

The Depravity Standard for Violent Crimes

International Journal of
Offender Therapy and
Comparative Criminology
1–24

© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: [10.1177/0366624X211058959](https://doi.org/10.1177/0366624X211058959)
journals.sagepub.com/home/ijo



Michael Welner¹, Kate Y. O’Malley¹ ,
James Gonidakis¹, and Samantha Blair¹

Abstract

In violent crime cases, aggravating factors in United States criminal codes, such as “heinous,” “atrocious,” or “depraved,” are used to distinguish elements of the crime warranting more severe sentencing. These aggravating terms are vaguely defined and applied arbitrarily in violent cases. This paper details the development of a 25 item Depravity Standard to operationalize an evidence-based approach to distinguishing the worst of violent crimes. The items were applied to 393 detailed case files drawn from several American jurisdictions to develop and refine the item definitions, determine interrater reliability, and mine for the frequency of each item’s occurrence. This information was combined with 1,590 participant responses ranking the relative depravity of each item to develop a straightforward scoring system for measuring depravity in violent cases. The Depravity Standard guide can seamlessly be applied to identify the worst violent crimes, and provide support for those cases that may deserve leniency or early-release.

Keywords

depravity standard, sentencing, parole, HAC statutes, violent crime

Introduction

Violence as a behavioral expression has many diverse manifestations and degrees of severity. For those whose violence leads to criminal prosecution, aggravating factors of the offense heighten the charges one faces and can lead to significantly more severe sentencing if convicted. What “aggravates” a crime, makes one robbery worse than another, one assault worse than another, one charge of menacing worse than another, also engages the depravity, or the severity, of a crime (Welner,

¹The Forensic Panel, New York City, NY, USA

Corresponding Author:

Michael Welner, The Forensic Panel, 515 N. Flagler Dr., West Palm Beach, FL 33401, USA.
Email: drwelner@forensicspanel.com

O’Malley, Gonidakis, & Tellalian, 2018). Any experienced law enforcement or justice professional readily appreciates that some violent crimes are more depraved than others.

How to classify violence, and by extension violent crime, is a long-standing interest of the scientific community. Some researchers have found it useful to determine violence as either predatory or reactionary/affective (Crick & Dodge, 1996; Meloy, 2006; Weinshenker & Siegel, 2002). Affective violence is preceded by high levels of arousal, is accompanied by intense feelings, and is characterized by impulsivity and hot-bloodedness to a real or perceived threat or fear (Weinshenker & Siegel, 2002). Predatory violence is characterized by the absence of threat and is planned in the absence of emotion as a goal-directed attack (Weinshenker & Siegel, 2002). Some violence may be a mixture of predatory and impulsive qualities.

Diagnostic distinctions can also inform investigative directions of the motives of violence. Whereas antisocial violence is instrumental in its intent, as well as the indifference of the actor, narcissistic features contribute to the grandiosity of the motivation, and paranoia feeds into prejudice motives. Violence is also distinguished for its intended effect on the victim. Sadistic violence encompasses intended and effected torture, as well as the relish of the victim’s suffering.

The law accounts for how some violent crimes are worse than others, and that there are varying levels of violence even within crime types. Most states separate crimes by level of severity such as Assault first, Assault second, or Assault third. These levels also determine whether a crime is a misdemeanor or a felony. Critics argue that prosecutors pay more attention to their cases when the victim is educated and middle- or upper-class than when the victim is poor and uneducated, and that lenient plea bargains are more available to wealthy defendants (Davis, 2008).

Statutes, additionally, contain large sentencing discretion, often without providing clear guidance to the judge and/or jury. For example, in Illinois, depending on the fact pattern of the crime, battery can be charged as battery, battery of an unborn child, domestic battery, aggravated domestic battery, aggravated battery, aggravated battery with a firearm, aggravated battery with a machine gun or a firearm equipped with any device or attachment designed or used for silencing the report of a firearm, aggravated battery of a child, aggravated battery of an unborn child, aggravated battery of a senior citizen, drug induced infliction of aggravated battery to a child athlete, and heinous battery, all carrying different recommended sentencing lengths.

Heinous battery in Illinois’ Criminal Code of 2012 (720 ILCS 5/12-4.1) is defined as a person who, in committing a battery, knowingly causes severe and permanent disability, great bodily harm or disfigurement by means of a caustic or flammable substance, a poisonous gas, a deadly biological or chemical contaminant or agent, a radioactive substance, or a bomb or explosive compound. This is the only guidance in the statute, with the trier of fact directed to determine a sentence of no less than 6 years and no more than 45 years. Who deserves 6 and who deserves 45 years? There are currently no standardized guidelines to assist the trier of fact in determining how one battery compares to other battery offenses to inform such a decision.

The U.S. Supreme Court has also concluded that states must define the crimes for which harsher penalties are appropriate, and that these definitions should obviate standardless sentencing discretion (Godfrey v Georgia, 1980). The statutory guidelines that assist the trier of fact in making determination of aggravators vary significantly from state to state. Aggravators sometimes include aspects of the severity of the intent (i.e., “The manner in which the crime was carried out indicates planning, sophistication, or professionalism” (California, C.R.C. §4.421(a)(8)), actions (i.e., “The harm, injury, loss, or damage suffered by the victim of an offense was greater than the elements necessary to prove the commission of the offense”; Indiana, I.C. §35-38-1-7.1(a)(1)(2)), or the victim targeted (i.e., “the defendant knowingly directed the conduct constituting the offense at a victim because of that person’s race, sex, color, creed, physical or mental disability, ancestry, or national origin; Alaska, A.S. §12.55.155(c)(22)).

However, a variety of other factors may also be included as aggravators, such as when a crime is heinous, atrocious, cruel, inhuman, vile, depraved, or other synonymous terms. For example, where “[a] person who, with the intent to cause cruel or extreme physical or mental pain and suffering, inflicts great bodily injury or severe mental pain or suffering upon another person within his or her custody or physical control commits torture and is guilty of a felony punishable by imprisonment for life or any term of years” (Michigan; M.P.C. §750.85 (1)(2)(a), and “[a] person is guilty of elevated aggravated assault if that person. . .engages in conduct that manifests a depraved indifference to the value of human life and that in fact causes serious bodily injury to another person with the use of a dangerous weapon” (Maine; M.R.S. §208-B (1)).

The U.S. Supreme Court has acknowledged the jury’s burden to weigh aggravating factors despite a lack of expertise, and that this quandary could be alleviated if the jury received guidance regarding factors of the crime and defendant that the State, representing organized society, deems relevant to sentencing (Gregg v. Georgia, 1976).

The aggravator of heinous, atrocious, and cruel (also known as HAC) primarily appears in murder cases. However, the notion that the distinction of a depraved assault or a depraved robbery is any less important is both short-sighted and unfair to the experience of both the victim and the perpetrator. If the depravity of a crime, or its lack of depravity, does not reflect upon its severity, then what is aggravating, anyway?

