

Examining the validity of the modified Taylor competitive reaction time test of aggression

Christopher J. Ferguson · Stephanie M. Rueda

Published online: 17 February 2009
© Springer Science + Business Media B.V. 2009

Abstract The use of behavioral tests of aggression has been a source of controversy for decades. Many critics of such measures note that validity studies for these measures are generally lacking. This manuscript describes two studies designed to test the validity of one of the most commonly used behavioral aggression measures, the modified Taylor competitive reaction time test (TCRTT). Participants in both studies were college undergraduates who were individually administered the TCRTT in a laboratory setting. In the first study ($n=103$), results on a standardized version of the modified TCRTT were examined for their convergent validity with expected measures of trait aggression and violent acts, including violent criminal behaviors and domestic violence. The second study ($n=101$) examined the validity of the modified TCRTT against neuropsychological outcomes that are predictive of impulsive violence. Both studies raised serious validity problems for the modified TCRTT. It is recommended that this measure not be adopted clinically as a behavioral measure of aggression and that other similar measures be more rigorously tested. Its use in research should be undertaken under advisement that results should not be generalized to serious acts of aggression or violence.

Keywords Aggression · Assessment · Mass media · Personality assessment · Violent behavior

Researchers have been examining the origins of violent behavior for decades. In order to test causal connections between external effects and violent outcomes, researchers have relied not on violent behaviors per se, but rather on a broader class of aggressive behaviors. Aggression has been defined as behavior intended to cause physical harm or humiliation to another organism which wishes to avoid the harm (Baron and Richardson 1994). Violent behaviors, by contrast, are typically restricted to acts which are intended to cause serious physical harm (for a discussion of methodological difficulties in operationally defining aggression see Savage 2004). Aggression as a class of behavior is much broader than violent behavior and can include numerous acts (i.e.,

C. J. Ferguson (✉) · S. M. Rueda
Department of Behavioral, Applied Sciences and Criminal Justice, Texas A&M International University, 5201 University Blvd., Laredo, TX 78041, USA
e-mail: CJFerguson1111@aol.com

giving research participants non-injurious ‘noise blasts’, insulting a person, writing a bad evaluation) which are neither physically injurious nor illegal. As such, aggressive behaviors can be studied in laboratory environments when violent behaviors cannot. Should these behavioral measures of aggression have high validity, it would be reasonable to conclude that their utility may extend beyond research settings into the clinical assessment of aggression and violence risk. However, behavioral measures of aggression have come under considerable criticism, both for the unstandardized way in which they are often employed (Ferguson 2007) and for the lack of controlled validity studies for these measures (Ritter and Eslea 2005; Tedeschi and Quigley 1996). Our study examined one of the most commonly used behavioral measures of aggression, the modified Taylor competitive reaction time test (TCRTT) in regards to its validity as a test of aggression.

The Taylor competitive reaction time test

The original version of the TCRTT (Epstein and Taylor 1967) was composed of participants playing a reaction time game against an ‘alleged’ human opponent, who, in reality, did not exist. Before each trial, the participant set an electric shock level, with the understanding that the opponent would receive that shock as punishment for losing. Alternatively, the participant would be shocked by the opponent if they themselves lost the competition. There was, in fact, no opponent, and the series of wins and losses were standardized as a means of provoking aggression in the participant. Several studies have supported the effectiveness of the electroshock version of the TCRTT as a measure of aggression (Giancola and Zeichner 1995; Taylor 1967), although the validity of the measure has also been questioned (Tedeschi and Quigley 1996).

The TCRTT was modified in later studies (e.g., Anderson and Dill 2000; Anderson and Murphy 2003) to use noise blasts instead of the electric shocks. Although the noise blasts are less aversive, they are easily adaptable to a computer-driven format and may raise fewer ethical concerns with institutional review boards. The noise blasts do not cause physical pain and may be less stressful either to be administered or received, leading to fewer ethical concerns than with electric shock. In all likelihood, however, the adoption of the ‘noise blast’ paradigm is as much practical as ethical. Unlike a shock machine, the variation in noise burst may be administered through a typical PC or Mac computer, requiring no additional machinery. The procedure is otherwise similar, with noise blasts serving as punishment for losing. These noise blasts can be varied in regards to both intensity and duration, thus producing multiple means of ostensibly measuring ‘aggression.’ This, in fact, has been one of the concerns raised by some researchers (e.g., Ferguson 2007): that there is no standardized measuring format for the modified TCRTT. The variable of ‘aggression’ can be measured through multiple methods. The varieties of total scores that can be derived are numerous. Given that different studies use different means of measurement (see Anderson and Dill 2000; Anderson and Murphy 2003; Bartholow et al. 2006; Carnagey and Anderson 2005 for four different ways of using the modified TCRTT), the opportunities for capitalization on chance are numerous. Indeed, this is unlikely to be a scenario unique to the TCRTT, and issues identified here might be true across numerous measures and research fields, greatly weakening the validity of much social science research. Researchers (or indeed

clinicians) could choose outcomes that best suit their hypothesis and ignore outcomes that do not. This issue was addressed by Ferguson et al. (2008), who developed a standardized and reliable version of the modified TCRTT. The validity of the measure remains to be adequately tested, however.

The most common use of the modified TCRTT in a criminological context is for studies examining the relationship between media violence and aggression or violent behavior. For example, regarding research on violent video-games, a majority of studies purporting to examine aggressive behavior experimentally used the modified TCRTT (Ferguson 2007). Criminologists have long speculated on the role of media violence as an agent causing aggressive or violent behavior (Surette 2007). In this regard, it may be little exaggeration to suggest that the modified TCRTT functions as a cornerstone of the causal argument for such a link, as it is the main experimental measure of aggressive behavior. The importance of our understanding the validity of the TCRTT for this research field, in particular, is evident.

