

## MobyAssessments: Psychometric Analysis

The reliability and precision of MobyAssessments are supported by comprehensive psychometric analyses. Psychometrics is a field of scientific study concerned with the measurement of mental capacities, including student knowledge and skills. MobyMax utilizes the statistical techniques developed by psychometric researchers to provide an objective method for evaluating and refining Moby Assessments. This process is carried out at two levels: at the *item level* and at the *assessment level*.

### Item Analysis

Each question, or item, in a Moby Assessment is evaluated statistically to ensure that it is fair, effective, and reliable. Several measures are used in this procedure.

- *Item discrimination*: The purpose of an item is to differentiate students based on their level of proficiency. This relationship, between performance on an item and overall proficiency, can be quantified using the point biserial correlation [1]. The point biserial correlation measures the degree to which performance on an item corresponds to performance on the assessment as a whole, and is represented as a value between -1 and 1. Items with higher point biserial correlations are better at discriminating between high and low performing students. Per psychometric guidelines, items with a point biserial correlation<sup>1</sup> above .3 are considered ‘good’, and items falling below that threshold are subject to revision [2].
- *Item difficulty*: It is important that items are rigorous, but also clear and appropriate. When they do not meet these standards, they are often either too easy or too difficult. The difficulty of an item is measured as the proportion of students who answer it correctly. Difficulty scores can therefore range between 0 and 1, with lower scores corresponding to more difficult items. Items with difficulty scores outside the recommended range of 0.2 to 0.8 are examined and improved in order to maintain optimal variability in student performance [3].
- *Distractor analysis*: For multiple-choice items, incorrect answer choices should be unambiguously incorrect, while at the same time maintaining plausibility [4]. To examine the effectiveness of incorrect answer choices, or distractors, the distribution of student responses across all answer choices is evaluated. Distractors that are infrequently chosen, and distractors that are too frequently chosen, are improved or replaced.

---

<sup>1</sup> When computing the point biserial correlation, a statistical correction is made which excludes a student’s score on an item from the student’s total score. This ‘corrected point biserial correlation’ is a more robust and reliable measure [7].

## Assessment Analysis

Each MobyMax Assessment as a whole is statistically evaluated to guarantee that it reliably and precisely assesses student ability. While many statistics are utilized when evaluating the Moby Assessments, the two key statistics for measuring reliability and precision are *coefficient alpha* and *standard error of measurement*.

- *Coefficient Alpha*<sup>2</sup>: For an assessment to be valid, it must be reliable. The most common way of evaluating the reliability of an assessment is to measure its *internal consistency* by examining the intercorrelations between items [6]. Coefficient alpha quantifies this internal consistency as a value between 0 and 1, with values above 0.8 indicating ‘good’ reliability [7]. All Moby Assessments have a coefficient alpha higher than 0.8.
- *Standard Error of Measurement (SEM)*: While it is crucial that an assessment is reliable, it should also be precise. A precise assessment is one with a small range of error. As such, the precision of an assessment can be evaluated by measuring the size of that range. SEM measures of the range of error, with smaller values indicating greater precision. Further, SEM is calculated in the same scale as the assessment results, thus allowing the SEM to be interpreted directly in conjunction with assessment scores. For example, an SEM of 3 would indicate that a student’s score on an assessment is likely within  $\pm 3$  points of their ‘true’ score. While there can be no fixed criteria for determining ‘good’ SEM scores across assessments, because criteria will depend on the scale, structure, and purpose of the assessment, the SEM of all Moby Assessments meet high standards of precision.

## Data Collection

The psychometric analysis of Moby Assessments is based on student response data collected across the nation, at every grade level, thus ensuring that the analyses are based on large and highly representative samples of students. Further, the process of data collection and analysis is continual, allowing for constant monitoring and refinement of all current and upcoming Moby Assessments.

## Conclusion

By utilizing the psychometric analyses described above, MobyMax is able to ensure the reliability and precision of each test included in the Moby Assessments suite at both the item and assessment level.

---

<sup>2</sup> There are multiple ways of calculating coefficient alpha. Because items in Moby Assessments are scored dichotomously – as correct or incorrect – coefficient alpha is calculated using Kuder-Richardson KR-20, which is a special case of Cronbach’s alpha [5].

## References

- [1] Hays, W. L. (1994). *Statistics*, 5th Edition. Florida: Harcourt Brace & Company.
- [2] Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill
- [3] Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., & Hagen, E. P. (1991). *Measurement and evaluation in psychology and education* (5thEd.). New York: MacMillan.
- [4] Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth: Holt, Rinehart, and Winston.
- [5] Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15, 309–334.
- [6] Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth: Holt, Rinehart, and Winston.
- [7] LeBlanc, Vincent, Cox, Michael A. A. (2017) interpretation of the point-biserial correlation coefficient in the context of a school examination, *The Quantitative Methods for Psychology*, 13(1), 46-56.
- [8] Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.