

EXHIBIT 7.9 High Coefficient Alpha but Low Agreement

Skill	Rater 1	Rater 2	Rater 3
Skill A	1	3	4
Skill B	2	4	5
Skill C	2	4	5
Skill D	1	3	4
Skill E	2	4	5

EXHIBIT 7.10 Low Coefficient Alpha but High Agreement

Skill	Rater 1	Rater 2	Rater 3
Skill A	2	2	2
Skill B	2	2	2
Skill C	3	3	3
Skill D	2	2	2
Skill E	2	2	2

Why? Remember that the strength of the correlation coefficient depends on how much variation there is in the items (in this case, skills). Since there is no variation in the items (e.g., skill A is rated 2 by all three raters, as is skill B), the average correlation among the items is 0 and coefficient alpha is 0.

When assessing the reliability of scores where multiple raters are assessing candidates on multiple items, keep in mind that coefficient alpha and interrater agreement are tapping into different aspects of reliability. Therefore, it is important to consider both.

Test-Retest Reliability. To assess test-retest reliability, the test scores from two different time periods are correlated through calculation of the correlation coefficient. The r may be calculated on total test scores, or a separate r may be calculated for scores on each item. The resultant r indicates the stability of measurement—the higher the r , the more stable the measure.

Interpretation of the r value is made difficult by the fact that the scores are gathered at two different points in time. Between those two time points, the attribute being measured has an opportunity to change. Interpretation of test-retest reliability thus requires some sense of how much the attribute may be expected to change,