



Watson-Glaser Validity Evidence

Alice Keane, CPsychol

A decorative graphic in the bottom right corner consisting of a cluster of overlapping circles in various shades of teal, yellow, and dark blue. A large, light blue circle is centered within this cluster, containing the text 'MORE INSIGHT MORE IMPACT™'.

MORE INSIGHT
MORE IMPACT™

“Validity is high if a test gives the information the decision maker needs”

(Cronbach, 1970)

One of the primary reasons psychometric tests are used is to predict a test taker’s potential for future success.

Criterion-related validity evidence is when a statistical relationship exists between scores on the test and one or more criteria such as job performance, supervisor ratings, training course grades etc. By collecting test scores and criterion scores, one can determine how much confidence may be placed on test scores in predicting outcomes of interest such as job success.

Cronbach (1970) characterised criterion-related validity coefficients (‘r’) of .30 or above as having “definite practical value” and The U.S. Department of Labor (1999) provides the following general guidelines for interpreting validity coefficients:

values above .35	=	very beneficial
.21-.35	=	likely to be useful
.11-.20	=	depends on the circumstances
below .11	=	unlikely to be useful

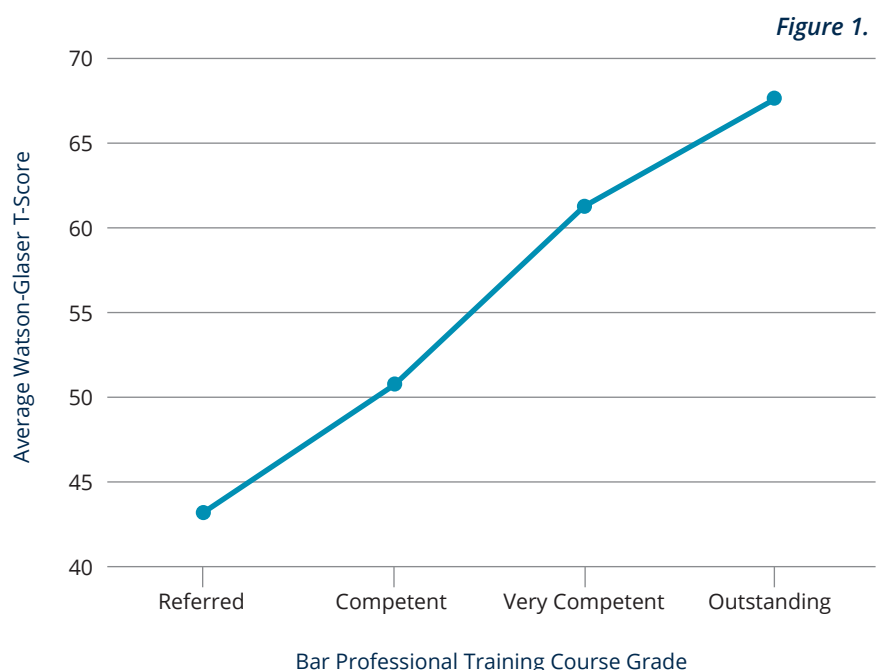
Bar Standards Board (2013)

A correlation of **0.62** was found between scores on the Watson-Glaser and average final exam grade, in a sample of 123 trainee Barristers. This is a very high correlation coefficient, suggesting a strong link between barrister training success and the Watson-Glaser. The final grade included written exams and ratings on vocational exercises such as writing opinions and arguing a case. Figure 1 below illustrates the average test score (T Score) for each category of student on the course.

In a further study sampling 988 participants, a correlation between average final exam grade and scores on items from the Watson-Glaser Unsupervised of **0.51** was found.

Furthermore, the Watson-Glaser was more predictive than A level points, degree class and whether the student attended a Russell Group university.

These studies provide strong evidence in support of the tool’s usage in the law industry, in particular for barrister training selection. Following on from this, anyone wishing to train as a barrister is required to complete the Bar Course Aptitude Test, which is composed of Watson-Glaser items.



