THE ROLE OF TABLET-BASED PSYCHOLOGICAL TASKS IN RISK ASSESSMENT

PABLO A. ORMACHEA
Center for Science and Law

BRIAN K. LOVINS
Harris County Community Supervision and Corrections Department

DAVID M. EAGLEMAN
Stanford University

SASHA DAVENPORT

ANNA JARMAN

GABE HAARSMA
Center for Science and Law

Risk assessment has become a prominent part of the criminal justice system in many jurisdictions, typically relying on structured questions and an interview. This approach, however, may not accurately assess certain psychological concepts correlated with reoffense, such as executive functioning, ability to plan, impulse control, risk-taking, aggression, and empathy. We hypothesized that using rapid-tablet-based neurocognitive tests would pay off in terms of objectivity, precision, and scalability when added to the existing risk assessment structure. We analyzed 240 observations from adult felony offenders from a large urban county in the South assessed by the Texas version of the Ohio Risk Assessment System (ORAS) risk tool. We identified significant differences in impulse control, planning, and reactive aggression between offenders and reoffenders. By combining these variables with the Texas Risk Assessment System (TRAS), we yielded significant improvements in risk prediction. We hope this will provide new inroads for actuarial assessments of reoffense risk that incorporate direct measurements of individual decision making.

Keywords: risk assessment; recidivism; impulse control; aggression; psychometrics

Over the past three decades, the proportion of criminal justice agencies that use structured risk assessment has increased significantly (Latessa & Lovins, 2010). Until the early 1970s, most courts and criminal justice agencies relied on professional judgment—a

AUTHORS' NOTE: *This material is based upon work supported by the National Science Foundation SBE Office of Multidisciplinary Activities under Grant No. 1519667. We would like to thank all of the current and former members of the Center for Science and Law who assisted with this project, including Angela Wilkins, Sean Ham, Mikal Lerma, Kristian Marlow, Anjelica Mendoza, Scott Novich, and Joshua Preston. We also thank our interns: Emily Paul, Alex Gajewski, Eleanor Faraguna, Hannah Weinstein, Cailin Frankland, Katherine Yao, and Shez Jafry. Finally, we thank two statisticians, Dr. Simon Sheather and Dr. Edward Jones, at Texas A&M for their guidance with analysis on a related project with similar statistical techniques.* *Correspondence concerning this article should be addressed to Pablo A. Ormachea, Center for Science and Law, 2405 Holcombe Blvd., Suite TMCx+115, Houston, TX 77021; e-mail: ormachea@scilaw.org.*

CRIMINAL JUSTICE AND BEHAVIOR, 2017, Vol. 44, No. 8, August 2017, 993–1008.
DOI: 10.1177/0093854817714018

single criminal justice actor collecting information and making a judgment on what was relevant, important, and correlated to future recidivism (Latessa & Lovins, 2010). Starting in the mid-1970s, correctional professionals began to use structured risk assessments to inform decisions (Baird, 2009).

Such early instruments were built from static risk factors, predominantly criminal history, to assist the criminal justice system in determining the likelihood of an offender to recidivate. Although these early generation tools were more successful than personal judgment, they have their limits in practice (Harris, Rice, Quinsey, & Cormier, 2015). The most significant limitation to static measures of risk is the inability to detect change over time. There is a growing literature suggesting that (a) risk assessments are important in identifying the right people for the right level of treatment and (b) reassessments are important to identify those areas that have been addressed and no longer pose a significant risk (Cohen, Lowenkamp, & VanBenschoten, 2016; Labrecque, Smith, Lovins, & Latessa, 2014; Vose, Smith, & Cullen, 2013). In contrast, there are several critiques of instruments that combine dynamic and static factors into a single risk assessment (Baird, 2009). One such critique is the ability of the assessor to collect and score the instrument accurately. This issue can be exacerbated by including items that are difficult to directly observe, and instead relying on a face-to-face interview—specifically the criminogenic domains focusing on antisocial attitudes, personality characteristics, and social skills (Kennealy, Skeem, & Hernandez, 2017). Although the research is clear that measuring dynamic risk is a key element in changing offender behavior, it has been generally silent on the methods to gather information accurately.

Beginning with the Wisconsin Risk Needs Assessment in the 1970s, correctional practitioners were trained to use information obtained from a file review and collateral information (Baird, 2009). With more complex assessments such as the Level of Service Inventory–Revised (LSI-R) and Ohio Risk Assessment System (ORAS), a structured interview component was added to help ensure information was collected consistently across interviews (Andrews & Bonta, 1995; Latessa, Lemke, Makarios, Smith, & Lowenkamp, 2009). In an effort to save time and resources, instruments like the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) developed a standalone self-report option, allowing offenders to answer a series of questions through an automated process (Northpointe, 2012).

LITERATURE REVIEW

THE TEXAS RISK ASSESSMENT SYSTEM

The Texas Risk Assessment System (TRAS), based on the ORAS, is considered a fourth generation risk assessment (Latessa et al., 2009). In fact, the TRAS is actually a set of seven separate assessments that are designed to measure risk at specific points in time for specific types of offenders. There are assessment tools within the TRAS to help assist with jail/bond decisions, sentencing, prison intake and release, and parole along with assessments specific for misdemeanants and felons (Lovins, Latessa, May, & Lux, 2017). The TRAS was designed using theoretically relevant measures and developed through a series of actuarial models.