Several studies (i.e., Anderson and Bushman 1997; Anderson et al. 1999; Giancola and Chermack 1998; etc.) suggest they provide evidence for the construct validity of the modified TCRTT as a measure of aggression. Yet, these studies typically use very indirect methodology to suggest that, as some studies of laboratory aggression effects have similar effects to some correlational studies of aggression, this is an indication of external validity. Yet, they actually provide no evidence that higher use of noise blasts is associated with any external indicator of aggression within individuals. In other words, evidence that the modified TCRTT predicts real world aggression or violence is absent. Several researchers have voiced these concerns regarding the modified TCRTT as well as other similar behavioral measures of aggression (Ferguson 2007; Ritter and Eslea 2005; Tedeschi and Quigley 2000, 1996). This paper describes the examination of convergent validity of the modified TCRTT through two studies. The first examined the convergent validity of the modified TCRTT with trait aggression and real world violent acts, including violent criminal acts and domestic violence. The second examined the correlation between the modified TCRTT and neuropsychological tests that have been demonstrated to predict aggression due to frontal lobe deficits. These two studies were designed to test for the validity of the modified TCRTT against both instrumental and hostile aggressive behaviors (Atkins et al. 1993; Buss 1961). Anastasi and Urbina (1996) discussed the issue of validity coefficients. They noted that validity coefficients as low as 0.2 or 0.3 were generally weak (although they may be used for some personnel selection purposes, which we did not believe applied here). Based on this discussion, we established a validity threshold of $r=0.40$ between the TCRTT and related outcome variables, which was used here as evidence for validity.

Study 1

Method

Participants

Participants included 103 young adults recruited from a Hispanic-serving public university in the south of the USA. Of these students 62 (60.2%) were male and 41

(39.8%) were female. Regarding ethnicity, 98 (95.1%) were Hispanic, three (2.9%) were (non-Hispanic) Caucasian and two (1.9%) declined to answer. These ethnicity data were reflective of the student body of the university. The use of a Hispanic majority sample necessitated limitations to the generalization of results. However, cautionary generalization may be attempted in this case, as previous authors of the study of aggression have asserted that similar aggression effects occur in a similar manner across population groups (see Grimes et al. 2008 for a discussion). The mean age of the sample was 23.6 years [standard deviation (SD)=5.82 years].

Materials

Demographic characteristics sheet On a single page, participants indicated their age, gender, self-described ethnicity, and education level.

Aggressive behavior This experiment used the modified version of the TCRTT (Anderson and Dill 2000; Ferguson et al. 2008). The modified TCRTT provides an opportunity for the participant to play a ‘reaction time game’ against a fictional opponent. Participants are asked to set the level of a noise blast that will serve as punishment for their competitor in the reaction time game. This noise blast can be varied, both in terms of intensity (loudness) and duration. There are 25 trials in the modified TCRTT, and the noise level and duration can be reset each time. For each of the 25 trials, participants are told that if they win, their opponent will hear the noise blast they have set, and, if they lose, they will hear a noise blast that their opponent has set for them. The pattern of wins and losses is actually preset in the computer, as there is no human opponent. The win and loss trials are standardized across all participants, regardless of reaction time. The first trial ends in loss for the participant, with the punishing noise blast set at maximum. This is designed to ‘provoke’ aggression from the participant. Noise blast levels range between 0 dB and 95 dB. Note that this is just over the United States Safety and Health Standards recommendations for *sustained 8-hour* exposure of 90 dB for full-time workers and is well under the pain threshold of 125 dB.

We used the internal consistency coefficient, alpha, of the 25 trials on the modified TCRTT to examine the reliability of this laboratory measure of aggression. The reliability of intensity scores was found to be high (alpha=0.94) for our sample. Coefficient alpha for the noise duration was likewise high (alpha=0.93). As such, reliability was not a problem for this standardized version of the modified TCRTT. The intercorrelation between the two measures was $r=0.76$, not as high as might have been expected, but high enough to indicate adequately that the two variables were tapping into the same construct.

Trait aggression To measure trait aggressiveness, we asked participants to complete the aggression questionnaire(AQ) short form (Buss and Warren 2000). The shortened version of the AQ consists of the summed scores of the first 15 items of the original 34-item version and was designed to measure the degree to which respondents endorse statements about their levels of aggression. Participants score the items, using a 5-point Likert scale, ranging from “not at all like me”, to “completely like me,” with higher scores indicating more aggressiveness. An

example item is, “At times I get very angry for no good reason.” Based on the normative sample reported in the manual, the AQ obtained an alpha coefficient of 0.90 for the total score. The AQ has been demonstrated to have good predictive validity (Felsten and Hill 1999) and convergent validity with other measures of trait aggression (Garcia-Leon et al. 2002). Within our sample, the AQ obtained an alpha coefficient of 0.85.

Violent criminal behavior Measurement of self-reported violent crime was obtained with the National Youth Survey (Elliot et al. 1985), a measure first developed in conjunction with the National Institute of Mental Health. This measure is a 45-item self-report measure of violent and nonviolent crimes in which individuals are asked to estimate how many times they have committed those behaviors. A violent crime index was derived from the sum of 11 items related to violent crime commission. Individual items were transformed into z-scores that equalized variance prior to summation. Items on this scale include estimates of how often in the past a respondent has committed acts such as “hit a parent or caregiver” or “attacked/seriously injured someone on purpose”. Coefficient alpha for this 11-item index of total past commission of violent crime with our sample was 0.83. Previous studies (e.g., Anderson and Dill 2000; Ferguson et al. 2008) have found this violence index to be a reliable and valid measure.