The Depravity Standard

A methodology to better guide decision-making about depravity in murder incorporates elements of intent, actions, victimology, and an offender’s attitude about the offense and has been validated for application to the American justice system. The Depravity Standard was developed (Welner, O’Malley, Gonidakis, Saxena, et al., 2018) to emphasize the use of evidence over impressionistic judgment, a reliable methodology over arbitrariness, incorporating data mining from numerous source materials of large numbers of adjudicated guilty murder cases, and focusing on the components of a crime rather than the offender’s background and other sources of bias.

The Depravity Standard’s research methodology also integrated large-scale public input with data mining from adjudicated case decisions. The research has

demonstrated content validity, construct validity, and interrater reliability for the assessment of the depravity of a given murder case. The degree of depravity of a murder case can be demonstrated, as can the depravity of one offender relative to another within the same case. The Depravity Standard as an evidence-based societal standard of depravity, can be used to guide justice officials when determining the most appropriate charge based on the severity of the crime compared to other similarly charged offenses. It can also be used to assist judges and juries make sentencing decisions when wide ranges are presented in statutes, and for assisting parole boards to determine eligibility for early release from prison.

Having recognized early-release decisions that typically only apply to non-murder crimes, additional research was undertaken to develop sex crimes, non-violent, and violent crimes versions of the Depravity Standard. This article presents the development and validation of a version of the Depravity Standard for use in violent (non-murder) felony crimes.

The Depravity Standard—Violent Crimes Evidence Guide was borne from the original Depravity Standard—Murder Evidence Guide (Welner, O’Malley, Gonidakis, & Tellalian, 2018; Welner, O’Malley, Gonidakis, Saxena, et al., 2018, 2018). The 25-item evidence inventory, with its highly detailed definitions and qualifying and disqualifying examples, was developed through several study phases. The first study involved the review of 110 appellate court cases that were upheld for being “heinous, atrocious, or cruel” or other similar terms like “evil” and “depraved.” This established an initial pool of items for consideration in a Depravity Standard, that was next supplemented with input from 91 law and psychology professionals to expand the possible items. Thematic analysis was employed to consolidate the examples provided, and the items fell into four different categories, namely the intent, actions, attitude, and victim choice of the offender.

In order to ensure that the possible items were a reflection of societal standards, an online public survey was next employed (see; Welner, O’Malley, Gonidakis, Saxena, et al., 2018). Over 25,000 participants from the U.S. rated each of the items for being “Not Depraved,” “Somewhat Depraved,” or “Especially Depraved.” By combining somewhat and especially depraved responses, all items ranged between 69.5% and 99.1% in support for being at least somewhat depraved, and as such all were included in the instrument (Welner, O’Malley, Gonidakis, Saxena, et al., 2018). These final 25 items were then applied to over 750 adjudicated guilty murder cases files to refine the qualifying and disqualifying examples (for more information see Welner, O’Malley, Gonidakis, Saxena, et al., 2018). The Murder Evidence Guide item definitions were then reviewed and adapted for non-murder violent crimes, and this development of the Violent Crimes Evidence Guide is outlined below.

Method

This section details the methodology employed for the item development, application of items to case files, inter-rater reliability, item weighing, and case scoring using the Depravity Standard.

Item Development

The detailed Depravity Standard—Murder Evidence Guide descriptions were adapted to apply to non-homicide violent crimes. Each of the 25-items from the Murder Evidence Guide (Welner, O’Malley, Gonidakis, & Tellalian, 2018; Welner, O’Malley, Gonidakis, Saxena, et al., 2018, 2018) was refined for distinction and definition to ensure applicability of the Depravity Standard across a representative sampling of violent crime felony cases. While the item name itself remained the same (i.e., *Item 1: Intent to emotionally traumatize the victim, maximizing terror, or through humiliation, or intent to create an indelible memory of the event* is the same item name in both Evidence Guides), the numerous qualifying examples and disqualifying examples were tailored to non-murder violent crimes specifically through extensive review of closed felony violent crime case files. The aim was to create definitions that were clear and easy for lay jurors to understand. Thorough descriptions of how each item might present in the range of violent cases would guide raters to reliably apply these definitions during case data mining. The preciseness of definitions and diversity of examples ensured that evidence of presence or evidence of absence of each item would not likely be overlooked.

The definitions for each item were formulated and refined such that when an item was present in a case, it was uncommon or even rare relative to the overall case sample of violent crimes. The exercise of piloting the Depravity Standard to guide case review aimed to demonstrate that the Depravity Standard delineated a narrowed subset of violent crime as more depraved.

Case files for adjudicated guilty violent crimes were obtained from jurisdictions across the United States. Eight disparate jurisdictions provided a combined 521 closed violent (non-homicide) criminal case files for our review. The process for obtaining case files has been described elsewhere (Welner, O’Malley, Gonidakis, Saxena, et al., 2018).

All crimes were required to have a victim. As such, gun or drug possession charges and others were removed from consideration. For categories that were not immediately clear, the Federal Bureau of Investigation determinations were consulted. As a result, simple robbery cases were included (Federal Bureau of Investigation, 2011), ensuring that our analysis would dovetail with the classification of violent crime as it currently exists in the American justice system. Each case was probed for elements of violence and only included when violent intent or action was present.

Data deriving from multiple and convergent sources of evidence parallels best practices across forensic sciences, and contributes to the certainty of justice (Meloy, 2006; Monahan et al., 2001). The case files relied upon for this study therefore drew from a range of data sources, including but not limited to crime scene photos, 911 calls, videotapes of the event, presentencing reports, criminal history, police reports, and victim, witness, and defendant statements.

Cases that upon closer review were victimless ($n=2$), or in which the defendant was found not guilty or incompetent to stand trial ($n=1$) were removed from analysis. Incomplete files were also removed ($n=10$). Cases were defined as incomplete if

information germane to the Depravity Standard was missing due to a clear lack of documented information about the crime, or if the documents were so heavily redacted that what actually took place remained uncertain.

The remaining batch of 506 cases were randomly assigned to raters. At the time of publication 393 had been rated and were included for analysis. Table 1 demonstrates the total number of cases rated by jurisdiction.

Procedure. Raters were research volunteers who completed a preliminary 2-hour training session before being assigned practice cases. The description of participants and their training has been previously detailed (Welner, O’Malley, Gonidakis, Saxena, et al., 2018). When training supervisors determined that raters had a keen understanding of the 25 Depravity Standard items, cases were assigned randomly from the overall sample of U.S. cases provided.

Raters reviewed each case and applied the definitions and qualifying and disqualifying examples of each of the 25 items. Each case was rated by two independent participants trained in the same protocol. The rating process involved reviewing a case, applying the definitions and qualifying and disqualifying examples, and charting the presence or absence of each of the 25 items within each of the documents of the case, one at a time. For each case document, a rating of present (Yes), absent (No), or insufficient data (ID) was recorded for each of the 25 items, along with the raters’ evidence for the score. For example, the rater would review the crime scene photos, and for each item, note Yes, No, or ID. The ID responses represented either a lack of information regarding depravity in the crime (i.e., a witness statement that only details that they saw a blue vehicle parked next door) or a rater’s uncertainty about the information presented (i.e., the defendant’s statement contradicts a co-defendant’s statement and there is no way to determine which is the true account from the provided materials without speculation).