In Texas, the TRAS–Community Supervision Tool (TRAS-CST) is used by probation departments to help place offenders in levels of supervision, as well as to ensure that criminogenic targets are being addressed. Ultimately, the TRAS-CST measures both static and dynamic factors—some of which are factors associated with cognition and cognitive skills. In all, the TRAS-CST has 34 items across seven domains. The domains cover the primary criminogenic risk areas identified in extant research (see Andrews & Bonta, 1995, for complete review): criminal history, employment/education, family, neighborhood factors, peers, substance abuse, and criminal attitudes. The TRAS-CST is scored as a composite risk assessment providing an overall score and level associated with general reoffending. Within the composite score, items are grouped together to create a risk level within each domain. The domain levels are useful for case planning and assisting criminal justice actors in identifying criminogenic targets.

Lovins and colleagues (2017) found that the TRAS was a valid measure of risk for the overall population as well as subpopulations of men, women, and people of color. The area under the curve (AUC) for the TRAS-CST in the validation study was .674 for males and .676 for females. Regarding race and ethnicity, the AUC for Hispanics was .693, Caucasians was .670, and African Americans was .652. Although the values of AUC are considered moderately strong and consistent with other risk assessments, there is considerable room to grow in measuring risk to reoffend.

RISK ASSESSMENT: INTEGRATION OF TECHNOLOGY

One such area of potential growth is how risk assessment information is collected. Traditionally, corrections and offender treatment programs have been slow to integrate technology with practice (Marsch, 2012). Some advancements have been made recently in surveillance strategies (e.g., GPS tracking) to supervise offenders, but there still remains a significant gap in integrating technology in other areas of community corrections (Geis, Gainey, & Healy, 2016). Integrated management systems that automate counts and inmate movement are rarely found in prisons. Case management systems that incorporate the probation officer's workflow are often not utilized, relying on older technology in which the officer uses the system as a secondary documentation system. As for conducting assessments, correctional professionals are often limited to hard copies of collateral information and traditional pen-and-paper interview structures instead of more technologically advanced methods of collecting information.

There are a limited number of studies that examine the use of technology to assist correctional professionals in gathering information to complete risk assessments. This is especially important in an era in which dynamic measures of risk have shown to be effective in managing offender behavior, and are necessary if agencies are interested in changing criminogenic behavior (Cohen et al., 2016). One of the barriers to successfully addressing these dynamic needs is the ability to accurately assess them in individuals especially as reassessments are conducted annually (Kennealy et al., 2017).

As the field moves to more dynamic measures of risk and, specifically, those measures that are examining cognition and cognitive skills, the ability to ascertain these measures in a face-to-face interview can become difficult. Probing a recalcitrant offender about his or her self-perception as a risk taker is a trial in an interview setting. Moreover, unlike criminal history or employment, where an assessor can validate the responses from collateral information or an employer if necessary, questions like risk taking are difficult to validate from a secondary source. Without a secondary source of information, it is difficult to ascertain whether the offender is being honest about his or her behavior or is attempting to appear more favorable in hopes to get lesser interventions.

Given some of the barriers to measuring cognition, we have developed a tablet-based suite of interactive, modular tests that are grounded in neurocognitive assessments to better measure personality traits, cognition, and problem solving. Instead of relying solely on face-to-face interviews conducted by correctional and administrative professionals, these neurocognitive tasks are designed to use technology to gather the offender's criminogenic traits through a game-like environment. This provides the benefits of objectively measuring the traits of interest in a way that is more organic and potentially more accurate than relying solely on the traditional face-to-face interactions.

DECISION-MAKING TRAITS ASSOCIATED WITH REOFFENSE

Previous research has identified several cognitive and emotional traits that are associated with criminal reoffense—from reduced impulse control (Pratt & Cullen, 2000) to empathy deficits (Wilson, Juodis, & Porter, 2011), among others. Neuroimaging research suggests these deficits may result from underdeveloped structures or underactive function within the brain, particularly in the dorsomedial prefrontal cortex, anterior insular cortex, caudate, and orbitofrontal cortex (Jankowiak-Siuda & Zajkowski, 2013; Rosell & Siever, 2015; Siever, 2008; Stein, Hollander, & Liebowitz, 1993). While neuroimaging would provide a means to measure brain functioning, it is too expensive to deploy broadly within the criminal justice system. Fortunately, there is a wide range of validated interactive assessments from neuropsychology that could be deployed to help assist assessors in capturing these characteristics.

To measure these characteristics, cognitive tasks were selected from the existing literature to identify effective, suitable, and interactive measurements for assessing self-control (Pratt & Cullen, 2000), executive function (Morgan & Lilienfeld, 2000), attentiveness (Meier, Perrig, & Koenig, 2012), empathy (Wilson et al., 2011), and aggressive behavior (Monahan, Steinberg, Cauffman, & Mulvey, 2009). Such behaviors have been linked with a heightened risk for engaging in criminal behavior. Specifically, the assessments were built as a combination of six areas:

Impulse Control

Perhaps the most well-researched offender deficit is impulse control. Specifically, research has generally found impulse control to be a strong and consistent predictor of criminal involvement (Pratt & Cullen, 2000). For instance, Carroll et al. (2006) found that deficits in impulsivity were related to repeat offending among several subcategories of offenders. In juveniles, there are higher levels of impulsivity and poor cognitive inhibitory control among early-onset offenders than late-onset offenders and controls.

