

Reliability and Standard Error of Measurement

Tables 5A, 5B, 5C, and 5D provide reliability estimates for GRE General Test, GRE Subject Tests, GRE Biology Test subscores, and GRE Psychology Test subscores, respectively. Reliability indicates the degree to which individual test takers would keep the same relative standing if the test were administered more than once to each test taker. The reliability index ranges from zero to one; a reliability index of one indicates that there is no measurement error in the test and therefore the test is perfectly reliable.

The Verbal Reasoning and Quantitative Reasoning measures of GRE General Test are intended to have reliabilities of at least .90. The reliability of the Analytical Writing measure is similar to the reliability for other writing measures where the reported score is based on a test taker’s performance on two tasks. Reliability is influenced by the consistency of the ratings assigned to each essay. Overall, the two ratings used in each essay score are in agreement about 85 percent of the time; they differ by one score point about 14 percent of the time; and they differ by two or more score points about one percent of the time.

The Subject Tests are intended to have reliabilities of at least .90 for the total test scores. For each of the Subject Tests, the reliability coefficient of the total scores is at least .90, and the reliability coefficient of the subscores is at least .80.

Tables 5A, 5B, 5C, and 5D also provide data on the standard errors of measurement (SEM) and SEM of score differences. SEM is an index of the variation in scores to be expected due to errors in measurement. For a group of test takers, it is an estimate of the average difference between observed scores and “true” scores (i.e., what test takers’ scores on a test would hypothetically be if there was no measurement error). Approximately 95 percent of test takers will have obtained scores that are within a range extending from two standard errors below to two standard errors above their true scores.

The SEM of score differences is an index used to determine whether the difference between two scores is meaningful. Small differences in scores may be due to measurement error and not to real differences in the abilities of the test takers. This index incorporates the error of measurement in each score being compared. To use the SEM of score differences, multiply the value by 2. Score differences exceeding this value are likely to reflect real differences in ability at approximately a 95 percent confidence level.

Table 5A: Reliability Estimates and Standard Errors of Measurement (SEM)^a for Individual Scores and Score Differences for the GRE[®] General Test

Score	Reliability Estimate	SEM of Individual Scores	SEM of Score Differences
Verbal Reasoning	0.92	2.4	3.4
Quantitative Reasoning	0.95	2.1	3.0
Analytical Writing	0.85	0.33	0.46

^aThe reliability estimates and SEMs for the computer-delivered Verbal Reasoning and Quantitative Reasoning measures of the General Test are based on item response theory (IRT). The reported values are an average of all the estimates obtained for all the multi-stage tests delivered between July 1, 2014, and June 30, 2017. The reliability estimates and SEMs for the computer-delivered Analytical Writing measure are computed based on split-half analyses using the performance of all individuals who tested between July 1, 2014, and June 30, 2017. The reliability estimates for the paper-delivered version of the measures are very close to those reported for the computer-delivered version.

Table 5B: Reliability Estimates and Standard Errors of Measurement (SEM)^a for Individual Scores and Score Differences for GRE[®] Subject Tests

Score	Reliability Estimate	SEM of Individual Scores	SEM of Score Differences	Sample Size
Biology Test	0.95	27	38	319
Chemistry Test	0.94	25	35	1,156
Literature in English Test	0.95	21	30	612
Mathematics Test	0.91	41	58	1,059
Physics Test	0.94	36	51	1,535
Psychology Test	0.96	22	32	917

^aThe reliability for all the Subject Tests scores are estimated using the Kuder-Richardson formula (KR-20). The reported reliability, SEM, and sample size values are based on a test edition that is representative of recent test editions.

Table 5C: Reliability Estimates and Standard Errors of Measurement (SEM)^a for Individual Scores and Score Differences for GRE[®] Biology Test Subscores

Subscore	Reliability Estimate	SEM of Individual Scores	SEM of Score Differences	Sample Size
Cellular and Molecular Biology	0.89	3.9	5.4	319
Organismal Biology	0.86	4.3	6.1	319
Ecology and Evolution	0.84	4.6	6.5	319

^aThe reliability for the Biology Test subscores are estimated using the Kuder-Richardson formula (KR-20). The reported reliability, SEM, and sample size values are based on a test edition that is representative of recent test editions.

Table 5D: Reliability Estimates and Standard Errors of Measurement (SEM)^a for Individual Scores and Score Differences for GRE[®] Psychology Test Subscores

Subscore	Reliability Estimate	SEM of Individual Scores	SEM of Score Differences	Sample Size
Biological	0.89	3.6	5.1	917
Cognitive	0.94	2.5	3.6	917
Social	0.91	3.3	4.6	917
Developmental	0.87	3.9	5.5	917
Clinical	0.92	3.1	4.4	917
Measurement/Methodology/Other	0.92	3.1	4.4	917

^aThe reliability of the Psychology subscores are estimated as the proportional reduction in mean square error (a subscore reliability statistic that provides estimates comparable to Cronbach's alpha). The reported reliability, SEM, and sample size values are based on a test edition that is representative of recent test editions.

Conditional Standard Errors of Measurement for the *GRE*[®] Verbal Reasoning and Quantitative Reasoning Measures

Tables 5E and 5F contain estimates of the conditional standard errors of measurement (CSEM) at selected reported scores for the GRE Verbal Reasoning and Quantitative Reasoning measures. While the SEMs presented in Table 5A address the average measurement precision of the test, the measurement precision actually varies across the score scale. The CSEM reflects this variation by indicating the amount of error in a reported score at a given point on the scale. Like the SEM, the CSEM can be used to compute a confidence band around a test taker’s score. Such a band would help to determine the score range in which the test taker’s “true” score probably lies. Unlike the SEM, the CSEM takes the variation in measurement precision across the score scale into account.

The CSEM of score differences incorporates the measurement error in each score. The CSEM of score differences should be used when comparing the scores of two individuals because small differences in scores may not represent real differences in the abilities of the two individuals. To use the CSEM of score differences, take the larger of the two values and multiply by 2. Score differences exceeding this value are likely to reflect real differences in ability at approximately a 95 percent confidence level.

Table 5E: Conditional Standard Errors of Measurement (CSEM) of Individual Scores at Selected Scores for the *GRE*[®] Verbal Reasoning and Quantitative Reasoning Measures^a

Measure	130	135	140	145	150	155	160	165	170
Verbal Reasoning	4.1	3.6	2.9	2.5	2.2	2.1	2.1	2.0	1.4
Quantitative Reasoning	3.6	2.8	2.3	2.2	2.1	2.0	2.0	2.1	1.0

Table 5F: Conditional Standard Errors of Measurement (CSEM) of Score Differences at Selected Scores for the *GRE*[®] Verbal Reasoning and Quantitative Reasoning Measures^a

Measure	130	135	140	145	150	155	160	165	170
Verbal Reasoning	5.9	5.1	4.1	3.5	3.1	3.0	2.9	2.9	2.0
Quantitative Reasoning	5.0	4.0	3.3	3.0	2.9	2.8	2.9	3.0	1.5

^aThe CSEM of individual scores and CSEM of score differences for the Verbal Reasoning and Quantitative Reasoning measures of the computer-delivered GRE General Test are based on item response theory (IRT). The reported values are an average of all the estimates obtained for all the multi-stage tests delivered between July 1, 2014, and June 30, 2017. The CSEM of individual scores and CSEM of score differences are not available for the Analytical Writing measure.