EyeDetect Audio Multiple-Comparison Test (AMCT)
Development and Validation Summary
June 5, 2021

Converus recently introduced the EyeDetect Audio Multiple Comparison Test (AMCT) to accommodate populations that can't read, or can't read well enough to take the standard Multiple Comparison Test. The AMCT covers up to four relevant issues. It can be used to decide if a person was deceptive to any of the relevant issues on the test, and it can be used to decide if the person was deceptive to each relevant issue individually. Our research suggests that AMCT decisions are approximately 81% correct, whether it is used to decide if the person was deceptive to any question on the test or to decide if the person was deceptive to each question individually.

Purpose

The purpose of this summary is to describe the steps used to develop and validate the AMCT.

Experiment

A mock crime experiment modeled after Cook et al. (2012) was conducted to collect the ocular-motor data needed to develop and cross-validate a statistical model of ocular-motor measures that computes a credibility score for each issue. In this experiment, there were four issues: theft of cash, theft of a gift card, theft of a cell phone, and theft of a headlamp.

One hundred and eighty subjects were recruited from the local community. They were told some subjects would commit one or more of the thefts, whereas others would be innocent, and would not commit any of the crimes. Subjects were arbitrarily assigned to one of three groups. One of two groups of guilty subjects stole $20 from a secretary (n=75). The other group of guilty subjects stole $20 from a secretary AND stole a gift card from a wallet (n=54). The third group of subjects was innocent of all four crimes (n=51).

After subjects carried out their instructions, they were given the AMCT. The AMCT contained eight T/F statements about each of the four topics (32 items) and the set of 32 items was repeated 5 times in different orders. A computer-generated synthetic voice presented instructions and test statements orally over headphones, while a Tobii eye tracker recorded gaze position and pupil size of left and right eyes. The computer also recorded response times and response errors. Subjects were told they should respond quickly and accurately to the statements or they would fail the test. They were paid for their time and paid an additional $30 bonus if they passed the test.

Analysis

The ocular-motor data were analyzed to identify features that discriminated between questions answered truthfully and deceptively. A set of ocular-motor features was identified that achieved over 80% accuracy on the complete set of relevant questions. Those features were weighted and combined by means of a logistic regression equation that generated a credibility score for each relevant question. The credibility index was the probability that the person was truthful about that topic. If the credibility

index was 0.5 or greater, the subject was classified as truthful about that issue. If the credibility index was less than 0.5, the subject was classified as deceptive about that issue.

K-Fold Validation

A statistical model that is optimal for classifying the cases in a particular experiment is rarely optimal for the population from which the subjects were sampled. The model is not optimal because the sample does not perfectly represent the more general population from which it was drawn. Consequently, we obtain biased estimates of accuracy if we test the model on the cases that were used to create the model.

Better estimates of accuracy can be obtained with k-fold validation. A k-fold validation divides the data set into *k* folds (subsets). The first subset comprises a hold-out subsample and is removed from the dataset. The remaining subsets are combined to create a training set. A logistic regression model is developed using the cases in the training set. That logistic regression model is then used to classify the cases in the hold-out subsample. The accuracy observed in the hold-out sample provides a less biased estimate of accuracy because the holdout cases were not used to optimize feature coefficients in the regression equation. The accuracy achieved in the hold-out sample is recorded.

This process continues for each partition of the data set. The first subset is returned to the training set, and the second subset is removed to serve as a new holdout sample. A new logistic regression model is created with all but the second subset of cases. That model is used to classify cases in the holdout sample, and its accuracy is recorded. This process is repeated for each of the remaining subsets. The best estimate of accuracy for the model is mean accuracy across the *k* holdout samples.

Validation of the AMCT

One hundred and eighty subjects were available to validate the AMCT. Each subject was truthful or deceptive to each of four relevant questions. That provided a total of 180 X 4 = 720 relevant questions where the person was truthful or deceptive.

An 8-fold validation was performed with the set of 720 relevant questions. The sample of 720 questions was split into 8 subsamples (folds). Each subsample consisted of 28.3% innocent, 41.7% cash, and 30% cash+card subjects. Table 1 shows percent correct for truthful and deceptive questions for each subsample as well as the mean accuracy across the eight folds.

On average, accuracy was slightly higher for questions answered truthfully (83.8%) than for questions answered deceptively (81.3%). At the level of individual relevant questions, mean accuracy on cross-validation was 82.6%. Based on these results, *we would expect the AMCT to produce 82.6% correct decisions on individual questions when the model is used in a new sample*.

Table 1. Percent correct decisions for questions answered truthfully or deceptively in 8-fold validation

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Mean
Truthful	83.1	90.9	82.1	86.6	80.3	70.6	85.1	91.7	83.8
Deceptive	75.0	100.0	63.6	68.2	70.8	95.5	82.6	95.5	81.3
								Mean Accuracy	82.6

The credibility scores for each holdout sample were used to classify subjects as either truthful to all relevant issues or deceptive to any one or more of the issues. Mean accuracy was 84.8%. Accuracy was higher for innocent (88.2%) than guilty subjects (81.4%). On average, *we would expect the AMCT to produce 84.8% correct decisions when the model is used to classify test subjects in a new sample*.

References

Cook, A. E., Hacker, D. J., Webb, A. K., Osher, D., Kristjansson, S., Woltz, D. J., & Kircher, J. C. (2012). Lyin' Eyes: Ocular-motor Measures of Reading Reveal Deception. *Journal of Experimental Psychology: Applied, 18(3),* 301-313.