Then, the rater completed an overall rating for each of the items, detailing the evidence for the rating. See Table 2 for an example of the overall results for Item 20 “Excessive response to trivial irritant” as assessed for one of the cases.

The full template contains a row for each of the 25 items, with two templates (one created by each rater) per case. These templates were reviewed by a Research Assistant who had extensive training in the protocols. This review formed a consensus rating for each of the 25 items per case. If both raters agreed on the rating for an item, that response was the consensus for the case. If the raters disagreed on the presence or absence of an item, that discrepancy was reviewed in depth. If it was determined that the discrepancy was present due to a clear misunderstanding of an item, additional training was provided to the rater. If the discrepancy identified an opportunity to refine the tool’s qualifying and disqualifying examples to add clarity, the question was reviewed by the entire active research team in a scheduled research team meeting. Adjustments to the definitions were made where appropriate and agreed upon by the whole team, and an additional rater would be assigned to rate the case using the updated definitions. A new consensus rating for that item in that case would then be made, and any further discrepancies were again treated as above. Raters were debriefed

Table 1. Breakdown of Cases Rated by Type Supplied and Included per Jurisdiction.

Type of crime	Christian county (MO)	Jackson county (MO)	Jacksonville (FL)	Jefferson parish (LA)	Kaua'i county (HI)	Lake county (IL)	Little Rock (AR)	St. Clair (IL)	Totals
Abuse of a child	1	1			16				2
Abuse of family/domestic abuse		15	3	1		19	1	31	30
Aggravated assault			2		6				64
Aggravated battery						2			53
Aggravated burglary									6
Aggravated domestic battery									12
Aggravated robbery									30
Armed robbery	10	14		1	1			4	6
Assault					35				62
Battery			6	6		10			22
Burglary with assault or battery			1						1
Carjacking				4					4
Child abuse	2	3		1					5
Criminal mischief									1
Cruelty to a juvenile second degree					1				1
Domestic assault	4	9		2	1		6	8	13
Domestic battery			1		1		2		20
Endangering the welfare of a child									4
False imprisonment/felonious restraint	2		1		1		1		5
Home invasion			1		1				6
Kidnapping							5		1
Poison food/water with intent to kill/injure persons								4	17
Robbery	3		2		4			10	10
Terroristic act						12	2		15
Terroristic threatening		1					1		1
Unlawful discharge of firearm									1
Vehicular assault	1								1
Total violent cases rated	22	46	26	33	68	33	112	53	393

Table 2. Example of case rating for Item 20.

Case # 24

Item 20: Yes, indicates the perpetrator assaulted the victim without being provoked because he thought the victim was looking at him the wrong way.

Police records	
Arrest report, arrest warrant, affidavit	ID, documents do not have information that can be used to prove or disprove the item.
Case summary, complaint, rap sheet	ID, documents do not have any details pertaining to the crime and the information provided cannot be used to prove or disprove the item.
Incident report, hate crime report	Yes, indicates the perpetrator assaulted the victim without being provoked.
Supplemental report	Yes, it appears that the suspect reacted to the way the victim looked at him.
Police intake screening form, prosecutor letter	Yes, indicates the perpetrator assaulted the victim without being provoked. Witnesses said that they do not know what triggered the suspect.
Photos	ID, photos cannot be used to prove or disprove the item.
Medical records	
Diagnostic report	Yes, the suspect said he attacked the victim because he was looking at him.
Victim status form	ID, does not contain relevant information.
Medical scan	ID, does not contain relevant information.
Statements and interviews	
Officer declaration	Yes, indicates the perpetrator attacked the victim when he thought the victim was looking at him.
Defendant statement	Yes, the suspect said he attacked the victim because he was looking at him.

regularly to ensure adequate supervision, to promote discussion about the cases, ensure quality, and that the nuances of instruction were being followed in a uniform fashion. These procedures have been previously described in detail (Welner, O’Malley, Gonidakis, Saxena, et al., 2018).

The protocol of lay rating of cases by the raters, a Research Assistant facilitating consensus, and collaborative discussion about each item, has refined the Depravity Standard and its Violent Crimes items into an inventory applicable to everyday case data. Qualifying and disqualifying examples for each item in the context of violent crimes have been finalized through this process in order to guide the determination of rating for an item’s presence or absence.

Applying the Definitions to Case Files

In addition to further refining the definitions, applying the Depravity Standard items to closed felony case files allowed for an assessment of whether the 25

items developed reliably measure depravity in actual violent crime cases. It was hypothesized that Depravity Standard items would only appear in a sub-group of the overall group of cases, and that multiple items would appear in still smaller subsets. If items appeared with high frequency, the Depravity Standard may not be narrowing enough to the worst of cases. If items never appeared, the definitions could be too restrictive, unless the rare presence of that item demonstrates it to be uncommon, but not impossible.

Items that seldom presented were reviewed to account for the possibility that definitions were not unduly restrictive and that items were not appearing because they were rarely seen, but still realistic possibilities. *Influencing criminality in others to avoid prosecution or penalty* does occur in certain violent attacks, such as in gang-related crimes where someone is recruited as the henchman so that a leader can maintain the veneer of not being involved. Our findings reflect upon this item as rare. This effort was therefore designed to additionally test the content validity of the Depravity Standard measure.

Inter-rater Reliability

An important aspect of validating the Depravity Standard for application to violent crimes cases was ensuring inter-rater reliability. From the 393 cases reviewed, a sub-sample of 250 cases were randomly selected for inclusion using a random sample generator (Haahr, 2006). Each of the 250 cases were rated twice by two independent groups of raters for the presence of the Depravity Standard items. Each case was assigned an overall rating of present (Yes), absent (No) or insufficient data (ID) for each of the 25 items.

Data was analyzed using IBM SPSS Statistics for Macintosh Version 23.0. The data was screened for coding errors, and responses for the presence of each item were coded as “Yes” or “No/ID.” Ratings of “No” and “ID” have the same functional significance in a criminal justice context as proof of guilt requires presence of evidence beyond a reasonable doubt. A court may only use aggravating factors, excepting prior convictions, to impose a harsher sentence than usual where the jury found those factors to be true beyond a reasonable doubt (i.e., Cunningham v. California, 2007). Therefore, absence of evidence equals evidence of absence for scoring purposes. As such, the ratings of “No” and “ID” were combined.

Upon preliminary analysis of Cohen’s Kappa (Cohen, 1960) for inter-rater reliability, high agreement between raters yielded a low or negative Kappa value. Manual review of the 250 cases demonstrated that ratings of “No/ID” (item is absent or can’t be determined) appeared with significantly more prevalence than ratings of “Yes” (item is present).