Executive Function

Offenders have frequently been identified as suffering from deficits in executive function, described as the ability to "adapt to novel and diverse situations while simultaneously inhibiting inappropriate behaviors" (Hancock, Tapscott, & Hoaken, 2010, p. 339). Executive function is an advanced construct that includes many subparts such as planning, organized search, and impulse control (Welsh, Pennington, & Groisser, 1991). A number of studies and at least one meta-analysis have demonstrated substantial deficits in these areas among offender samples, including difficulties in set shifting (task switching), concentration and organization, as well as the ability to focus on goal-directed behaviors (Broomhall, 2005; Hoaken, Shaughnessy, & Pihl, 2003; Marsh & Martinovich, 2006; Raaijmakers et al., 2008; Séguin, Nagin, Assaad, & Tremblay, 2004; Shallice & Burgess, 1991). In a recent study, Hancock et al. (2010) found that executive functioning deficits were related to violent offending, but not to nonviolent criminal involvement. Given the link between executive functioning deficits and violent crimes, there is a need to assess the depth and breadth of executive functioning deficits within the criminal population.

Empathy

Researchers identify two types of empathy: affective and cognitive. Affective empathy refers to the capacity to respond with an appropriate emotion to another's emotional state. Cognitive empathy is the capacity to understand another's emotions. Both types of empathy are associated with offending (Wilson et al., 2011). A number of studies point to deficits in cognitive empathy among offenders, particularly in the perception of anger, sadness, fear, and disgust (Blair, Colledge, Murray, & Mitchell, 2001; Blair, Mitchell, Peschardt, & Perrett, 2004; Blair et al., 2002).

Aggression

Research has found that the inability to suppress aggression is associated with increased odds of reoffending (Monahan et al., 2009). The struggle with empathy may explain the heightened aggression that has been identified among persistent recidivists (Monahan et al., 2009), as it has been suggested that aggressive offenders may lack the cognitive empathy necessary to prevent or avoid difficult situations (Louise von Borries et al., 2012).

The two most experimental metrics in the tablet are below. There is only limited research supporting the links between these traits and recidivism, suggesting a potential inroad for new behavioral markers.

Risk-Taking

Researcher discussions with local police officers and probation officers indicated that they considered risk-taking was a potential factor influencing future criminal activity. Although some recent studies have suggested that risk-taking can be a predictor of recidivism (Latessa & Lovins, 2010; Lovins et al., 2017), there is a dearth of research exploring the relationship between risk-taking and criminal reoffense, specifically. Here, the game selection was guided primarily by research with other populations, including adolescents, which often relies on the Balloon Analogue Risk Task (Lejuez et al., 2002; Swogger, Walsh, Lejuez, & Kosson, 2010). This test has served as an effective predictor for real-world risk-taking behavior (Lejuez, Aklin, Zvolensky, & Pedulla, 2003).

Planning and Spatial Problem Solving

Initial research has suggested that some offenders, particularly domestic violence offenders, have a marked decrease in planning ability as measured by the Tower of London task (Mintz, 2008). However, Schiffer and Vonlaufen's (2011) research using 30 child molesters did not find significant between-group differences in the Tower of London task compared to the controls.

TABLE 1: Sample Statistics

Demographic variable	*n*	%
Gender		
Male	165	68.8
Female	70	29.2
Unreported	5	2.1
Race/ethnicity		
Black, Non-Hispanic	81	33.8
White, Non-Hispanic	56	23.3
Hispanic	85	35.4
Other/not identified	13	5.4
Age		
18 to 20	31	12.9
21 to 24	43	17.9
25 to 29	45	18.8
30 to 39	61	25.4
40+	53	22.1
Unreported	7	2.9

Given the growing importance on measuring more dynamic risk factors and the difficulty of correctional professionals to measure these factors through face-to-face interviews, we looked to develop a more comprehensive method of assessing these factors. This study examines the ability of the structured tablet games to (a) measure factors related to reoffending and (b) determine whether these measures could be used in conjunction with the TRAS to enhance the current ability to predict recidivism.

METHOD

PARTICIPANTS

The initial step for study participants was to complete a battery of assessments conducted by the local probation department. These assessments included a series of self-report questionnaires and the TRAS-CST. Upon completion, participants were randomly selected to participate in completing the interactive games on the tablet. Specifically, researchers would enter the building 3 times a week for at least 4 h at a time. During that time, every second individual who was undergoing an initial assessment by the local probations department was offered the opportunity to consent to participation in the study. As part of the tablet process, participants completed a demographic survey before completing each assessment in the following order: Eriksen Flanker (executive function), Balloon Analogue Risk Task (risk-taking), Go/No-Go (impulse control), Point-Subtraction Aggression Paradigm (reactive aggression), Reading the Mind Through the Eyes (cognitive empathy), and the Tower of London (planning). The tasks are explained in detail below.

Table 1 presents the descriptive characteristics of our sample. There were 240 offenders in the study. The average age of the participants was 31.8 and ranged from 17 to 68 years. In all, 23% of the sample was White/non-Hispanic, 34% were identified as Black/non-Hispanic, and 35% were identified as Hispanic.

MEASURES

Recidivism

The variable of interest in this study is a new arrest following the assessment until our cut-off date of January 2017. Assessments were administered between August 2015 and January 2016, which means time at risk was a mean of 439 days, with a minimum of 350 days and a maximum of 496 days. Overall, 20% of the sample had reoffended by the cut-off date.

