EyeDetect Audio Multi-Issue Comparison Test (AMCT)

Development and Validation Summary

J.C. Kircher, Ph.D.

March 13, 2020

Converus recently introduced the EyeDetect Audio Multi-Issue Comparison Test (AMCT) to accommodate populations that can't read or can't read well enough to take a standard Multi-Issue Comparison Test (MCT). The AMCT covers up to four relevant issues. It can be used to decide if a person is deceptive to any of the relevant issues on the test, and it can be used to decide if a person is 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: (1) theft of cash, (2) theft of a gift card, (3) theft of a cell phone, and (4) 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=74). The other group of guilty subjects stole $20 from a secretary AND stole a gift card from a wallet (n=55). The third group of subjects was innocent of all four crimes (n=51).

After subjects completed their instructions, they were given the AMCT. The AMCT contained eight True/False statements about each of the four topics (32 items) and the set of 32 items was repeated five times in different orders. A computer-generated digital 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 the number of questions answered incorrectly. Subjects were told they should respond quickly and accurately to the statements or they would fail the test. Subjects were paid for their time and were paid an additional $30 bonus if they passed the test.

Analysis

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 better than 80%

accuracy on the complete set of relevant questions. For each subject, those features were weighed 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 to questions 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. K-fold validation was completed with 296 of those 720 relevant questions. To achieve a proper balance of truthful and deceptive cases and deceptive answers by guilty subjects who committed one or two crimes, the number of questions for the validation process was limited by the number of guilty subjects who committed one crime (n = 74). All 74 of the questions answered deceptively were included in the validation sample. Then, an equal number of deceptive questions was obtained from a random sample of 37 of the 55 guilty subjects who committed two crimes. Each of those 37 guilty subjects contributed two cases to the deceptive validation sample (2 X 37 = 74) because they lied to two of the four questions on the test. Since the validation sample now contained 2 x 74 = 148 deceptive cases, an equal number of questions truthfully was randomly selected from the entire sample of 180 subjects. That subsample of 148 truthful cases contained a random mix of questions answered truthfully by people who were truthful to all relevant questions (innocent) or were truthful to some questions but not others (guilty).

An 8-fold validation was performed. The sample of 296 questions was split into 8 subsamples of either 36 or 38 questions. Half of the questions in each subsample were questions answered truthfully, and half were questions answered deceptively. Half of the questions answered deceptively were from subjects who committed one crime and half were from subjects who committed two crimes. Table 1 shows percent correct for truthful and deceptive questions for each fold, as well as the mean accuracy across the eight folds.

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**
n	36	38	36	38	36	38	36	38	296
Truthful	88.9	89.5	83.3	84.2	77.8	78.9	66.7	84.2	81.7
Deceptive	88.9	89.5	55.6	78.9	83.3	73.7	83.3	84.2	79.7
								Mean Accuracy	80.7

On average, accuracy was slightly higher for questions answered truthfully (81.7%) than for questions answered deceptively (79.7%). At the level of individual relevant questions, mean accuracy on cross-validation was 80.7%. Based on these results, *we would expect the AMCT to produce 80.7% correct decisions when the model is used in a new sample*.

A new logistic regression equation was developed using all 720 questions from the 180 subjects. The observed mean accuracy at the level of individual questions was 82.2%. When the model was used to decide whether a subject was deceptive to any one or more of the relevant questions, its mean accuracy was 80.7%. The model is slightly more accurate for deceptive subjects (83.1%) than for truthful subjects (78.4%).

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.