Domestic violence perpetration The conflict tactics scale (CTS; Straus et al. 1996) is one of the most widely used measures of domestic violence perpetration. Respondents report on the frequency with which they both commit and are victims of a wide range of physical assaults, psychological abuse and sexual coercion. Measures of perpetration of domestic violence included in our study were scales for physical assault of partner, from which one poorly inter-correlated item (“grabbed partner” #45) was dropped ($\alpha=0.64$), and psychological abuse ($\alpha=0.75$). The alpha values reported here were for our sample.

Procedure Participants were tested in a standardized laboratory setting. They were given an informed consent form and told that they would be playing a reaction time game against a human opponent. The questionnaires were administered first, followed by the modified TCRTT. Following the procedure, participants were debriefed, informed of the deception in the modified TCRTT, queried for suspiciousness and invited to ask questions. All procedures were designed to comply with American Psychological Association (APA) standards for the ethical treatment of human participants and passed before the relevant institutional review boards (IRBs).

Results

Means and standard deviations for all included measures are presented in Table 1.

In examining the validity of the TCRTT, we used the validity threshold of $r=0.40$ (see Anastasi and Urbina 1996 for a discussion) as the criterion for evidence of validity. The use of this criterion focuses on an appropriate estimate of the effect

Table 1 Means and standard deviations for outcome measures (*WAIS* Wechsler adult intelligence schedule)

Measure	Mean	Standard Deviation
Study 1		
TCRTT intensity	6.12	1.99
TCRTT duration	5.14	1.88
Trait aggression	28.20	9.40
Violent crime	-0.16	6.50
Physical assault	1.81	5.64
Psychological abuse	13.86	23.07
Study 2		
1. TCRTT intensity	6.08	1.50
2. TCRTT duration	3.00	1.21
3. Trait aggression	30.35	9.48
4. Stroop interference	53.37	8.21
5. Trails executive score	25.31	16.39
6. WAIS	104.27	12.52

size, rather than on statistical significance. Statistical significance is easily swayed by sample size, resulting either in the rejection of perfectly acceptable validity coefficients in small samples, or the citing of unacceptable validity coefficients as ‘evidence’ for validity as very small effects become statistically significant due to large sample size. Thus, in the interpretation of results, significance or non-significance is of comparatively little value, whereas interpretation of effect size is inherently more valuable (see Cohen 1994).

We examined correlations between the intensity and duration measures of the modified TCRTT and trait aggression, violent criminal acts and domestic violence, as well as gender, using simple bivariate correlations. The results are displayed in Table 2. As can be seen from the results, the intensity and duration measures of the

Table 2 Bivariate correlations between the TCRTT and aggression/violence outcomes

Variable	1	2	3	4	5	6	7
1. TCRTT intensity	1.00	0.76**	-0.09	0.15	-0.05	0.13	-0.07
2. TCRTT duration		1.00	-0.08	0.13	0.01	0.11	0.00
3. Gender			1.00	-0.08	-0.19*	0.15	0.07
4. Trait aggression				1.00	0.17	0.21*	0.25**
5. Violent crime					1.00	0.00	0.14
6. Physical assault						1.00	0.18
7. Psychological abuse							1.00

* $P < 0.05$

** $P < 0.01$

modified TCRTT were not related to any violent outcomes, and coefficients consistently fell beneath $r=0.40$.

As male individuals engage in higher amounts of aggressive and violent behavior (Archer and Coyne 2005), one reasonable suggestion for the apparent low validity of the modified TCRTT might be differential validity. In other words, it is possible that, as aggression (or at least directly aggressive) and violent behaviors are comparatively uncommon among women, aggression measures such as the modified TCRTT are not valid for use on female subjects. For men, among whom direct aggression and violence is more common, the modified TCRTT may, nonetheless, be valid. To test this, we once again ran bivariate correlations between the modified TCRTT intensity and duration measures against trait aggression, violent behaviors and executing functioning to test the validity of the modified TCRTT for men specifically. The results are presented in Table 3. Once again, the interference and duration scores of the modified TCRTT correlated well with each other ($r=0.76$) but not with any variables related to aggression or violence. With women, the TCRTT intensity score was moderately correlated with domestic physical assaults as measured by the CTS ($r=0.34$), although this still fell under the validity threshold (of $r=0.40$). Neither the duration nor the intensity score was significantly correlated with other outcomes, however, including violent crimes.

Discussion

Results of the first study suggest that modified TCRTT performance is not related to aggression or violent acts. The two measures taken from the modified TCRTT correlated well with each other, suggesting that they were both tapping into the same construct; however, the construct that they tapped into appeared to be unrelated to the aggression for which the modified TCRTT is primarily employed and on the population of young college adults upon which it is most commonly employed (e.g., Anderson and Dill 2000; Bartholow et al. 2006; etc.). This proved to be true for young adults in general, as well as for male subjects specifically.