The lower occurrence of Yes ratings can be attributed to the extensive development and validation phases of the research, where items were refined to describe specific criteria of intent, actions, attitudes, and victimology that reflect depravity, or the “worst-of-the-worst” violent crimes. The thorough development process provided very specific criteria for each of the items to the participant raters to ensure that, when

scored, the items reflected exceptional qualities in a crime. Raters therefore had the necessary guidance to highlight only a smaller sub-set of offenses for which any of the items were present. Moreover, raters were instructed to rate items as present only when they were certain there was sufficient evidence to support such a judgment. For this reason, the most depraved violent crimes appeared only as a small percentage of overall cases, and this discriminant sensitivity was the hypothesized result.

The AC1 statistic (Gwet, 2008) was applied instead to each of the 25 items across the 250 random cases in IBM SPSS version 22.0 to determine the level of agreement between independent raters. The AC1 statistic is noted by Gwet (2008) to measure “true” inter-rater reliability in that it removes the evaluation of all agreements that occurred by chance. Since Kappa statistics can lead to low or negative values in data sets with agreement yields above or below 50%, the AC1 statistic was more appropriate (Gwet, 2008).

Weighting the Items

During the initial public survey during the Depravity Standard’s development, the researchers established that public respondents affirm the 25 Depravity Standard items as being somewhat or especially representative of depravity (Welner, O’Malley, Gonidakis, Saxena, et al., 2018). The study presented each of the items with examples of how they may present in a range of felonies. Examples of how each item might manifest in violent (non-murder) cases were presented.

An additional study was undertaken to delineate public attitudes about each of the items as they may manifest only violent crimes, murder, sex crimes, and non-violent felonies. We undertook this additional study specifically to refine the Depravity Standard as a societal standard and measure the degree of depravity the public attributes to each item relative to other items. This study was used to inform a valid scoring mechanism for how each Depravity Standard item should be weighed when present.

Weighting specific qualities of depravity assists inexperienced and experienced triers of fact alike about the severity of a given offense. Furthermore, in cases of multiple or collaborating perpetrators and conspirators, weighing items of the Depravity Standard informs appraisal of the relative culpability of co-defendants, especially because it explores intent and therefore accounts for those who are a crime’s unseen hand. It also allows for assessment of inter-item correlations. By focusing on the quality of depravity as opposed to quantity of depraved items alone, the features of criminal behavior that society deems most and less severe can impact sentencing to a degree more in accord with societal preferences.

IRB was obtained for this study from The Forensic Panel’s Institutional Review Board.

Participants. Participants were 14,206 respondents from around the world who took part in a voluntary online survey. The survey was accessible to anyone who elected to take part, however, recruitment efforts targeted United States residents aged 18 years

or older. Participants were randomly assigned to one of four categories; Murder/Attempted Murder, Violent Crimes, Non-Violent Crimes, or Sex Crimes. 1,590 participants were retained for analysis of their responses to items occurring in violent (non-homicide) crimes.

Of the participants retained for analysis, 40.3% percent were male and 59.7% were female. The most represented states were Ohio (12.5%), California (8.7%), Texas (7.8%), New York (5.3%), Florida (5.1%), and Virginia (4.4%). The majority of participants were White (82.1%), and most participants reported living in suburban areas (53.6%).

Participation in the Depravity Standard research was voluntary. Participants self-selected to complete the survey by visiting the website www.depravitystandard.org or by searching for the research online. To raise awareness of the research and encourage participants to visit the website and complete the survey, several recruitment techniques were utilized. During media appearances on forensic and investigative topics, including the research, the principal investigator encouraged viewers and readers to participate in the online survey. Additionally, several prior articles have been written about the project (Welner, 2003, 2009, 2013) as well as media articles about the research. Participants, upon completion of the online survey, were also able to enter the email addresses of friends and family who may be interested in participating. A Twitter account (@whatisdepraved) and Facebook account (www.facebook.com/DepravityStandard/) highlighted crimes with thought-provoking fact patterns to encourage readers to contribute their own input on what makes a crime depraved by participating in the surveys. Participants from an earlier phase of the research (Welner, O'Malley, Gonidakis, Saxena, et al., 2018) were emailed and invited to complete the latest survey online. A number of interested parties posted a link to the research on their own social media accounts. Members of the research team placed posters about the research on their college research recruitment boards with permission from their schools. A number of teachers and professors also called attention to the research in their classrooms, and others learned of the research through depictions of the research in fictional television shows (Goodman, 2016).

No compensation was offered for survey completion. Participants read the statement, "By entering the survey I agree that I have read and understood the informed consent information," and were provided with two hyperlinks to view the informed consent information. Participants also received a copy of the informed consent form to their email address when they started the survey. For minors who actively looked up the website and opted to complete the survey, the researchers felt it unnecessary to obtain parental consent since this would have been impractical. Additionally, the content of the survey was concise and far less detailed and graphic than other readily viewable internet content such as news media, entertainment, and gaming sites (Yoon & Somers, 2003).

Procedure. Participants signed in using an email address on the website landing page. This form of entry enabled the researchers to forward informed consent information to them (along with contact information for several helplines, should any

feelings of discomfort emerge). The email address submitted by participants also reduced the likelihood that participants would take the survey twice, with the site rejecting attempts to enter the survey with an email address already used to participate.

Participants then proceeded to an initial starting page that generated a random number to assign them to one of four crime categories: 1 = *Assault/Violent Crimes*; 2 = *Murder/Attempted Murder*, 3 = *Non-Violent Crimes*, and 4 = *Sex Crimes*. The participants were unaware as to which category each number referred. Each was asked to choose their assigned number from a drop-down box and read the participant instructions, including information about consent. Informed consent was established by proceeding and subsequently completing and submitting the survey.

This paper reports on participants who completed the *Violent Crimes* category. There were 25 survey questions, each representing one of the 25 items of the *Depravity Standard*. Participants were instructed to read each item, use a drop-down menu to view longer descriptions and examples of each item, and to then rate the item on a scale of 1 = *Depraved to the least degree* to 100 = *Depraved to the most extreme degree* by dragging a bar until the desired number was reached on the scale of 1 to 100. The order of the 25 items was randomly generated for each participant to control for order effects.

Instructions encouraged participants to use the entire range of 1 to 100 as they believed the items should be scored, in order to discourage homogeneous response styles that would make distinctions among items more difficult. Participants could adjust earlier answers as they proceeded through rating each of the 25 items. Once participants were satisfied with their answers, they clicked a button to submit their results and proceeded to a demographics questionnaire.

Participants were queried about numerous different demographics. Researchers collected this demographic data in order to compare results across sub-groups, to be presented in a subsequent manuscript, and to explore how representative the sample was in comparison to the general United States population. After answering the demographic questions, participants reviewed and submitted their results, and could enter email addresses for friends and family that may wish to take the survey.

