

TrackTest English test-retest reliability- Cronbach’s Alpha and ICC for the test with the dynamically generated questions

Institution: TEK Kampüs Okulları, a public school in Arsuz, Hatay, Turkey.

Participants: 306 students- 5th-10th graders (11-17 years old) which participated in all 3 sessions, and achieved a relevant result (above 30%), 918 tests in total.

Dates: 3 assessment sessions in 2018- February 16, March 23, April 27.



The test-retest reliability measurement or repeatability is a method for testing the stability and reliability of an assessment instrument over time. When same students repeat the test more times, they should get the questions of the same difficulty and achieve similar results. This is quantifiable by statistical methods with the reliability coefficient on a scale from 0 (no correlation) to 1 (best). A reliability coefficient of .70 or higher is considered “acceptable” in most social science research situations. The alpha coefficient above .80 means that items have high consistency.

For the purpose of this analysis, we have used the standard English core test results provided by TrackTest for one public school in Turkey.

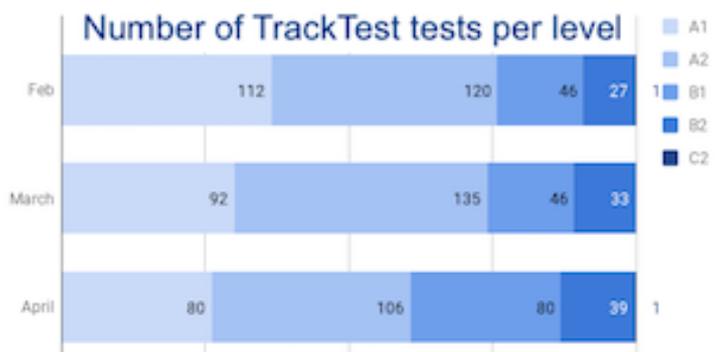
English assessment settings:

TrackTest English test was set in “random mode”, i.e. each student during each assessment session got a **different set of questions** and tasks of the same difficulty. Tests were taken on tablets in the proctored environment organized by PlusEd Turkey.

Students were allowed to progress to a higher level in the next assessment. For instance, they completed successfully the A1 level test in the first assessment and took the A2 level test in the second assessment.

This would usually cause a major problem when making a comparison across the levels. For this reason, the TrackTest system provides also the

internal TrackTest Score (TTS) which automatically recalculates the level test percentage score results in the universal TTS points. These TTS scores (0-1200 points) were used for this test-retest reliability analysis.



We had to take into account also the usual biases in the test-retest-retest scenarios, e.g. time difference between tests during which students attended the English language tuition in their school as well as the novelty factor of the first test vs. repeated tests and carryover effect.

However, five weeks gap between the tests is still quite short for the significant improvement in English language and the retests were provided with a different set of questions. Therefore there might be some degree of error but there was no sign of intervening factors which would compromise the overall test-retest reliability.

Two test reliability coefficients were calculated, a standard Cronbach's Alpha () and Intraclass Correlation (ICC). Since the variance component equalled 0, the result for both was the same.



Test reliability results:

Cronbach's Alpha: 0.816 High

Cronbach's Alpha Based on Standardized Items: 0.817

N of Items: 3

Intraclass Correlation-ICC (Average Measures): 0.816

Two-way mixed model, type: Consistency

95% Confidence Interval (lower-upper bound): 0.777-.0849