Structured Risk Assessment

The TRAS-CST is comprised of 34 items across seven domains ranging from criminal history and employment measures to criminal attitudes and cognitive skills. The information to score the TRAS-CST is generally collected through a combination of a face-to-face interview, self-report questionnaire, collateral interviews, and criminal justice records. For misdemeanors and felonies, the risk scores are summed and then binned into three or four risk categories, respectively. It is the Texas implementation of the ORAS-CST, which has 23%, 31%, 44%, and 52% recidivism rates for the low, moderate, high, and very high risk categories, respectively (Latessa, Lovins, & Lux, 2014).

The interactive tasks were selected because they measured well-established variables in which previous literature suggested there was a link between the trait and criminal reoffense:

Executive Functioning

The Eriksen Flanker task was used to measure deficits in executive functioning. Each user sees a school of fish that points left or right. The middle fish may point in the same direction (congruent) or a different direction as the school (incongruent). As quickly as possible, his or her task is to indicate whether the middle fish is pointed left or right. Performance is then scored using a National Institute of Health formula based on reaction time and accuracy. The closest item on the TRAS is the measure quantifying the offender's belief that he can control his response to external stimuli.

Planning

In the Tower of London task, each user is presented with three poles that hold three colored discs. The task is to make the fewest moves possible to match the target condition. The number of correct trials was tracked as well as the number of extra moves required to complete them. The TRAS does not currently have a measure for planning.

Impulse Control

The tablet provides two metrics for impulse control. First, we use a variable from the Tower of London: median time before the first move. Second, we rely on an established assessment called the Go/No-Go task. In this task, a participant is tasked with tapping the screen as quickly as possible any time he sees a carrot. Then, 20% of the time, an eggplant appears and the participant must inhibit the impulse to tap. We measure the number of correct taps on the go-stimulus as well as the number of times the participant correctly inhibits on a no-go stimulus. For the TRAS, the closest item for impulse control is the measure for an offender's ability to respond effectively to situations outside his control.

Risk-Taking

In the Balloon Analogue Risk Task, each user is presented with 30 balloons. His task is to push the "inflate" button to blow up the balloon as large as the participant dares before hitting the "collect points" button. As the balloon increases in size, so do the odds of popping, which means the participant receives no points. The number of taps and the number of balloon pops are tracked so as to create a measure for a risk-taking. The TRAS examines risk taking directly, measuring the offender's engagement in risk-taking situations.

Reactive Aggression

In the Point-Subtraction Aggression Paradigm, users are tasked with tapping a "grow" button to accumulate as many "dollars" as possible. They are set against a fictitious second player who is busy performing the same task. However, the "second player" eventually "punishes" the user by destroying one of the dollars. The user then has a choice: to avenge the punishment or to ignore it and continue "growing." The optimal strategy (i.e., the one that is consistent with the instructions to make the most money) is to ignore the insult and continue growing (i.e., pushing the "grow" button over and over). Therefore, the amount of their time that they chose to tap the alternate button—otherwise "punish" their opponent—serves as a measure of reactive aggression. The number of times a user grows, punishes, or protects the dollars is tracked. The TRAS relies on two items to measure aggression: attitude toward fighting and predilection for preemptive aggression.

Cognitive Empathy

We use the Reading the Mind Through the Eyes task to measure empathy. In this task, users are presented with the upper half of 30 different faces (including the eyes). They are tasked with selecting the word that best describes each face's emotional state. We track the number of incorrect trials. The TRAS measures the capacity of the client to show concern for others through the face-to-face interview.

DEMOGRAPHIC VARIABLES

For analysis of both the tablet data and the TRAS data, the following demographic variables were included in the initial model, and stepwise logistic regression was used for variable selection to decide whether to keep them in the final model.

Race/Ethnicity

Race and ethnicity were measured as a self-report question in which the offender was asked to self-identify race and ethnicity. Nonrecidivators were comprised of 33% Black, 23% White, and 41% as Other. Recidivators, by contrast, were 36% Black, 23% White, and 40% Other.

Ethnicity

The offender was asked to self-identify as Hispanic or Non-Hispanic. Nonrecidivators and recidivators were similar, with 36% and 34%, respectively, self-reporting as Hispanic.

Gender

Offenders were asked to self-identify as male or female. Males were overrepresented in the sample of recidivators, with 74% compared with 67% of nonrecidivators.

Age

Age was calculated from date of birth and represents the age in years at the time of the assessment. The sample of nonrecidivators had a mean age of 33.03 with a standard deviation of 11.59. The recidivators were younger with a mean age of 26.85 and a standard deviation of 7.54.

Primary Language

Our assessments were exclusively provided in English. Houston, though, has large Hispanic and Vietnamese populations. To enable statistical tests that evaluate whether performance on our assessments was affected by having English as a Second Language— particularly the Reading the Mind Through the Eyes task — we had users report their primary language. For nonrecidivators, 90%, 6%, and 1% reported their primary languages as English, Spanish, and Vietnamese, respectively. For recidivators, there were slightly more English speakers (96%) and fewer Spanish speakers (4%).