Although the original version of the TCRTT (Epstein and Taylor 1967) has been the source of some controversy regarding the validity of its electroshocks to measure

Table 3 Bivariate correlations between the TCRTT and aggression/violence outcomes (by gender). Results for men appear above the diagonal line; those for women appear below the diagonal line

Variable	1	2	3	4	5	6
1. TCRTT intensity	1.00	0.76**	0.14	-0.07	-0.06	-0.02
2. TCRTT duration	0.74	1.00	0.14	-0.01	-0.03	0.13
3. Trait aggression	0.13	0.11	1.00	0.19	0.11	0.20
4. Violent crime	-0.08	0.09	0.10	1.00	0.07	0.22
5. Physical assault	0.34*	0.31	0.34*	-0.04	1.00	0.15
6. Psychological abuse	-0.14	-0.24	0.36*	-0.13	-0.21	1.00

* $P<0.05$

** $P<0.01$

aggression (Giancola and Chermack 1998; Ritter and Eslea 2005; Tedeschi and Quigley 1996), results on this modified noise-blast version are perhaps more disappointing. It is possible that, in further removing the protocol from the actual simulation of causing harm, the validity has been further reduced.

One possible caveat to our results is that perhaps our study, focused as it was on trait aggression and the relative frequency of violent events, missed a more latent propensity for violence that may be assessed by the modified TCRTT. Perhaps the modified TCRTT is not effective at detecting instrumental/trait aggression, but it could be argued that impulsive/hostile aggression may still be measured by the modified TCRTT. This possibility bears further investigation, as does the subject of the second study. A further caveat is that cultural issues and perception of violence might have influenced some results, such as those for domestic violence. Although there was no clear indication that this was the case, and previous researchers have argued against wide differences between groups in aggression effects (see Grimes et al. 2008), cultural comparisons may make a worthy avenue for future research.

Although, in the first study, the modified TCRTT was not correlated with actual violent acts, it remains possible that the modified TCRTT could be related to poor impulse control, which could lead to violence. There has been previous research to suggest that frontal lobe deficits associated with poor executive functioning might be, in part, responsible for aggressive or antisocial behavior (Hare 1993). The effect of this executive functioning deficit on violence has been found in both mentally ill (Kumari et al. 2006) populations and in those not mentally ill (Soderstrom et al. 2002). Donovan and Ferraro (1999) found that measures of executive functioning such as the Stroop and the Trails B test distinguished domestic violence perpetrators from a matched sample of non-violent controls. It has been theorized that low cortical arousal in the frontal lobes results in deficits in executive functioning, which, in turn, limit control of aggressive and violent impulses (Elliot and Mirsky 2002; Hare 1993). A catalyst model of aggression (Ferguson et al. 2008) has been suggested as an evolutionary model of violence and aggression. Among other things, this model has suggested that individuals with impaired executive functioning are more prone to aggressive acts.

As such, and given that measures of executive functioning are predictive of violent behavior, it would be expected that valid behavioral measures of aggressive behavior should demonstrate some associated relationship with measures of executive functioning. Thus, perhaps, the modified TCRTT is a better measure of impulsive/hostile aggression than it is of instrumental/trait aggression. This study was designed to examine that possibility.

Study 2

Method

Participants

Participants included 101 young adults recruited from two public universities in the midwest and south of the USA. Of these students, 46 (45.5%) were male and 55 (54.5%) were female. Regarding ethnicity, 42 (41.6%) were Caucasian, 49 (48.5%) were

Hispanic, seven (6.9%) were African–American, two (2%) were Asian and one (1%) was listed as “other.” The mean age of the sample was 23.9 years ($SD=3.70$ years).

Materials

Demographic characteristics sheet On a single page, participants indicated their age, gender, self-described ethnicity, and education level.

Aggressive behavior The modified TCRTT, as described above, was used as the measure of aggressive behavior. For our sample, internal consistency of the intensity measure was $\alpha=0.90$, and, for the duration measure, $\alpha=0.98$). However, in this sample, the intensity and duration measures were not highly correlated ($r=0.29$), raising some concern for the compatibility of the two measures.

Trait aggression As in the first study, the AQ was used as a measure of trait aggression. Within the second study’s sample, the AQ obtained an alpha coefficient of 0.85.

Executive functioning Executive functioning and planning associated with low cortical arousal in the frontal lobe and aggression were measured with the Stroop color and word test (Stroop) (Golden and Freshwater 1998). The Stroop test presents information to participants in three formats, black and white printed words (red, green, blue), colored Xs, and colored printed words (red, green, blue). Participants are asked either to read the words aloud or to state the color of the ink that the words are printed in, aloud and as quickly as they can. This test measures a participant’s ability to select appropriate stimuli and eliminate distraction. Test–retest reliability studies for the Stroop test range between 0.70 and 0.89. Low interference scores in the Stroop test have been associated with brain injuries, including in the prefrontal cortex (Golden and Freshwater 1998).

Executive functioning/mental flexibility The second measure used in this study for executive function is the trail making test, versions A and B (TrailsA, TrailsB) (Reitan and Wolfson 1985). The trails test requires participants to connect numbered (TrailsA) or interchanging numbered and lettered (TrailsB) dots on a page of paper as quickly as possible. These tests are designed to measure attention, mental flexibility and visual search functions. Numerous studies have reported satisfactory inter-rater and alternate forms reliability for the trail making test (see Spreen and Strauss 1998: 535 for a full discussion). The trails tests have been found to be valid indicators of brain damage (Leininger et al. 1990) and frontal lobe deficits (Lezak 1983; D’Esposito et al. 1996). Measures of mental flexibility and executive functioning include time score on TrailsB, as well as the difference between the time score on TrailsB and TrailsA (referred to below as the executive score). Higher scores are indicative of greater dysfunction.

