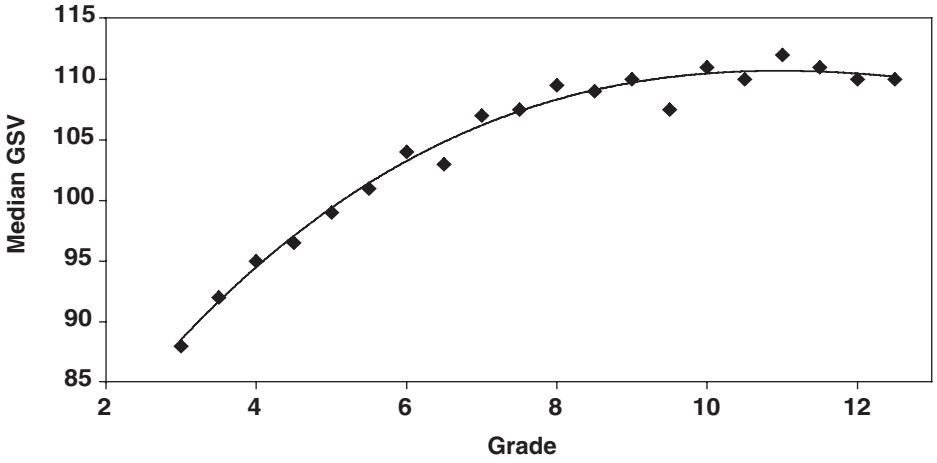
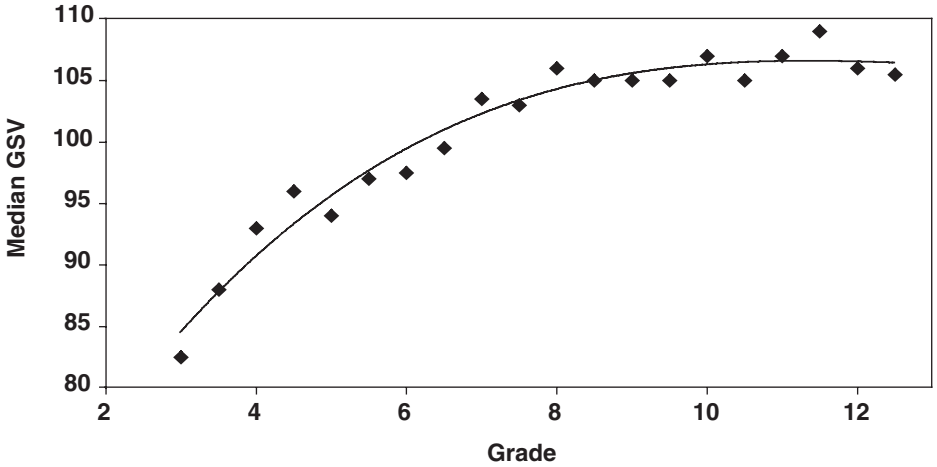


FIGURE 3. Median GSV by Grade: Math Total



Using GSV Scores to Measure Growth

As noted previously, the primary rationale for the GSV scale is to support the measurement of change over time in the performance of a student or group. Because the GSV scale reflects absolute performance (rather than relative position with respect to a norm group), any positive or negative change in the GSV score indicates an increase or decrease in measured achievement on that subtest or composite. However, users will often need additional information to help them interpret an observed GSV score change, addressing such questions as whether the change is statistically significant (i.e., unlikely to be solely the result of measurement error) and how the amount of change compares to the typical growth shown by students at that grade level over that period of time.

Table 2 shows the size of difference between GSV scores required for statistical significance at the .10 level (two-tailed) for each subtest and composite at each grade. (This table applies to individual GSV scores, not group averages. For determining the statistical significance of a difference between group means, a standard *t* test may be used.) If a student's GSV score has changed by as much as or more than the value shown in Table 2, there is less than a 10% chance that it is the result of measurement error. To create this table, the standard deviation of GSV scores (shown in Table 1) and the internal-consistency reliability at each grade and season (obtained from the BASI Comprehensive Form manual) were used to calculate the standard error of measurement in GSV units. This value was multiplied by 1.64 to produce the critical value for a .10 significance level. Finally, the two critical values within a grade were averaged.

TABLE 2. Statistically Significant GSV Differences ($p < .10$) by Grade

	Subtest						Composite		
Grade	Vocab.	Reading Comp.	Spelling	Lang. Mech.	Math Comp.	Math App.	Reading Total	Written Language Total	Math Total
3	7	7	7	7	8	9	5	5	6
4	9	7	7	8	9	10	5	5	7
5	7	7	8	7	9	8	5	5	6
6	8	8	9	9	10	9	6	7	7
7	8	9	8	8	8	9	6	6	6
8	9	9	9	9	11	10	6	6	8
9	7	7	8	8	9	11	4	5	6
10	7	6	8	8	10	9	5	5	7
11	7	7	7	10	9	8	5	5	6
12	7	7	7	9	9	9	4	5	6

Table 3 shows the average amount of change in GSV scores in the norm sample for each subtest and composite over four different time intervals (9, 18, or 27 weeks, or 1 year). These results are derived from the smoothed rates of growth such as those depicted in Figures 1 through 3. When interpreting an increase in the GSV score of an individual or in the average score of a group, users may find it helpful to compare that growth to the amount that would be expected over that period of time for individuals in the general population.