ANALYSIS

Data were collected and analyzed to determine the degree to which the TRAS-CST and the tablet battery were effective in predicting future offending. The TRAS-CST assigns a categorical response to each item of interest (0-1 or 0-1-2) for recorded responses. The tablet, by contrast, records a series of continuous variables (Table 2). Therefore, to fit the tablet within the existing TRAS-CST paradigm and risk assessment in general, binary thresholds were identified for each of the continuous variables. We turned to recursive partitioning, an implementation of Classification and Regression Trees (CART; Breiman, Firedman, Olshen, & Stone, 1983). The tree is built by first identifying the single variable that best splits the data into two groups based on reoffense. Then, for each subgroup, the process is repeated with all remaining variables recursively until no further improvement can be made (Therneau, Atkinson, & Ripley, 2015). Finally, the algorithm relies on cross-validation to prune back the tree to identify only significant differences (Therneau et al., 2015). This process not only returns optimal thresholds but also performs variable selection by returning only the variables that provided significant differences between the two samples. These thresholds are used to convert the continuous tablet variables into binary variables similar to the subitems in the TRAS.

Upon identifying specific items and determining the appropriate thresholds, bivariate tests were conducted to determine whether the new items were significantly correlated with

TABLE 2: Performance Statistics

	Nonrecidivators				Recidivators			
Tablet variable	*M*	*SD*	Minimum	Maximum	*M*	*SD*	Minimum	Maximum
Ericsson-Flanker task								
Executive effect (seconds)	0.03	0.3	−1.31	2.07	0.07	0.29	−0.63	1.56
National Institute of Health score	8.61	1.35	3.75	10	9.00	0.81	5.8	10
Go/No-Go tasks								
Incorrect no-go (trials)	4.49	4.03	0	18	5.02	4.03	0	15
Incorrect go (trials)	12.26	11.84	0	55	7.85	8.42	0	40
Balloon Analogue Risk task								
Balloons popped	9.56	2.98	2	22	9.85	3.87	5	30
Button press time	62.26	18.91	9.51	110.2	64.91	19.83	24.08	120.1
Point-Subtraction Aggression Paradigm								
Punishes	9.04	9.34	0	34	10.4	9.84	0	34
Protects	8.01	6.3	0	21	8.04	5.92	0	23
Reading the Mind Through the Eyes task								
Correct trials	20.92	5.63	4	31	21.49	5.16	8	31
Tower of London task								
Trials solved	17.37	2.91	10	24	17.09	4.11	5	24
Time until first move	5.22	1.62	2.41	13.56	5.75	2.01	3.09	12.13

recidivism. Specifically, we tested the binary splits with independent sample *t* tests to quantify the significance of the difference between those who reoffended and those who did not. Once these items were identified, each item was matched with existing TRAS items that measured similar concepts. The fifth step was to replace the TRAS items with the new items to determine whether the TRAS results were stronger for participants in a combined, substituted model. To determine whether the results were stronger, a receiver operating characteristic (ROC) curve analysis was conducted and AUC scores were examined. The same process was conducted in the sixth step, except that instead of replacing the items, the new tablet items were simply added to the TRAS structure, creating a TRAS + Tablet model. As the overall composite score was increased from 47 to 52 points, the categories were redefined using the ROC analyses and adjusting the cutoffs to meet the most appropriate level of specificity and sensitivity (Hajian-Tilaki, 2013). To maintain consistency with standard practice for TRAS AUC measurements, AUC scores were calculated using the categorical scale of risk levels.

RESULTS

IS THE TRAS A VALID MEASURE OF RECIDIVISM FOR THIS SAMPLE?

Our first analysis set out to determine whether the TRAS was a valid predictor of recidivism for the subject group. The TRAS successfully separated the sample into four distinct categories ranging from 5.3% rearrest rate for low risk up to 35.5% rearrest for the highest risk group (Table 3). Moreover, the AUC was .635, suggesting the TRAS demonstrated modest capabilities in predicting new arrest over chance.

TABLE 3: Recidivism Rates for the TRAS

Risk category	*n*	% rearrested
Low risk	19	5.26
Low/moderate risk	88	13.64
Moderate risk	102	22.55
High risk	31	35.48

Note. AUC = .635; $p \geq .001$. TRAS = Texas Risk Assessment System; AUC = area under the curve.

TABLE 4: Reoffense Statistics for the Tablet Variables

Tablet variable	*n*	% rearrested	Chi-square *p* value
Flanker Executive Effect (executive functioning)			
0 (<0.10)	192	18.8	
1	47	23.4	.472
GNG Incorrect Go (accuracy; impulse control)			
0	85	7.06	
1 (<11)	155	26.5	<.001
PSAP Grows (reactive aggression)			
0 (<32.00)	233	18.0	
1	7	71.4	<.001
PSAP Punishes (reactive aggression)			
0 (<33.50)	237	19.0	
1	3	66.7	.039
TOL Solved (planning; set shifting)			
0 (>10.50)	193	16.1	
1	47	34.0	.005
TOL Time to First Move (impulse control)			
0 (<8.39)	168	15.5	
1	72	29.2	.014

Note. GNG = Go–No Go; PSAP = Point-Subtraction Aggression Paradigm; TOL = Tower of London.

WHAT FACTORS FROM THE TABLET ASSESSMENT WERE ASSOCIATED WITH RECIDIVISM?

Once it was determined that the TRAS was a valid measure of risk for this sample, the second step was to examine bivariate relationships between tablet measurements and rearrest. In the judgment of one of the designers of the TRAS, six tablet metrics appeared to measure similar concepts as the TRAS. Of the six items, the Flanker Executive Effect was the only one that did not demonstrate a significant relationship (Table 4). The other five measures were independently related to rearrest.

DOES THE TRAS PERFORM BETTER WHEN THE TABLET MEASURES REPLACE SPECIFIC TRAS ITEMS?