Major Law Firm

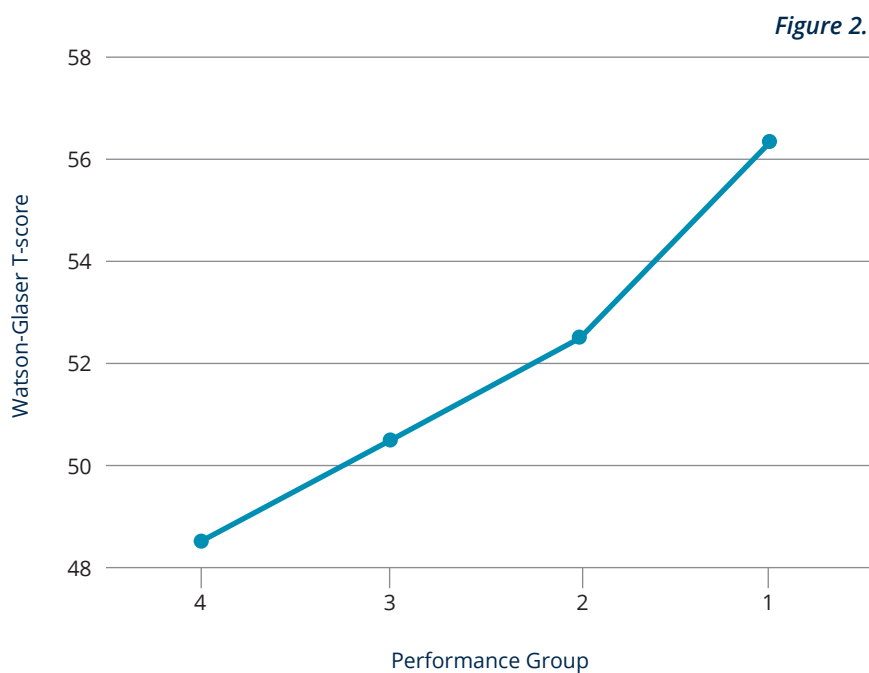
A study in 2013 set out to examine the relationship between scores on the Watson-Glaser test of Critical Thinking Ability obtained by graduate employees within a major law firm at recruitment and subsequent performance over two years in their roles.

As a result of this data demonstrating the predictive nature of Watson-Glaser, the firm have continued to use the test as a major part of their sifting-out and selecting in stages of recruitment.

Data from 250 graduate employees was examined and a summary of the main findings is as follows:

- The Watson-Glaser (written paper and pencil version) was taken under supervision during the assessment process.
- On the whole, the graduate employees had a high level of Watson-Glaser scores compared to the general population and other private sector graduates

- The employees completed four six-month placements within the business and their performance was measured and rated at the end of each placement. Ratings ranged from Level 1 (exceeds expectations) to Level 4 (meeting some expectations, but underperforming in some areas).
- Scores on the Watson-Glaser were found to be predictive of task performance in the role, with a correlation of **0.44** (see Figure 2). Analysis of those employees with consistent performance grades over the two year period showed that the consistent top performers (level 1) achieved the highest average Watson-Glaser score (at recruitment). The next highest score was achieved by those performing at Level 2, then Level 3, then Level 4 (consistent bottom performers). However, there were some small sample sizes in the very top and bottom groups (4 and 1) as not many of the group were consistently scoring at levels 1 and 4.
- There were no gender differences evident in either the Watson-Glaser data or the performance data.



Other

Analysts from a U.S. government agency (discussed in Watson and Glaser, 2006) had Watson-Glaser Short Form scores that correlated moderately with supervisory ratings of:

- (a) Analysis and Problem Solving behaviours ($r=0.40$) and
- (b) Judgment and Decision Making behaviours ($r=0.40$)

Scores also correlated moderately with supervisory ratings on a dimension composed of behaviours dealing with Professional/Technical Knowledge and Expertise ($r=0.37$) as well as with "Total Performance" ($r=0.39$).

Using a sample of leadership assessment centre participants, Kudish and Hoffman (2002) reported that Watson-Glaser 80-Item (U.S. form) scores had a large correlation with ratings of Analysis ($r=0.58$) and a moderate correlation with ratings of Judgement ($r=0.43$). Ratings on Analysis and Judgment were based on participants' performance across assessment centre exercises including a coaching meeting, in-basket exercise or simulation, and a leaderless group discussion.

References

- Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed.). New York: Harper & Row.
- Kudish, J. D., & Hoffman, B. J. (2002, October). *Examining the relationship between assessment center final dimension ratings and external measures of cognitive ability and personality*. Paper presented at the 30th International Congress on Assessment Center Methods, Pittsburgh, PA.
- U.S. Department of Labor. (1999). *Testing and assessment: An employer's guide to good practices*. Washington, DC: Author.
- Watson, G., & Glaser, E. M. (2006). *Watson-Glaser Critical Thinking Appraisal, Short Form manual*. San Antonio, TX: Pearson.