Intelligence All participants were assessed for general cognitive ability with the verbal intelligence scale portion of the Wechsler adult intelligence schedule (WAIS) (Wechsler 1997). The testing manual for this cognitive test reports good test–retest and coefficient alpha reliability as well as a number of supportive validity studies for the verbal portion of this test. Elliot and Mirsky (2002) note that low verbal intelligence is

associated with violent criminal behavior, although intelligence tests are likely less sophisticated in predicting violence than are tests of executive functioning. This measure is used here as an overall indication of cognitive functioning, and it is expected that valid measures of aggression should show some degree of negative correlation with verbal intelligence, if smaller than for other measures.

Procedure

Participants were tested in a standardized laboratory setting. They were given an informed consent form and told that they would be playing a reaction time game against a human opponent. The executive functioning measures were administered first, along with the AQ and the WAIS, followed by the modified TCRTT. Following the procedure, participants were debriefed, informed of the deception in the modified TCRTT, queried for suspiciousness and invited to ask questions. All procedures were designed to comply with APA standards for the ethical treatment of human participants and had been passed before relevant IRBs.

Results

Means and standard deviations for all included measures are presented in Table 1.

The modified TCRTT intensity and duration scores were correlated against the executive functioning measures (Stroop and trails), the WAIS, and the AQ. Results are presented in Table 4. Results for the modified TCRTT intensity score were better for the second sample than for the first. This measure did not relate to executive functioning, but it did correlate with trait aggression and gender in the expected directions for an aggression measure. All convergent correlations were small, however, lower than the $r=0.40$ threshold.

Modified TCRTT duration scores demonstrated more problematic results. The duration measure did correlate with the trails executive score as well as verbal intelligence. However, in both cases, the correlations were in the opposite direction

Table 4 Bivariate correlations between the TCRTT and executive/cognitive outcomes. Trails executive score = TrailsB–TrailsA. The Stroop and trails scores work at inverse with each other; thus, a negative correlation is expected

Variable	1	2	3	4	5	6	7
1. TCRTT intensity	1.00	0.29**	0.25**	0.02	0.01	−0.03	−0.30**
2. TCRTT duration		1.00	0.02	0.09	−0.21*	0.33*	−0.22*
3. Trait aggression			1.00	0.04	0.00	0.18*	−0.25**
4. Stroop interference				1.00	−0.19*	0.09	−0.11
5. Trails executive Score					1.00	−0.16	−0.03
6. WAIS						1.00	−0.28**
7. Gender							1.00

* $P < 0.05$

** $P < 0.01$

from expected. In other words, individuals who used longer durations (ostensibly indicative of aggression) were less impulsive and higher in intelligence. Although it should be noted that these results as well do not cross the $r=0.40$ threshold, they are worrisome results for the validity of the modified TCRTT duration measure.

As with the first study, it was possible that male participants would show differential validity in contrast to female participants. Once again, the analysis was limited to male participants only. Results, as presented in Table 5, were actually less encouraging for male participants alone. More consistent with study 1, neither modified TCRTT intensity nor duration was correlated with trait aggression. The modified TCRTT maintained its mis-directed relationships with the trails executive score and verbal intelligence scores.

Discussion

Results from study 2 provided little encouragement for the use of the modified TCRTT as a measure of aggression. For the entire sample, the modified TCRTT interference score was related to trait aggression, but this relationship appeared to have been fueled by female participants and did not hold for young men, the population most likely to act aggressively. Modified TCRTT duration scores did correlate with trails executive score and verbal intelligence, but in the opposite direction from that which would have demonstrated good validity. Thus, the modified TCRTT cannot be said to be a measure of impulsive aggression. It may be that, for women, the modified TCRTT intensity score may tap into a construct related to aggression, competitiveness perhaps. Yet, this relationship is weak and does not hold for men.

General discussion

Taken together, the results from study 1 and study 2 provided little support for the convergent validity of the modified TCRTT as a measure of aggression. Consistently

Table 5 Bivariate correlations between the TCRTT and executive/cognitive outcomes (by gender). Trails executive score = TrailsB–TrailsA. The Stroop and trails scores work at inverse with each other; thus, a negative correlation is expected. Results for male participants appear above the diagonal line, those for female participants appear below the diagonal line

Variable	1	2	3	4	5	6
1. TCRTT intensity	1.00	0.33*	0.09	0.09	0.07	-0.16
2. TCRTT duration	0.12	1.00	-0.10	0.07	-0.22	0.42**
3. Trait aggression	0.26	0.05	1.00	0.12	0.09	0.14
4. Stroop interference	-0.16	0.05	-0.12	1.00	-0.28*	0.19
5. Trails executive score	-0.07	-0.21	-0.11	-0.06	1.00	-0.19
6. WAIS	-0.09	-0.02	0.11	-0.18	-0.16	1.00

* $P<0.05$

** $P<0.01$

across both studies, the modified TCRTT failed to perform as expected of a valid measure of aggression. The modified TCRTT was not sufficiently associated with violent criminal behaviors, domestic violence, or executive functioning measures that have previously been found to be predictive of aggression and violence. In study 2 the modified TCRTT appeared to be correlated with trait aggression (although not at the $r=0.40$ threshold), but this relationship did not hold for young men, arguably the population at greatest risk for aggression.