Scoring Cases with the Depravity Standard

In order to provide a practical everyday application of the validated *Depravity Standard* to cases in the justice system, we devised a percentile scoring system. This method ensured the variable weight of different items would directly reflect the input of the general public, as well as the frequency with which items appeared in the violent crimes case sample.

The 393 violent crimes cases from jurisdictions across the U.S were merged with the overall percentages scored for each of the 25 items by U.S. participants 18 years and older, to determine overall scores for each of the cases. The public survey data weighted by gender and ethnicity were the final sample used. By merging survey data with case analysis, a percentile scoring system was devised.

Table 3. 25 Depravity Standard Items for the Violent Crimes Evidence Guide.

1	Intent to emotionally traumatize the victim, maximizing terror, through humiliation, or intent to create an indelible emotional memory of the event
2	Intent to maximize damage or destruction, by numbers or amount if more than one person is victimized, or by degree if only one person is victimized
3	Intent to cause permanent physical disfigurement
4	Intent to carry out a crime for excitement of the criminal act
5	Targeting victims who are not merely vulnerable, but helpless
6	Exploiting a necessarily trusting relationship to the victim
7	Influencing depravity in others in order to destroy more
8	Crime reflects intent of progressively increasing depravity
9	Carrying out a crime in order to terrorize others
10	Carrying out crime in order to gain social acceptance or attention, or to show off
11	Influencing criminality in others to avoid prosecution or penalty
12	Disregarding the known consequences to the victim
13	Intentionally targeting victims based upon prejudice
14	Prolonging the duration of a victim's physical suffering
15	Unrelenting physical and emotional victimization; amount of victimization
16	Exceptional degree of physical harm; amount of damage
17	Unusual and extreme quality of suffering of the victim, including terror and helplessness
18	Indulgence of actions, inconsistent with the social context
19	Carrying out crime in unnecessarily close proximity to the victim
20	Excessive response to trivial irritant; actions clearly disproportionate to the perceived provocation
21	Pleasure in response to the actions and their impact
22	Falsely implicating others, knowingly exposing them to wrongful penalty and the stress of prosecution
23	Projecting responsibility onto the victim; feeling entitlement to carry out the action
24	Disrespect for the victim after the fact
25	Indifference to the actions and their impact

Results

Item Development

The final 25 items for the Violent Crimes Depravity Standard Evidence Guide are presented in Table 3.

Each of the 25 items were elaborated to include extensive definitions, with several qualifying and disqualifying examples per item, by applying them to actual closed felony violent crimes case files. For example, criteria for Item 1, *Intent to emotionally traumatize the victim* are met when efforts are made to humiliate the victim in addition to the actual attack, where such efforts are superfluous to the act being carried out. Forcing the victim to strip naked, or urinating on the victim, would be clearly emotionally traumatic and would qualify Item 1 as being present.

Table 4. Number of Items Present in Cases.

Number of items	Total cases with the number of items present
0	129 (32.82%)
1	119 (30.28%)
2	52 (13.23%)
3	45 (11.45%)
4	19 (4.83%)
5	17 (4.33%)
6	4 (1.02%)
7	4 (1.02%)
8	2 (0.51%)
9	1 (0.25%)
13	1 (0.25%)

On the other hand, frightening a victim into compliance or silence does not meet criteria for Item 1. Threatening a victim is common even when the intent is not to traumatize, but rather to carry out the crime and/or to not be apprehended and the victim's compliance is deemed essential. Intimidation to prevent disclosure or to engender silence is not the same as intent to emotionally traumatize. Therefore, in such an instance this item would be rated as being not present.

A qualifying example for *intent to cause permanent physical disfigurement* would be when the perpetrator's objective was to scar a victim indefinitely or permanently, even if the goal was not accomplished. A deep knife wound to the face, amputations, genital mutilation, acid-attacks, and other such assaults aiming to render ugliness or disability are examples, and would qualify this item for being present. The item would be present if, for example, an acid attack aimed at the face, even if by happenstance the perpetrator's aim was poor or the impact lessened because the victim happened to dodge the corrosive with lightning speed.

This item, however, would be disqualified if permanent disfigurement happens to occur as a direct or indirect result of the attack, but does not specifically reflect the goal and intent of the attacker. This precision of item description minimizes arbitrariness and promotes consistent application in casework, however common or obscure the potential fact pattern.

Applying the Definitions to Case Files

Review of the 393 cases that met criteria for inclusion demonstrated that the highest number of cases had no elements of depravity present (32.82%). Cases had between 0 and 9 items present, with one exceptional case containing 13 items. As hypothesized, there is a decline in total cases as the number of items present increased. For example, there were 45 cases (11.45%) with three items present, and 19 cases (4.83%) with four items present. Cases with six items or more were rare. Table 4 lists the number of items present in the cases, and to what frequency.

The percentages of frequency with which each item occurred were assessed in the overall sample across all jurisdictions that provided a random sample of violent crimes cases. All items of the Depravity Standard were represented rarely to infrequently. Item frequencies ranged from .25% (Item 3, *Intent to cause permanent physical disfigurement*) to 31.55% (Item 20, *Excessive response to trivial irritant; actions clearly disproportionate to the perceived provocation*) in incidence. Only three items occurred in more than ten percent of cases, with the most frequent being Item 20. Item 25, *Indifference to the actions and their impact*, was by far the most frequent item relating to attitude of the offender. The most frequently present intent item was Item 2, *Intent to maximize damage or destruction*, present in nearly 9% of cases. The results further demonstrate the utility of the Depravity Standard in creating a narrowed class of the worst of violent crime offenders.

Interrater Reliability

The AC1 statistic was calculated for Items 1 through 25 of the Depravity Standard for the 250 random cases, with Cohen's Kappa reported for comparison (see Table 5).

The AC1 statistic may be interpreted using the cut-off ranges proposed by Landis and Koch (1977), Altman (1990), or Fleiss et al. (2003). Depending on the cut-off scale used, 23 Depravity Standard items demonstrated "almost perfect," "very good," or "excellent" agreement ($0.81 < \kappa < 1.00$), one item had "substantial" or "good," agreement ($0.61 < \kappa < 0.80$), and one item had "moderate" agreement ($.41 < \kappa < .60$). All items were therefore retained for the Violent Crimes Evidence Guide.

Weighting the Items

Participants rated the level of depravity for each item surveyed on a scale of 1 to 100 (1 is the least depraved and 100 is the most depraved an item could be) and completed a series of demographic questions. Data was screened to exclude respondents who only filled out the demographic information but did not complete the questionnaire items in the survey. Outliers were also removed. Participants who completed fewer than 19 (75%) of the 25 questionnaire items were also excluded from analysis, since the weighting of items was relative to how a participant perceived each item on a scale of 1 to 100. If fewer than 19 items were accounted for, or if there was little or no variance in scores (i.e., answering "100" or "0" for every item), this did not represent a sufficient comparative assessment of depravity for the purpose of weighting. Any participants with clearly fake demographics were also removed.