To address this question, tablet items that were identified as statistically significant were matched with TRAS items that measured the same concept. As noted in the “Method” section, the aggression measures from the tablet replaced the questions regarding attitude toward fighting and preemptive anger from the TRAS, and the impulsivity measure from the tablet replaced the impulsivity question from the TRAS. Table 5 provides the results of

TABLE 5: Percentage of Offenders by Risk Categories: TRAS + Replacement

Risk category	*n*	% rearrested
Low	19	5.26
Low/moderate	91	9.89
Moderate	105	23.81
High	25	48.00

Note. AUC = .692; $p \geq .001$. TRAS = Texas Risk Assessment System; AUC = area under the curve.

TABLE 6: Percentage of Offenders by Risk Categories: TRAS + Tablet Measures

Risk category	*n*	% rearrested
Low	75	5.33
Moderate	129	22.48
High	36	38.89

Note. AUC = .683; $p \geq .001$. TRAS = Texas Risk Assessment System; AUC = area under the curve.

the TRAS plus replacement items. As noted, the TRAS with replacement items demonstrated a clear differentiation of risk categories. Moreover, the AUC improved to .692, suggesting that this model is more effective at measuring risk than the TRAS alone.

DOES THE TRAS PERFORM BETTER IN CONJUNCTION WITH THE TABLET BATTERY?

When combining the assessments, we used the TRAS score and the five tablet items for each participant that were significantly correlated with reoffending. Given that the potential raw score increased from 47 to 52 points, the data were reanalyzed to determine the best fit. In developing cutoffs for risk assessments, in general, a preconceived number of categories does not exist, more so because the data drive the development of the threshold cutoffs for each risk category. Although it is preferred to create cutoffs based on a construction sample and then apply the results to a validation sample, we were limited in this study to developing the identified cutoffs based on a single sample. Based on the ROC analyses, it was determined that the TRAS plus the tablet measures provided a clear three-category fit instead of the original four category fit.[1] Table 6 provides the results of the final categorization of the TRAS + Tablet items. As noted, the AUC for the combined model is .683, which suggests this model performs slightly worse than when the tablet tasks replace the TRAS subitems.

DISCUSSION

The goal of this study was to explore the possibility of using computer-based interactive games to measure the more difficult items around cognition, aggression, and impulsivity. First, we examined the ability of the TRAS-CST to measure future recidivism. It was found that the TRAS-CST, using traditional means of face-to-face interviews, was a valid measure of risk. Second, we examined the validity of measures extracted from the tablet games to measure recidivism independently. Although there were some measures that were not predictive of reoffending, there were five measures that demonstrated a bivariate relationship with reoffending.

With a base TRAS-CST completed and six measures of risk associated with the tablet games, we then explored whether the cognitive tablet games could replace similar measures from the TRAS and produce a better score or be combined with the TRAS to strengthen the measure of risk. As discussed earlier, we chose to compare the values of AUC across results to determine whether there was a clear distinction between the strength of any one method. We found that adding the TRAS-CST items with the cognitive tablet measures did increase the AUC of the standalone TRAS-CST from .635 to .683 and replacing the TRAS-CST measures with the tablet measures increased the AUC to .692.

The two models in which the tablet measures were included did show substantive increases in the raw AUC scores. There are several reasons in which the tablet games may have produced more effective measures of risk than the TRAS alone. First, the information gathered in an interview is limited to the ability of the assessor to ask relevant and appropriate questions. If assessors are not engaged and asking open-ended questions to elicit conversation, often these items are answered and scored through short, direct answers. For example, an interviewer exploring whether an offender is likely to use aggression to solve a disagreement is going to receive significantly different information if they asked “Tell me about a time when you used anger to solve a problem,” than if they asked the question “Have you ever used anger to solve a problem.” The former question would elicit a response that would allow for the assessor to probe and receive more information, whereas the latter question would get a single yes/no response.

Second, questions that measure feelings and cognitive processes are often difficult to probe through question/answer interactions. For example, trying to measure impulsivity through a series of questions is much more difficult than measuring it through observations. If an offender was placed in a lab for 2 weeks, an assessor could observe the offender’s behavior and determine if impulsivity was an issue, but often an assessor is limited to an hour discussing someone’s impulsivity in a vacuum.

Third, offenders are shaped to answer questions around impulsivity and aggression in the negative—knowing that the answers will cause additional interventions. For instance, if asked “Are you a violent person?” the most typical response would be “No.” They know the answer that keeps them from further scrutiny. When using interactive games, the assessment of these characteristics is done through observations tied to the games—and therefore are potentially more difficult to “fake” by the offender.

LIMITATIONS

There were several limitations to this study that should be noted. First, the sample size was relatively small and the distribution of cases across risk categories was limited. Expanding the number of offenders will provide a more robust sample and will allow for future studies to continue to explore the improvement of cognitive risk assessment measures. Second, arrests were limited to the offender’s initial county of conviction. Whereas this does limit the percentage of those reoffending, there is no evidence that expanding the arrests statewide would change whether a characteristic was predictive of recidivism. Third, the current nature of risk assessment is to use discrete measures of risk. Using interactive games to measure risk expands these measures to more continuous variables allowing for a more robust measure of differences. While this study condensed the measures to discrete measures, future studies should explore how to integrate continuous variables into risk assessment.