Given the results of this study, it is recommended that research studies which use the modified TCRTT as a laboratory measure of aggression interpret their results with some caution. Specifically, results from the modified TCRTT should not be extended to serious aggressive or violent acts. The modified TCRTT has not seen clinical use as a measure of aggressiveness. Had this study been more promising in its results, it could have been recommended that the modified TCRTT be used in some clinical settings. Given its ease of use, public domain status and relatively quick administration (15 minutes), it could easily have been fitted into a violence risk assessment. However, given the results, such clinical use is highly not recommended. The modified TCRTT failed to demonstrate validity on the population for which it was intended. There seems little reason to believe that adapting the modified TCRTT to clinical populations would be feasible. There is, of course, no particular movement to use the modified TCRTT in clinical settings. If the measure were valid, it would be reasonable to ask why there is not such a movement. Results from this study provide some indication, namely that the measure simply has limited clinical utility. Although clinical standards might arguably be higher than those for research studies, given the extent to which clinical implications are made from research results using this measure (i.e., that watching violent television increases pathologically aggressive behavior), a discussion of clinical utility is warranted. The issue here is less that there is a meaningful clinical movement to adopt the TCRTT and more to note that the failure of a *social* measure of aggression to meet *clinical* standards of validity is a serious issue, particularly when clinically relevant conclusions about pathological aggression are being made based on results obtained from this measure. Given the results of this study, it is recommended that making pathological conclusions based on this aggression measure should be revised. Further, given that ethical concerns may persist regarding the deliverance of noise blasts, even if non-painful, the benefits of this procedure appear not to outweigh potential concerns.

Caution should be undertaken to note that these results should not be generalized to all aggression measures. Aggression measures that more consistently adhere to common definitions of aggression (e.g., Baron and Richardson 1994) may demonstrate greater efficacy.

Why does the modified TCRTT not work?

As other authors have noted (Ritter and Eslea 2005; Tedeschi and Quigley 1996, 2000), developing behavioral measures of aggression has proven to be difficult. There are several reasons why the TCRTT may fail to demonstrate adequate validity, and these issues may prove relevant to the development of future behavioral aggression measures.

Behaviors are not 'proxy' enough to actual aggression The first issue, and perhaps most evidence, is that the modified TCRTT, like other attempts to measure the aggression construct, may use behaviors that are too distant from actual aggressive behavior. As noted earlier, the noise blasts used in the modified TCRTT are obviously (to the participant) not harmful, and so the participant has no real expectation of causing actual harm to another individual, no matter how loud the blasts are set. Perhaps even more critically, the individual has no reason to expect that the hypothetical opponent is attempting to avoid the harm.

Konijn et al. (2007) attempted to fix this potential weakness by informing child participants that the highest level blasts (e.g., 8,9,10) were potentially damaging to hearing. However, it remains unclear whether participants believed this to be true, particularly since the participants themselves were exposed to noise blasts that were clearly not harmful. Participants may also find it perplexing that their 'harmed' opponents never complain, are heard to cry out, or cease participating in the procedure. Participants may also not believe that an authority figure (i.e., the examiners) would actually allow them to cause harm. Furthermore, the revision by Konijn et al. (2007) attempts to fix the validity problem by introducing yet another unstandardized version of the modified TCRTT without providing additional data that validated the effectiveness of their version. Lastly, this measure cannot be said to be a proxy for violent behavior, because, unlike in the real world, there is no physical danger (did the participants believe that they might receive damaging noise blasts too? It is unlikely, as this would seem to be unethical), nor repercussions for violence (i.e., legal or social sanctions).

Absence of physical, legal or social sanctions Aggression measures in the laboratory perform as poor stand-ins for violence, as there are no consequences for the 'aggressive' acts. Other scholars have commented that, in fact, participants may feel that their actions are sanctioned, or even demanded, by the research examiner (Ritter and Eslea 2005; Savage 2004). Violence in the real world carries risks of physical harm, legal repercussions and social sanctions. Participants in laboratory experiments experience none of these.

Absence of alternatives to non-aggressive behavior Ideally, a measure of aggressive behavior would allow individuals the choice between aggressive and non-aggressive alternatives. For such measures, aggressive individuals would be expected to choose aggressive alternatives more often than would non-aggressive individuals (although this would have to be validated, of course). However, the modified TCRTT does not allow for non-aggressive alternatives for dealing with provocation. In other words, options to respond to provocation through means other than aggression (i.e., diplomacy, withdrawal) are not allowed on the modified TCRTT (a noise blast may be set at zero, but this is simply ignoring a provocation, not 'dealing' with it). This may effectively set up 'demand characteristics', in effect shunting even non-aggressive individuals into the direction of behaving with more aggressive responses. By limiting the repertoire of potential behavioral responses to provocation, the modified TCRTT becomes isolated from real world behavior. In other words, individuals taking the modified TCRTT are forced to respond to provocation differently from how they might respond in the real world, thus reducing validity.

Absence of a clinical cut-off point Unlike most effective clinical measures of psychopathology [such as, for example, the Minnesota multiphasic personality inventory (MMPI-2) or the Beck depression inventory], the modified TCRTT provides no clinical cut-off value for ‘aggression’. In other words, the modified TCRTT does not provide guidelines for the type of score that might indicate a highly aggressive individual. Although aggression exists along a continuum, so, in effect, do most variables related to psychopathology (e.g., depression or anxiety). However, certainly, levels of these constructs are known to be related to negative outcomes. The same, likely, is true for aggression.