TABLE 3. Average GSV Growth by Grade and Time Interval

	Interval: 9 Weeks								
	Subtest						Composite		
Grade	Vocabulary	Reading Comp.	Spelling	Language Mech.	Math Comp.	Math App.	Reading Total	Written Language Total	Math Total
3	1.4	0.7	1.3	0.8	1.5	0.7	1.2	1.2	1.3
4	1.3	0.7	1.2	0.7	1.2	0.7	0.9	0.9	1.0
5	1.0	0.6	1.0	0.5	0.9	0.6	0.7	0.8	0.7
6	0.6	0.5	0.8	0.5	0.7	0.4	0.5	0.6	0.4
7	0.4	0.5	0.6	0.3	0.5	0.4	0.5	0.5	0.4
8	0.3	0.3	0.3	0.2	0.3	0.3	0.3	0.3	0.3
9	0.2	0.2	0.3	0.2	0.2	0.2	0.2	0.2	0.2
10	0.2	0.1	0.3	0.1	0.2	0.1	0.1	0.1	0.2
11	0.1	0.1	0.2	0.0	0.2	0.1	0.1	0.1	0.1
12	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.1	0.1

	Interval: 18 Weeks								
	Subtest						Composite		
Grade	Vocabulary	Reading Comp.	Spelling	Language Mech.	Math Comp.	Math App.	Reading Total	Written Language Total	Math Total
3	2.8	1.4	2.6	1.6	2.9	1.4	2.4	2.4	2.6
4	2.6	1.4	2.4	1.4	2.4	1.4	1.7	1.7	2.1
5	2.1	1.2	2.1	1.0	1.7	1.2	1.4	1.6	1.4
6	1.2	1.0	1.6	1.0	1.4	0.9	1.0	1.2	0.9
7	0.9	1.0	1.2	0.7	1.0	0.9	1.0	1.0	0.9
8	0.7	0.7	0.7	0.3	0.7	0.5	0.5	0.5	0.5
9	0.3	0.3	0.5	0.3	0.3	0.3	0.3	0.3	0.3
10	0.3	0.2	0.5	0.2	0.3	0.2	0.2	0.2	0.3
11	0.2	0.2	0.3	0.0	0.3	0.2	0.2	0.2	0.2
12	0.2	0.2	0.2	0.0	0.2	0.2	0.2	0.2	0.2

TABLE 3. Average GSV Growth by Grade and Time Interval (continued)

Interval: 27 Weeks

	Subtest						Composite		
Grade	Vocabulary	Reading Comp.	Spelling	Language Mech.	Math Comp.	Math App.	Reading Total	Written Language Total	Math Total
3	4.2	2.1	3.9	2.3	4.4	2.1	3.6	3.6	3.9
4	3.9	2.1	3.6	2.1	3.6	2.1	2.6	2.6	3.1
5	3.1	1.8	3.1	1.6	2.6	1.8	2.1	2.3	2.1
6	1.8	1.6	2.3	1.6	2.1	1.3	1.6	1.8	1.3
7	1.3	1.6	1.8	1.0	1.6	1.3	1.6	1.6	1.3
8	1.0	1.0	1.0	0.5	1.0	0.8	0.8	0.8	0.8
9	0.5	0.5	0.8	0.5	0.5	0.5	0.5	0.5	0.5
10	0.5	0.3	0.8	0.3	0.5	0.3	0.3	0.3	0.5
11	0.3	0.3	0.5	0.0	0.5	0.3	0.3	0.3	0.3
12	0.3	0.3	0.3	0.0	0.3	0.3	0.3	0.3	0.3

Interval: 1 Year

	Subtest						Composite		
Grade	Vocabulary	Reading Comp.	Spelling	Language Mech.	Math Comp.	Math App.	Reading Total	Written Language Total	Math Total
3	8.0	4.0	7.5	4.5	8.5	4.0	7.0	7.0	7.5
4	7.5	4.0	7.0	4.0	7.0	4.0	5.0	5.0	6.0
5	6.0	3.5	6.0	3.0	5.0	3.5	4.0	4.5	4.0
6	3.5	3.0	4.5	3.0	4.0	2.5	3.0	3.5	2.5
7	2.5	3.0	3.5	2.0	3.0	2.5	3.0	3.0	2.5
8	2.0	2.0	2.0	1.0	2.0	1.5	1.5	1.5	1.5
9	1.0	1.0	1.5	1.0	1.0	1.0	1.0	1.0	1.0
10	1.0	0.5	1.5	0.5	1.0	0.5	0.5	0.5	1.0
11	0.5	0.5	1.0	0.0	1.0	0.5	0.5	0.5	0.5
12	0.5	0.5	0.5	0.0	0.5	0.5	0.5	0.5	0.5

College Reference Scores

College reference scores are provided to enable students, parents, and counselors to compare a student's achievement scores to those obtained by students enrolled in two-year or four-year colleges or universities. These scores serve an informational purpose and are not designed to predict academic success in postsecondary schooling. Each of the two reference groups (two-year and four-year schools) consists of a diverse set of institutions, but the number of schools in each group is small and the reference groups are not necessarily representative of the general population of college and university students. Nevertheless, for secondary students or graduates who have taken the BASI, these reference scores may provide a helpful benchmark for the level of academic achievement in various content areas of students in community colleges, technical schools, and four-year colleges and universities.