Participants from outside the United States, or who did not answer the question for their country or Zip Code, were also removed from analysis. Data was also removed from participants under the age of 18 and therefore unable to serve on an American jury, and from participants who did not consider themselves fluent in English; the final sample comprised 1,590 participants.

Based on judicial demands for a jury of peers (Duren v. Missouri, 1979; reaffirmed in Berghuis v. Smith, 2010), Chi-square goodness-of-fit analyses were employed to

Table 5. Inter-Rater Reliability Results for Violent Crimes.

	ACI	COHEN'S
Item 1	0.913	0.384
Item 2	0.797	0.336
Item 3	0.996	0
Item 4	0.992	0.496
Item 5	0.979	0.437
Item 6	0.838	0.282
Item 7	0.967	0.184
Item 8	0.835	0.071
Item 9	0.996	0.798
Item 10	0.984	0.325
Item 11	0.996	0
Item 12	0.861	0.42
Item 13	1	1
Item 14	0.929	0.351
Item 15	0.785	0.527
Item 16	0.93	0.562
Item 17	0.845	0.263
Item 18	0.881	0.181
Item 19	0.93	0.24
Item 20	0.482	0.429
Item 21	0.883	0.08
Item 22	0.971	0.447
Item 23	0.836	0.411
Item 24	0.93	0.3
Item 25	0.613	0.262

compare our sample against national data for ethnicity (Kaiser Family Foundation, 2019a) and gender (Kaiser Family Foundation, 2019b). Review of the chi-square goodness-of-fit statistical results revealed that males were underrepresented in our sample, $\chi^2 (1, N=1,590)=48.626, p<.001$, as were most ethnic minority groups, $\chi^2 (6, N=1,590)=387.314, p<.001$. Data was therefore weighted 1.2 for males and 0.9 for women, with ethnicity weights ranging and from 0.7 for White participants to 3.3 for Hispanic participants, combined into a new gender \times ethnicity variable.

To analyze the overall ranking of the Depravity Standard items, and specifically what their weighted score was compared to other items in the measure, the 1 to 100 responses were averaged across participants to generate a percentage. Analyses were run on the weighted and unweighted data to determine the overall score for each of the 25 items. The average level of depravity was determined for each item in the form of a percent. Summary *t*-tests were run to compare the weighted and unweighted totals with no item values being statistically significantly different (equal variances assumed, $p<.001$) (See Table 6).

Table 6. Depravity Rating by Item.

Item	Description	N	Unweighted percentage	Weighted percentage	Summary t-test significance
1	Intent to emotionally traumatize the victim, maximizing terror, through humiliation, or intent to create an indelible emotional memory of the event	1,590	77.92	77.49	0.58
2	Intent to maximize damage or destruction, by numbers or amount if more than one person is victimized, or by degree if only one person is victimized	1,590	74.04	74.33	0.72
3	Intent to cause permanent physical disfigurement	1,590	82.33	82.50	0.80
4	Intent to carry out a crime for excitement of the criminal act	1,590	64.96	66.12	0.22
5	Targeting victims who are not merely vulnerable, but helpless	1,590	82.19	81.78	0.5
6	Exploiting a necessarily trusting relationship to the victim	1,589	71.87	71.61	0.75
7	Influencing depravity in others in order to destroy more	1,590	67.34	67.81	0.58
8	Crime reflects intent of progressively increasing depravity	1,588	71.17	71.39	0.78
9	Carrying out a crime in order to terrorize others	1,589	71.46	72.04	0.48
10	Carrying out crime in order to gain social acceptance or attention, or to show off	1,589	53.89	54.98	0.25
11	Influencing criminality in others to avoid prosecution or penalty	1,590	59.22	59.86	0.49
12	Disregarding the known consequences to the victim	1,590	68.48	68.57	0.92
13	Intentionally targeting victims based upon prejudice	1,590	66.34	67.85	0.10
14	Prolonging the duration of a victim's physical suffering	1,590	86.42	86.11	0.58
15	Unrelenting physical and emotional victimization; amount of victimization	1,589	78.14	77.60	0.46
16	Exceptional degree of physical harm; amount of damage	1,590	78.05	77.96	0.90
17	Unusual and extreme quality of suffering of the victim, including terror and helplessness	1,590	85.17	84.21	0.13
18	Indulgence of actions, inconsistent with the social context	1,588	65.92	65.58	0.70
19	Carrying out crime in unnecessarily close proximity to the victim	1,590	65.79	66.16	0.68
20	Excessive response to trivial irritant; actions clearly disproportionate to the perceived provocation	1,584	50.77	50.96	0.83
21	Pleasure in response to the actions and their impact	1,584	75.50	75.51	0.99
22	Falsely implicating others, knowingly exposing them to wrongful penalty and the stress of prosecution	1,582	62.00	62.61	0.51
23	Projecting responsibility onto the victim; feeling entitlement to carry out the action	1,582	66.08	66.02	0.95
24	Disrespect for the victim after the fact	1,582	68.66	68.87	0.82
25	Indifference to the actions and their impact	1,582	67.23	67.60	0.70

^aIndicates significant result at $p < .001$.

Table 7. Percentiles of Depravity Scores.

25th	30th	35th	40th	45th	50th	55th	60th	65th	70th	75th	80th	85th	90th	95th	99th +
0.00	0.00	50.96	50.96	66.02	67.60	75.16	77.60	118.56	128.92	150.57	195.16	212.67	271.48	355.50	567.11+

Inter-item correlations were additionally performed to allow for the assessment of reliability through examination of internal consistency of the items. Cronbach's alpha was determined for the weighted data ($\alpha = .95$), falling above the recommended minimum standard of .90 for applied research (Nunnally, 1978). The average inter-item correlation was 0.434. Ideally, the average inter-item correlation for a set of items should be between .15 and .50 (Briggs & Cheek, 1986). For a valid measure of a narrow construct, like depravity in crime, a much higher mean intercorrelation of .40 to .50 is recommended (Clark & Watson, 1995), thereby suggesting that the Depravity Standard items for violent crimes are reasonably homogenous while containing sufficiently unique variance so as not to be isomorphic, or a different expression of the same item (Piedmont, 2014). The weighted data was kept for further analysis.

Scoring Cases with the Depravity Standard

Percentages were first determined based on the survey responses. For example, Item 2 has an average score of 74.33 (out of a possible 100). If this was the only item present in a case, this would be the overall Depravity Score for that case. If a case had two items present, for example Item 1 (74.49) and Item 16 (77.96), that case would have an overall Depravity Score that is the sum of those two items, or 155.45. Based on these Depravity Scores, we are able to ascertain overall percentiles, to mark individual Depravity Scores for a case against other violent crimes. The percentiles are demonstrated in Table 7.