FUTURE DIRECTION

This study suggests that there is room for improvement in risk assessment. Risk assessment instruments are becoming significantly more important in the implementation of correctional services and a greater emphasis is being placed on cognitive measures of risk. With the increase in complex concepts, we must continue to explore ways to assess individuals accurately. Moreover, these concepts even become more difficult to measure through face-to-face interviews as offenders are reassessed multiple times.

For risk assessment, the inclusion of interactive, tablet-based games is a step forward in decision-making theory and in the study of recidivism. Improved knowledge in either of these areas could inform scientifically based practices aimed at the prevention and control of crime that reflect individual offender differences. Our findings move us one step closer toward providing an ability to base sentencing decisions on direct, proven, open-source assessments of criminal propensity. In the long term, we hope that quantifying the relationship between criminal behavior and cognitive or empathic traits will lead to new alternative and individualized sentencing strategies where sentence length and program eligibility incorporate a careful understanding of the offender's decision making.

NOTE

1. The values of area under the curve (AUC) are sample specific.

REFERENCES

Andrews, D. A., & Bonta, J. (1995). *LSI-R: The Level of Service Inventory–Revised*. Toronto, Ontario, Canada: MultiHealth Systems.

Baird, C. (2009). *A question of evidence: A critique of risk assessment models used in the justice system*. Madison, WI: National Council on Crime & Delinquency.

Blair, R., Colledge, E., Murray, L., & Mitchell, D. (2001). A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies. *Journal of Abnormal Child Psychology*, *29*, 491-498.

Blair, R., Mitchell, D. G., Peschardt, K. S., & Perrett, D. I. (2004). Reduced sensitivity to others' fearful expressions in psychopathic individuals. *Personality and Individual Differences*, *37*, 1111-1122.

Blair, R., Mitchell, D. G., Richell, R. A., Kelly, S., Leonard, A., Newman, C., & Scott, S. K. (2002). Turning a deaf ear to fear: Impaired recognition of vocal affect in psychopathic individuals. *Journal of Abnormal Psychology*, *111*, 682-686.

Breiman, L., Firedman, J. H., Olshen, R. A., & Stone, C. J. (1983). *Classification and regression trees*. Belmont, CA: CRC Press.

Broomhall, L. (2005). Acquired sociopathy: A neuropsychological study of executive dysfunction in violent offenders. *Psychiatry, Psychology and Law*, *12*, 367-387.

Carroll, A., Hemingway, F., Bower, J., Ashman, A., Houghton, S., & Durkin, K. (2006). Impulsivity in juvenile delinquency: Differences among early-onset, late-onset, and non-offenders. *Journal of Youth and Adolescence*, *35*, 519-529.

Cohen, T., Lowenkamp, C., & VanBenschoten, S. (2016). Does change in risk matter? *Criminology & Public Policy*, *15*, 263-296.

Geis, S., Gainey, R., & Healy, E. (2016). Monitoring high-risk sex offenders with GPS. *Criminal Justice Studies: A Critical Journal of Crime, Law and Society*, *29*, 1-20.

Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine*, *4*, 627-635.

Hancock, M., Tapscott, J., & Hoaken, P. (2010). Role of executive dysfunction in predicting frequency and severity of violence. *Aggressive Behavior*, *36*, 338-349.

Harris, G., Rice, M., Quinsey, V., & Cormier, C. (2015). *Violent offenders: Appraising and managing risk* (3rd ed.). Washington, DC: American Psychological Association.

Hoaken, P., Shaughnessy, V., & Pihl, R. (2003). Executive cognitive functioning and aggression: Is it an issue of impulsivity? *Aggressive Behavior*, *29*, 15-30.

Jankowiak-Siuda, K., & Zajkowski, W. (2013). A neural model of mechanisms of empathy deficits in narcissism. *Medical Science Monitor*, *19*, 934-941.

Kennealy, P., Skeem, J., & Hernandez, I. (2017). Does staff see what experts see? Accuracy of front line staff in scoring juveniles' risk factors. *Psychological Assessment*, *29*, 26-34.

Labrecque, R., Smith, P., Lovins, B., & Latessa, E. (2014). The importance of reassessment: How changes in the LSI-R risk score can improve the prediction of recidivism. *Journal of Offender Rehabilitation*, *53*, 116-126.

Latessa, E., Lemke, R., Makarios, M., Smith, P., & Lowenkamp, C. (2009). The creation and validation of the Ohio Risk Assessment System (ORAS). *Federal Probation*, *74*(1), 16-22.

Latessa, E., & Lovins, B. (2010). The role of offender risk assessment: A policy maker guide. *Victims & Offenders*, *5*, 203-219.

Latessa, E., Lovins, B., & Lux, J. (2014). *The Ohio Risk Assessment System Misdemeanor Assessment Tool (ORAS-MAT) and Misdemeanor Screening Tool (ORAS-MST)*. Cincinnati, OH: Center for Criminal Justice Research, University of Cincinnati.

Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., & Pedulla, C. M. (2003). Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviours. *Journal of Adolescence*, *26*, 475-479.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., . . . Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, *8*, 75-84.

Louise von Borries, A. K., Volman, I., de Bruijn, E. R., Bulten, B. H., Verkes, R. J., & Roelofs, K. (2012). Psychopaths lack the automatic avoidance of social threat: Relation to instrumental aggression. *Psychiatry Research*, *200*, 761-766.

Lovins, B., Latessa, E., May, T., & Lux, J. (2017). *Exploring the validation of the Ohio risk assessment system community supervision tool with a diverse sample of offenders in Texas*. Manuscript in preparation.