GSV scores are well suited to this application. One reason is that GSV scores are independent of the form and level taken by the student. That is, a student would be expected to obtain the same GSV score on a subtest or composite regardless of which BASI form or level he or she took. A second advantage is that GSV scores, like ACT® or SAT® scores, are not norm-referenced. A student in any grade or of any age may compare his or her GSV score to the college reference values with the knowledge that the comparison is one of absolute performance rather than relative standing within a group. The use of norm-referenced scores for this purpose would be confusing because there are two reference groups (the grade-based or age-based norm sample and the college reference sample).

Samples

Eleven institutions (listed in Table 4) agreed to take part in this study, which was conducted between 2002 and 2004. Eight of the 11 schools were public, and 3 were private. Of the 672 students who participated, 206 attended two-year schools and 466 attended four-year schools. The demographic characteristics of these samples are reported in Table 5.

The four-year college sample includes students in all four years of school because analysis of their BASI scores revealed average scores to be similar. That is, there was no tendency for average scores to increase as students got older. By contrast, a sample of 62 students in their second year of a two-year program scored higher than first-year students on all BASI subtests and composites. (The higher scores of second-year students could be due to learning during the first year or to attrition.) It was decided to exclude the second-year students from the two-year college reference group based on the assumption that individuals would be most interested in comparing their BASI scores to those of entering students.

Analysis

GSV scores were obtained for each BASI subtest and composite. For each of the two samples, the median, 25th percentile, and 75th percentile GSV scores were computed. These form the midpoint and the upper and lower endpoints of the box plot displayed on the BASI College Report. The box represents the middle 50% of scores for college students. The horizontal line to the right of the box shows the range of the upper 25% of scores, and the horizontal line to the left of the box shows the range of the lower 25% of scores. The diamond in each plot represents the student's GSV score in comparison to the college students' scores. Table 6 shows all of these values for each sample (two-year and four-year schools) for each BASI subtest and composite.

ACT is a registered trademark of ACT, Inc. SAT is a registered trademark of The College Board.

TABLE 4. Participating Institutions

Two-Year Colleges and Technical Schools

Johnston Community College, Smithfield, NC
Minneapolis Community College, Minneapolis, MN
Palo Verde College, Blythe, CA
Southwest Florida Community College, Fort Meyers, FL

Four-Year Colleges

The Citadel, Charleston, SC
Miami University, Oxford, OH
Ohio State University, Columbus, OH
State University of New York, Oswego, NY
University of Hartford, Hartford, CT
University of Houston–Victoria, Houston, TX
University of Maryland, College Park, MD

TABLE 5. Demographic Characteristics of the College Samples

		School Type	
		Two-Year	**Four-Year**
Gender	Female	68%	52%
	Male	32%	48%
Race/Ethnicity	African American	21%	4%
	Hispanic	17%	5%
	White	58%	85%
	Other	4%	6%
Year in College	First	206	157
	Second	—	116
	Third	—	100
	Fourth	—	93
n		206	466

TABLE 6. College Reference GSV Scores

	Percentiles									
	Two-Year Colleges					Four-Year Colleges				
Subtest or Composite	Low	25	50	75	High	Low	25	50	75	High
Vocabulary	86	107	112	121	142	100	118	124	130	148
Reading Comprehension	80	102	111	117	140	100	113	118	122	136
Spelling	94	112	118	124	142	101	120	125	133	153
Language Mechanics	87	102	109	112	127	96	111	116	121	136
Math Computation	87	100	106	109	123	90	112	118	127	150
Math Application	88	98	102	105	116	91	104	109	113	127
Reading Total	85	104	110	117	137	101	114	118	123	137
Written Language Total	92	107	111	117	132	102	115	119	124	138
Math Total	91	100	103	106	115	97	109	112	117	129

References

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Test of Educational Achievement* (2nd ed.). Minneapolis, MN: NCS Pearson, Inc.

Linacre, M. J. (2003). Winsteps Rasch measurement software [Computer software]. Chicago: Winsteps.

Williams, K. (2001). *Group Reading Assessment and Diagnostic Evaluation.* Minneapolis, MN: NCS Pearson, Inc.

Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests, Revised–Normative update.* Minneapolis, MN: NCS Pearson, Inc.

Woodcock, R. W., & Dahl, M. N. (1971). A common scale for the measurement of person ability and test item difficulty (AGS Paper No. 10). Minneapolis, MN: NCS Pearson, Inc.