Therefore, if a hypothetical case had three of the more depraved items present such as Items 3 (82.50), 14 (86.11), and 17 (84.21), the Depravity Score for that case would be 252.82. Based on the percentiles, this case would be more depraved than over 85% of violent crimes in the U.S. If, however, a case was present with items that the survey participants scored for being less depraved, four or more items would need to be present to reflect an 85th percentile. For example, if Items 10 (54.98), 11 (59.86), 20 (50.96), and 22 (62.61), were present in a case, the Depravity Score would be 228.41. These four items in combination would still total less than the three items present in the first case. The scoring system directly reflects the impact of societal input, embedded in the experience of violent crimes cases. In converging findings from case data with public survey data, the Depravity Standard scoring assists as a guide to experienced and inexperienced triers of fact.

This merging of survey data with case analysis to devise percentiles also allows for a quick and easy comparison between violent crimes for severity. Our data demonstrates that cases of six items or more are so seldom found as to demonstrate a truly unusually rare degree of depravity beyond that point. With a 99+ percentile

demonstrated at almost 570, scores beyond this measure are so extreme to make comparison beside the point.

Nevertheless, scores in lower ranges are influenced by the specific items present. The specific nature of intent, action, attitude, and victimology in these crimes may vividly illustrate significant differences in depravity between crimes, and within the same crime among co-defendants with different roles, different actions, and different reactions.

Discussion

The Depravity Standard is a guide to aid the trier of fact in assessing evidence pertinent to determining charges, sentencing decisions, and early release decisions. Apart from an expansive effort to incorporate public attitudes, and to account for demographic differences, the research has used a large number of highly sourced and adjudicated guilty violent crimes cases, in order to establish content validity. The researchers were fortunate enough to have been given access to these hundreds of pertinent cases from diverse jurisdictions.

The Depravity Standard is an evidence-based inventory of specific components of criminal depravity. Applied to the data of a case, the Depravity Standard enables justice participants to identify evidence of whether a crime, or even a particular perpetrator among co-defendants in a crime, demonstrated depravity and to what degree. It promotes investigative diligence beyond determinations of guilt.

Prosecuting authorities and grand juries need no longer assign aggravated charges without specific guidance as to whether a crime demonstrates evidence for depravity. These decisions can and should be informed by sufficient diligence at the investigative level even before choices about charges are made. The Depravity Standard assists investigative decision-making to enhance the quality and quantity of evidence informing prosecutor decision-making, triers of fact, and eventually, those aiming to make fair and equitable early-release decisions.

This advance reduces the likelihood that defendants are overcharged because of bias or other forces that impact justice that have little to do with the case facts. Defense and prosecution are both able to ascertain and establish the presence or absence of evidence for 25 specific examples of intent, actions, attitudes, and victimology in sentencing trials. Judges, juries, parole boards, and governors with access to such guidance are far better informed to assess the depravity of violent crime, and to better compensate for their lack of experience.

We acknowledge certain shortcomings of the research to date. Minority participants in the online survey were underrepresented. Weighing our responses to compensate for this underrepresentation did not demonstrate a significant difference between the weighted and unweighted data. However, future research post-implementation of the instrument in criminal justice settings will evaluate whether this underrepresentation impacts on its application by gender and ethnically diverse juries.

Much higher participant numbers may change the understandings within sub-groups, and as such the online surveys of the general public remain open. We will

continue to seek a data pool that is fully informative of all demographics, including continued efforts to recruit more minority participants. We will regularly review participant responses and any changes in societal attitudes relative to the items of the Depravity Standard will be reflected in future versions.

The studies presented in this article reflect a review of violent crimes. There will be types of violent crimes that have not been accounted for in this analysis due to not being present in the cases provided for review by the different jurisdictions. Rarer forms of violent crimes may indicate differences within their sub-niches to be incorporated in future versions of the violent Depravity Standard instrument; we now have the methodology to assess and drill down further on these subcategories. The Depravity Standard can be used by the scientific and legal community as a magnification tool to scrutinize subtypes of crimes, enhancing fairness in justice.

Applications

The qualitative and quantitative guidance about violent crimes from the Depravity Standard informs determinations of depravity at all stages of the justice system, and has particular value in guiding early release decisions. Parole and corrections officials confront difficult choices in an era of prison overcrowding and mandates to release large numbers of inmates. Elected officials considering pardons or commutations also deliberate without guidance. Pressures to further lower census due to COVID risk has further accelerated early-release and interest in alternatives to incarceration (Franco-Paredes et al., 2021).

The Depravity Standard informs early release decisions with guidance about the degree of depravity in one violent crime relative to others, and aids the decision of whether and when leniency is fair. Institutional conduct, predictions of future conduct, sentence length, program participation, and past criminal record are some of the strongest factors associated with parole release in states such as Oklahoma, Connecticut, Texas, Pennsylvania, and Nebraska (Caplan, 2007). Carroll and Burke (1990) found that Wisconsin parole board members place weight on punishment issues such as crime seriousness and prior record.

Proctor (1999) found that inmates were four times less likely to be granted parole if there was public opposition at the parole hearing. Evidence-based release decisions limit bias, promote fairness and enhance public trust in justice. Those officials who utilize the Depravity Standard seek to fairly assess pardon decisions rather than acting out of cronyism or other qualities that undermine public confidence in justice. The Depravity Standard brings objectivity and integrity to early-release determinations.

Many violent criminals have identical risk profiles for re-offense in the community. Which person convicted of battery is more appropriate for early release? Which person convicted of robbery? The Depravity Standard further focuses the pool of those eligible for early release after risk assessment measures have been used. If one person's crime is in the 40th percentile of violent crimes for depravity, and they are compared to another whose crime was in the 90th percentile, it becomes clearer as to who is more deserving of leniency. The injection of evidence collected with the Depravity

Standard regarding what a person intended to do as well as what one did will lead responsible parties to more just consideration of such cases.

The Depravity Standard does not replace judicial decision making and serves only as a guide to those deliberating charges, sentences, and early release decisions. There is otherwise no validated inventory or methodology to inform the elements of depravity in violent crime. By focusing inquiry on intents, actions, victimology, and attitudes of a violent crime, the Depravity Standard minimizes reliance on subjective judgments and narrows a class of the worst of violent offenders.

Assessment of depraved intent, action, victimology, and attitude factors can be separated from potential biases about who perpetrated the crime or why it was carried out. The Depravity Standard does not negate the importance of character evidence. However, it enables the before, during, and aftermath of crime to be carefully scrutinized with the focus needed to distinguish the severity of a crime by its unique features, while eliminating bias based on gender, ethnicity, socio-economic status or sexual orientation.

Professionals from multiple disciplines can be trained to utilize the Depravity Standard as a valuable guide, without requiring expert witnesses to interpret fundamental and straightforward evidence. The scoring process is straightforward and integrates public opinion with case file data to establish weights for each of the 25 items. The numerical weights of each of the items present are added together, enabling a case to be compared to other violent crimes for severity of depraved elements. This process for scoring ensures the distinction of the worst of crimes by a fair and informed decision-maker. The Depravity Standard is also separately being validated for application to sex crimes like rape and sexual assaults, and non-violent crimes such as property offenses, arson, fraud, and embezzlement. The training package for the four crime categories and the release of the Depravity Standard Evidence Guide's will enable application to the justice system as has been reliably demonstrated here.

