

Automatic Scoring of the Alternative Uses Task

Charlotte Tanis, Claire Stevenson, Matthijs Baas & Han van der Maas

Compare reliability and validity of automated AUT scoring to subjective scoring methods

Alternative Uses Task

The AUT measures divergent thinking. Participants are asked to come up with as many creative uses as possible for a common object such as a brick.

-  2 minutes
-  111 first year psychology students
-  session 1: brick and fork, session 2: paperclip and towel
-  Dutch

Scoring methods

Fluency
Number of responses

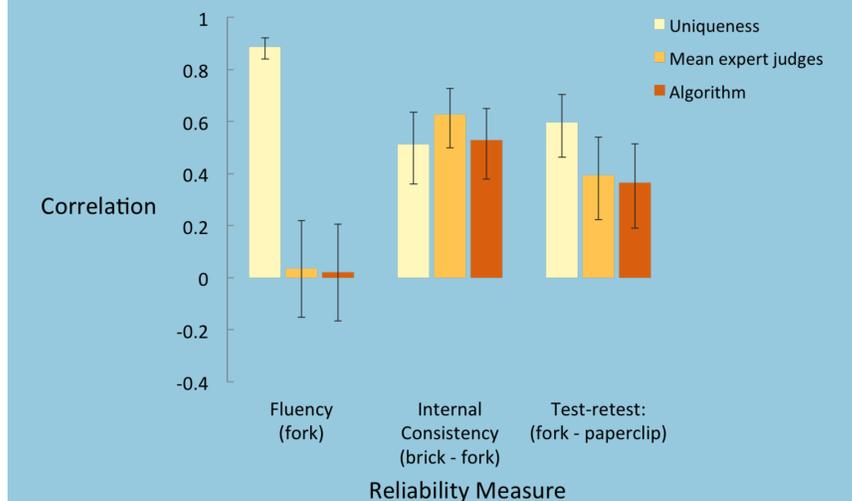
Uniqueness
Each response given by < 5% of sample receives 1 point
Creative potential: Sum of points

Mean of experts
Expert judges score each response, then calculate median
Creative potential: Average all responses

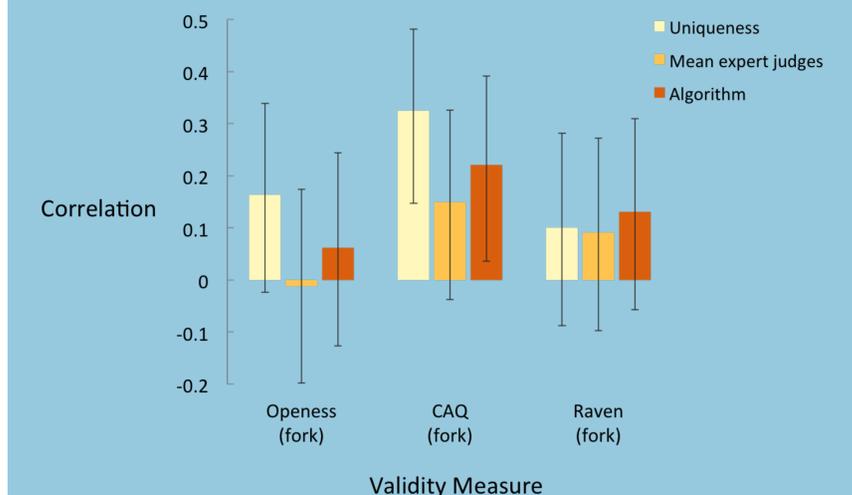
Algorithm
Algorithm scores each response
Creative potential: Average all scored responses

Results

Reliability of scoring methods



Validity of scoring methods



Algorithm

The algorithm is an expert system based on a selective consensual assessment technique. An existing response database is mined to score new solutions.

Database



Scoring new data



Reliability and validity

Reliability

1. Correlation with fluency: should be 0
2. Internal consistency: different object, same session
3. Test-retest reliability: different object, different session
4. Inter-rater reliability

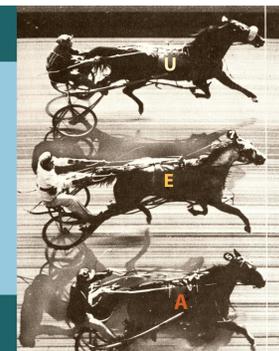
Validity

Correlations between AUT scores and related measures:

1. Openness to Experience
2. Creative Achievement Questionnaire
3. Raven IQ

And the winner is...

All methods in part... No method consistently performs better than others.
Improvements algorithm: more efficient data cleaning & larger database
Benefits algorithm: faster and unbiased scoring



Contact

Would you like more information on this project? Or get access to the database and polished version of the algorithm by contributing your data? Contact Charlotte Tanis:
amsterdamaut@gmail.com



UNIVERSITEIT VAN AMSTERDAM