Reliability and Standard Error of Measurement

Tables 4A and 4B provide reliability estimates for the GRE General Test and GRE Subject Tests, 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 reliability of the Analytical Writing measure is influenced by the consistency of the ratings assigned to the Issue essay task. Overall, the two ratings are in agreement about 98 percent of the time; they differ by one score point about 2 percent of the time; and they differ by two or more score points less than one percent of the time.

Tables 4A and 4B 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 4A: 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	0.87	3.2	4.5
Quantitative Reasoning	0.93	2.6	3.7
Analytical Writing	0.76	0.43	0.61

[a] The reliability estimates and SEMs for the 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 September 2023 and May 2024 to reflect the reliability of the shortened GRE. The reliability estimates and SEMs for the Analytical Writing measure are computed based on test-retest analyses using the performance on the Issue task only of all repeaters who tested between July 1, 2020, and June 30, 2023.

Table 4B: 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
Mathematics Test	0.92	43	61	1,219
Physics Test	0.93	44	62	664
Psychology Test	0.94	26	37	564

[a] The 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 between September 2023 and April 2024.

Conditional Standard Errors of Measurement for the GRE Verbal Reasoning and Quantitative Reasoning Measures

Tables 4C and 4D 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 4A 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 individual scores 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 4C: 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	4.3	3.9	3.3	2.9	2.7	2.5	2.5	2.6	1.7
Quantitative Reasoning	4.2	3.6	3.0	2.8	2.7	2.6	2.7	2.6	1.2

[a] The CSEM of individual scores and CSEM of score differences for the Verbal Reasoning and Quantitative Reasoning measures of the 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 September 2023 and May, 2024. The CSEM of individual scores and CSEM of score differences are not available for the Analytical Writing measure.

Table 4D: 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	6.1	5.5	4.7	4.1	3.8	3.6	3.5	3.6	2.4
Quantitative Reasoning	5.9	5.1	4.3	3.9	3.8	3.6	3.8	3.6	1.7

[a] The CSEM of individual scores and CSEM of score differences for the Verbal Reasoning and Quantitative Reasoning measures of the 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 September 2023 and May 2024. The CSEM of individual scores and CSEM of score differences are not available for the Analytical Writing measure.