Marsch, L. (2012). Leveraging technology to enhance addiction treatment and recovery. *Journal of Addictive Diseases*, *31*, 313-318.

Marsh, N., & Martinovich, W. (2006). Executive dysfunction and domestic violence. *Brain Injury*, *20*, 61-66.

Meier, N., Perrig, W., & Koenig, T. (2012). Neurophysiological correlates of delinquent behaviour in adult subjects with ADHD. *International Journal of Psychophysiology*, *84*, 1-16.

Mintz, S. R. (2008). *An analysis of executive functioning in a sample of first offense domestic violence offenders* (Unpublished doctoral dissertation). Alliant International University, San Diego, CA.

Monahan, K., Steinberg, L., Cauffman, E., & Mulvey, E. (2009). Trajectories of antisocial behavior and psychosocial maturity from adolescence to young adulthood. *Developmental Psychology*, *45*, 1654-1668.

Morgan, A., & Lilienfeld, S. (2000). A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function. *Clinical Psychology Review*, *20*, 113-136.

Northpointe. (2012). *COMPAS Risk & Need Assessment System: Selected questions posed by inquiring agencies*. Traverse City, MI. Retrieved from http://www.northpointeinc.com/files/downloads/FAQ_Document.pdf

Pratt, T., & Cullen, F. (2000). The empirical status of Gottfredson and Hirschi's general theory of crime: A meta-analysis. *Criminology*, *38*, 931-964.

Raaijmakers, M. A., Smidts, D. P., Sergeant, J. A., Maassen, G. H., Posthumus, J. A., van Engeland, H., & Matthys, W. (2008). Executive functions in preschool children with aggressive behavior: Impairments in inhibitory control. *Journal of Abnormal Child Psychology*, *36*, 1091-1107.

Rosell, D. R., & Siever, L. J. (2015). The neurobiology of aggression and violence. *CNS Spectator*, *20*, 254-279.

Schiffer, B., & Vonlaufen, C. (2011). Executive dysfunctions in pedophilic and nonpedophilic child molesters. *The Journal of Sexual Medicine*, *8*, 1975-1984.

Séguin, J., Nagin, D., Assaad, J., & Tremblay, R. (2004). Cognitive-neuropsychological function in chronic physical aggression and hyperactivity. *Journal of Abnormal Psychology*, *113*, 603-614.

Shallice, T., & Burgess, P. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, *114*, 727-741.

Siever, L. J. (2008). Neurobiology of aggression and violence. *American Journal of Psychiatry*, *165*, 429-442.

Stein, D. J., Hollander, E., & Liebowitz, M. R. (1993). Neurobiology of impulsivity and the impulse control disorders. *Journal of Neuropsychiatry Clinical Neuroscience*, *5*, 9-17.

Swogger, M. T., Walsh, Z., Lejuez, C. W., & Kosson, D. S. (2010). Psychopathy and risk taking among jailed inmates. *Criminal Justice and Behavior*, *37*, 439-452.

Therneau, T., Atkinson, B., & Ripley, B. (2015). *Recursive partitioning and regression trees* (R package v. 4.1-10). Retrieved from https://CRAN.R-project.org/package=rpart

Vose, B., Smith, P., & Cullen, F. (2013). Predictive validity and the impact of change in total LSI-R score on recidivism. *Criminal Justice and Behavior*, *40*, 1383-1396.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology*, *7*, 131-149.

Wilson, K., Juodis, M., & Porter, S. (2011). Fear and loathing in psychopaths: A meta-analytic investigation of the facial affect recognition deficit. *Criminal Justice and Behavior*, *38*, 659-668.

Pablo A. Ormachea, JD, MS, serves as codirector of the Center for Science and Law, where he manages several private and federally funded projects sharing the goal of constructing a cost-effective legal system with higher utility and lower cost. He holds a JD from Harvard Law School and an MS from Texas A&M University's Department of Statistics. His research focuses on building a scientific, data-driven criminal justice system that harnesses new technologies.

Brian K. Lovins, PhD, is the Assistant Director for Harris County Community Supervision and Corrections Department. He earned his PhD in Criminology from the University of Cincinnati, School of Criminal Justice. His recent publications focus on effective intervention for community corrections, the development and implementation of risk assessment, and the delivery of cognitive-behavioral interventions.

David M. Eagleman, PhD, is a neuroscientist at Stanford University and the founder and director of the Center for Science and Law in Houston, Texas. He is also the creator and host of PBS's *The Brain and the Author of Incognito: The Secret Lives of the Brain* (Pantheon, 2011). His areas of research include plasticity, time perception, vision, synesthesia, and the intersection of neuroscience with the legal system.

Sasha Davenport is the Director of Clinical Operations at the Center for Science and Law. She manages data collection efforts and staff development, and contributes to research involving forensic neuropsychology and psychometrics. She has a background in law and behavioral neuroscience, and her research interests include decision making in offender populations, juvenile justice, and criminal justice policy reform.

Anna Jarman is a former research fellow with the Center for Science and Law. She is currently a technology and investment analyst for Shell Technology Ventures, the corporate venture capital arm of Royal Dutch Shell and a master's student in Texas A&M University's Analytics program. Her research interests include comparative effectiveness and health policy, neuroscience, and applications of advanced analytics across industries.

Gabe Haarsma, PhD, is the Chief Data Scientist for the Center for Science and Law. He earned his PhD from Texas A&M University, and his research interests are in statistical inference, big data analysis, and data visualization.