Conclusion

The Violent Crimes version of the Depravity Standard has been developed to meet a long-standing need within the American criminal justice and corrections system to identify the worst-of-the-worst non-murder violent crimes for informing justice decision-making. Juries deliberating depraved elements of a crime will be able to consider the presence or absence of relevant evidence informing specific items of the Depravity Standard, as opposed to relying on gut instinct or court theatrics.

States that are now aggressively lowering their incarcerated populations by early release have not yet established any mechanism for providing sentencing leniency to those whose crimes are more deserving. Parole and corrections officials will benefit from the use of evidence-based guidelines that minimize bias and incidental influence.

The reliability and validation studies published here demonstrate how the Depravity Standard acts as a non-denominational guide that promotes fairness and a reliance on facts. The methodology can easily be applied to investigate, assess, score, and inform

whether a violent crime is exceptional, and to what degree. In that regard, the Depravity Standard promotes the fairness of our justice system and in turn, the public's confidence in it.

Author Note

Michael Welner is now affiliated to Icahn School of Medicine at Mount Sinai, New York, NY, USA.

Acknowledgments

The authors would like to acknowledge the many research assistants and interns who supported the development of this project. We are also grateful to the jurisdictions across the country that provided the closed felony cases for analysis.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Kate Y. O'Malley  <https://orcid.org/0000-0003-2136-1713>

References

Altman, D. G. (1990). *Practical statistics for medical research*. Chapman & Hall.

Berghuis v. Smith. (2010). 559 U.S. 314.

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality and Individual Differences*, 54, 106–148. <https://doi.org/10.1111/j.1467-6494.1986.tb00391.x>

Caplan, J. M. (2007). What factors affect parole: A review of empirical research. *Federal Probation*, 71(1), 16–19.

Carroll, J. S., & Burke, P. A. (1990). Evaluation and prediction in expert parole decisions. *Criminal Justice and Behavior*, 17(3), 315–332. <https://doi.org/10.1177/0093854890017003006>

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309–319. <https://doi.org/10.1037/1040-3590.7.3.309>

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>

Crick, N. R., & Dodge, K. A. (1996). Social information-processing mechanisms in reactive and proactive aggression. *Child Development*, 67(3), 993–1002. <https://doi.org/10.2307/1131875>

Cunningham v. California. (2007). 549 U.S. 270.

Davis, A. J. (2008). *Arbitrary justice: The power of the American prosecutor*. Oxford University Press.

Duren v. Missouri. (1979). 439 U.S. 357.

Federal Bureau of Investigation. (2011). *Crime in the United States*. <https://ucr.fbi.gov/crime-in-the-u-s/2011/crime-in-the-u-s-2011/violent-crime/violent-crime>

Fleiss, J., Levin, B., & Paik, M. (2003). *Statistical methods for rates & proportions* (3rd ed.). Wiley & Sons.

Franco-Paredes, C., Ghandnoosh, N., Latif, H., Krsak, M., Henao-Martinez, A. F., Robins, M., Vargas Barahona, L., & Poeschla, E. M. (2021). Decarceration and community re-entry in the COVID-19 era. *The Lancet Infectious Diseases*, 21(1), E11–E16. [https://doi.org/10.1016/S1473-3099\(20\)30730-1](https://doi.org/10.1016/S1473-3099(20)30730-1)

Godfrey v. Georgia. (1980). 446 U.S. 420.

Goodman, B. (2016). Turn it upside down [television series episode]. In doherty, R. (created by), Elementary. Columbia broadcasting system.

Gregg v. Georgia. (1976). 428 U.S. 153.

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, 61(1), 29–48. <https://doi.org/10.1348/000711006x126600>

Haahr, M. (2006). *Random.org: True random number service*. <http://www.random.org>

Kaiser Family Foundation. (2019a). *Population distribution by race/ethnicity 2019*. <http://kff.org/other/state-indicator/distribution-by-raceethnicity/>

Kaiser Family Foundation. (2019b). *Population distribution by gender 2019*. <http://kff.org/other/state-indicator/distribution-by-gender/>

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>

Meloy, J. R. (2006). Empirical basis and forensic application of affective and predatory violence. *Australian and New Zealand Journal of Psychiatry*, 40, 539–547. <https://doi.org/10.1080/j.1440-1614.2006.01837.x>

Monahan, J., Steadman, H., Silver, E., Appelbaum, P., Robbins, P., Mulvey, E., Roth, L., Grisso, T., & Banks, S. (2001). *Rethinking risk assessment: The MacArthur study of mental disorder and violence*. Oxford University Press.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Piedmont, R. L. (2014). Inter-item correlations. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-being Research* (pp. 3303–3304). Springer.

Proctor, J. L. (1999). The “new parole”: An analysis of parole board decision making as a function of eligibility. *Journal of Crime and Justice*, 22(2), 193–217. <https://doi.org/10.1080/0735648x.1999.9721100>

Weinshenker, N. J., & Siegel, A. (2002). Bimodal classification of aggression: Affective defense and predatory attack. *Aggression and Violent Behavior*, 7, 237–250. [https://doi.org/10.1016/s1359-1789\(01\)00042-8](https://doi.org/10.1016/s1359-1789(01)00042-8)

Welner, M. (2003). Response to Simon: Legal relevance demands that evil be defined and standardized. *The Journal of the American Academy of Psychiatry and the Law*, 31, 417–421.

Welner, M. (2009). The justice and therapeutic promise of science-based research on criminal evil. *The Journal of the American Academy of Psychiatry and the Law*, 37(4), 442–449.

Welner, M. (2013). Classifying crimes by severity: from aggravators to depravity. In In. J. E. Douglas, A. W. Burgess, A. G. Burgess, & R. K. Ressler (Eds.), *Crime Classification Manual* (3rd ed., pp. 55–72). John Wiley & Sons.

Welner, M., O’Malley, K. Y., Gonidakis, J., & Tellalian, R. E. (2018a). The Depravity Standard I: An Introduction. *Criminal Justice Journal*, 55, 1–11. <https://doi.org/10.1016/j.jcrimjus.2017.12.008>

Welner, M., O’Malley, K. Y., Gonidakis, J., Saxena, A., & Burnes, J. (2018b). The Depravity Standard II: Developing a Measure of the Worst of Crimes. *Criminal Justice Journal*, 55, 25–34. <https://doi.org/10.1016/j.jcrimjus.2017.12.009>

Welner, M., O’Malley, K. Y., Gonidakis, J., Saxena, A., & Stewart-Willis, J. (2018c). The Depravity Standard III: Validating an Evidence-Based Guide. *Criminal Justice Journal*, 55, 12–24. <https://doi.org/10.1016/j.jcrimjus.2017.12.010>

Yoon, J. S., & Somers, C. L. (2003). Aggressive content of high school students’ TV viewing. *Psychological Reports*, 93, 949–953. <https://doi.org/10.2466/pr0.2003.93.3.949>