As a matter of contrast, the MMPI-2 (Hathaway and McKinley 1989) provides t-score cut-off values of 65, above which scores are highly indicative of some form of psychopathology. Although MMPI-2 responses are to be judged in accordance with other clinical information, such scores have been empirically demonstrated to be associated with elevated risk of psychopathology (Hathaway and McKinley 1989). Should one person obtain a t-score of 45 on, say the MMPI-2 2-scale (depression), and another person obtain a score of 55, it would be concluded that neither is at particularly high risk for mood-related psychopathology, despite the difference in their scores, because neither score crosses the clinical cut-off value. By contrast on the modified TCRTT, no such clinical cut-off points exist. Should, for example, one person use average noise blast intensities of 4.0 (out of 10), and a second person 8.0, we have no evidence that the second person’s higher score is indicative of higher aggression, as no clinical cut-off values are provided. There is no evidence currently that even maximal scores (10 out of 10 average intensity) are particularly indicative of higher aggression risk. As a related issue, the sensitivity and specificity (see Ferguson and Negy 2006) of the modified TCRTT remain unknown.

Lack of standardized use This issue has been mentioned previously in research (e.g., Ferguson 2007), but it bears repeating here. The utility of the modified TCRTT has been limited by its unstandardized usage in the literature. No manual exists for the modified TCRTT, differing authors use differing instructions in giving the modified TCRTT (e.g., Anderson and Dill 2000; Konijn et al. 2007) and, as noted earlier, use differing measures of aggression from the modified TCRTT. Naturally, a test cannot be valid until it is reliable, and it cannot be reliable unless it is standardized. Future behavioral aggression measures would best focus early on standardized use.

Concluding remarks

Designing workable aggression measures for use in laboratory paradigms is a valuable undertaking. As noted by Ritter and Eslea (2005), recent attempts to improve designs such as the modified TCRTT have proven to suffer similar validity problems as the modified TCRTT. Future designs may benefit from our looking at ways wherein participants have the opportunity to aggress, but are not explicitly invited to do so. Similarly, laboratory designs often provide considerable distance between the participant and their ‘victim’ (i.e., having an opponent in another room or otherwise out of sight), whereas physical aggression or violence in the real world typically takes place face-to-face. Developing laboratory paradigms that allow for

greater face-to-face contact between participants may help in increasing the validity of such paradigms.

Research which has used the modified TCRTT to make conclusions about serious aggression or violence (e.g., Anderson and Dill 2000) should be re-examined, as conclusions based upon this measure may be seriously flawed. In a broader context, social scientists need to exercise greater care in generalizing results from 'proxy' measures to real-world phenomena. In certain areas, such as media violence research, basic tenets of good measurement appear to have been abandoned. To the extent that public policy debates may focus on these research findings (Grimes et al. 2008), this is an issue of serious concern.

The studies discussed here are not without limitations. Both studies consisted of homogeneous college-students samples, although this is the most common use for the modified TCRTT in most other studies as well. Study 1 employed predominantly Hispanics, limiting the generalizability of the study to other ethnic groups. Further research regarding the validity of the modified TCRTT is certainly warranted.

Results from our study suggest limitations in the use of the modified TCRTT as a behavioral measure of aggression. Problems with the modified TCRTT may prove to be endemic to behavioral aggression measures in general, which, some authors have noted, have widespread validity problems (Ritter and Eslea 2005; Tedeschi and Quigley 1996, 2000). It may be that the designing of behavioral measures of aggression that are valid and effective, while also mindful of ethical restraints, is an unfeasible task. When such efforts are undertaken in the future, they would likely be enhanced by focusing on standardization and reliability early in the design process, provide non-aggressive behavioral alternatives, and validate clinical cut-off points that may be illustrative of actual clinically significant aggression, rather than small fluctuations within the normal range of aggression.

References

- Anastasi, A., & Urbina, S. (1996). *Psychological testing*. New York: Prentice Hall.
- Anderson, C. A., & Bushman, B. J. (1997). External validity of "trivial" experiments: The case of laboratory aggression. *Review of General Psychology, 1*, 19–41.
- Anderson, C., & Dill, K. (2000). Video games and aggressive thoughts, feelings and behavior in the laboratory and in life. *Journal of Personality and Social Psychology, 78*, 772–790.
- Anderson, C., & Murphy, C. (2003). Violent video games and aggressive behavior in young women. *Aggressive Behavior, 29*, 423–429.
- Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science, 8*, 3–9.
- Archer, J., & Coyne, S. (2005). An integrated review of indirect, relational and social aggression. *Personality and Social Psychology Review, 9*, 212–230.
- Atkins, M., Stoff, D., Osborne, M., & Brown, K. (1993). Distinguishing instrumental and hostile aggression: Does it make a difference? *Journal of Abnormal Child Psychology*. Retrieved 2/8/07 from http://www.findarticles.com/p/articles/mi_m0902/is_n4_v21/ai_13240551.
- Baron, R., & Richardson, D. (1994). *Human aggression*. New York: Plenum Press.
- Bartholow, B., Bushman, B., & Sestir, M. (2006). Chronic violent video game exposure and desensitization to violence: Behavioral and event-related brain potential data. *Journal of Experimental Social Psychology, 42*, 532–539.
- Buss, A. (1961). *The psychology of aggression*. New York: Wiley.
- Buss, A. H., & Warren, W. L. (2000). *Aggression Questionnaire manual*. Los Angeles: Western Psychological Services.

- Carnagey, N., & Anderson, C. (2005). The effects of reward and punishment in violent video games on aggressive affect, cognition and behavior. *Psychological Science, 16*, 882–889.
- Cohen, J. (1994). 'The earth is round ($p < .05$).'. *American Psychologist, 49*, 997–1003.
- D'Esposito, M., Alexander, M., Fischer, R., McGlinchey-Berroth, R., & O'Connor, M. (1996). Recovery of memory and executive function following anterior communicating artery aneurysm rupture. *Journal of the International Neuropsychological Society, 2*, 565–570.
- Donovan, W., & Ferraro, R. (1999). Frontal lobe deficits in domestic violence offenders. *Genetic, Social and General Psychology Monographs, 125*, 71–102.
- Elliot, A., & Mirsky, A. (2002). Cognitive antecedents of violence and aggression. In J. Glicksohn (Ed.), *The Neurobiology of Criminal Behavior*. Dordrecht, Netherlands: Kluwer.
- Elliot, D., Huizinga, D., & Ageton, S. (1985). *Explaining delinquency and drug use*. Beverly Hills, CA: Sage.
- Epstein, S., & Taylor, S. (1967). Instigation to aggression as a function of degree of defeat and perceived aggressive intent of the opponent. *Journal of Personality, 35*, 265–289.
- Felsten, G., & Hill, V. (1999). *Aggression Questionnaire* hostility scale predicts anger in response to mistreatment. *Behavior Research and Therapy, 37*, 87–97.
- Ferguson, C. J. (2007). Evidence for publication bias in video game violence effects literature: A meta-analytic review. *Aggression and Violent Behavior, 12*, 470–482.
- Ferguson, C. J., & Negy, C. (2006). Development and preliminary validation of a defendant and offender screening tool for psychopathology in inmate populations. *Criminal Justice and Behavior, 33*, 1–22.
- Ferguson, C. J., Rueda, S., Cruz, A., Ferguson, D., Fritz, S., & Smith, S. (2008). Violent video games and aggression: Causal relationship or byproduct of family violence and intrinsic violence motivation? *Criminal Justice and Behavior, 35*, 311–332.
- Garcia-Leon, A., Reyes, G., Vila, J., Perez, N., Robles, H., & Ramos, M. (2002). The Aggression Questionnaire: A validation study in student samples. *Spanish Journal of Psychology, 5*, 45–53.
- Giancola, P. R., & Zeichner, A. (1995). Construct validity of a competitive reaction-time aggression paradigm. *Aggressive Behavior, 21*, 199–204.
- Giancola, P. R., & Chermack, S. T. (1998). Construct validity of laboratory aggression paradigms: A response to Tedeschi and Quigley (1996). *Aggression and Violent Behavior, 3*, 237–253.
- Golden, C., & Freshwater, S. (1998). *Stroop color and word test: A manual for clinical and experimental uses*. Wood Dale, IL: Stoelting.
- Grimes, T., Anderson, J., & Bergen, L. (2008). *Media violence and aggression: Science and ideology*. Thousand Oaks, CA: Sage.
- Hare, R. (1993). *Without conscience*. New York, NY: Guilford Press.
- Hathaway, S., & McKinley, J. (1989). *MMPI-2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press.
- Konijn, E. A., Nije Bijvank, M., & Bushman, B. J. (2007). I wish I were a warrior: The role of wishful identification in effects of violent video games on aggression in adolescent boys. *Developmental Psychology, 43*, 1038–1044.
- Kumari, V., Aasen, I., Taylor, P., Ffytche, D., Das, M., Barkataki, M., et al. (2006). Neural dysfunction and violence in schizophrenia: An fMRI investigation. *Schizophrenia Research, 84*, 144–164.
- Leininger, I., Gramling, S., & Farrell, A. (1990). Neuropsychological deficits in symptomatic minor head injury patients after concussion and mild concussion. *Journal of Neurology, Neurosurgery and Psychiatry, 53*, 293–296.
- Lezak, M. (1983). *Neuropsychological assessment*. New York, NY: Oxford University Press.
- Reitan, R., & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery*. Tucson, AZ: Neuropsychology Press.
- Ritter, D., & Eslea, M. (2005). Hot sauce, toy guns and graffiti: A critical account of current laboratory aggression paradigms. *Aggressive Behavior, 31*, 407–419.
- Savage, J. (2004). Does viewing violent media really cause criminal violence? A methodological review. *Aggression and Violent Behavior, 10*, 99–128.
- Soderstrom, H., Hultin, L., Tullberg, M., Wikkelsö, C., Ekholm, S., & Forsman, A. (2002). Reduced frontotemporal perfusion in psychopathic personality. *Psychiatry Research: Neuroimaging, 114*, 81.94.
- Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests*. Oxford: Oxford University Press.
- Straus, M., Hamby, S., Boney-McCoy, S., & Sugarman, D. (1996). The revised Conflict-Tactics Scale: Development and preliminary psychometric data. *Journal of Family Issues, 17*, 283–316.
- Surette, R. (2007). *Media crime and criminal justice: Images, realities and policies*. Belmont, CA: Wadsworth.

- Taylor, S. P. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality, 35*, 297.
- Tedeschi, J., & Quigley, B. (1996). Limitations of laboratory paradigms for studying aggression. *Aggression & Violent Behavior, 2*, 163–177.
- Tedeschi, J., & Quigley, B. (2000). A further comment on the construct validity of laboratory aggression paradigms: A response to Giancola and Chermack. *Aggression & Violent Behavior, 5*, 127–136.
- Wechsler, D. (1997). *WAIS-III: Administration and scoring manual*. San Antonio: The Psychological Corporation.

Christopher J. Ferguson, Ph.D. is an assistant professor of clinical and forensic psychology at Texas A&M International University, USA. He holds a doctorate in clinical psychology from the University of Central Florida, USA. His research has focused on the etiology of violent criminal behavior and the positive and negative influences of playing violent video-games.

Stephanie M. Rueda is currently a Masters student in counseling psychology at Texas A&M International University, Laredo, Texas, USA. Her current research involves the influence of alcohol on aggressive behavior.