Development and Validation of the Delinquency Reduction Outcome Profile (DROP) in a Sample of Incarcerated Juveniles: A Multiconstruct/Multisituational Scoring Approach

Baptiste Barbot
Yale University

Gerald J. Haeffel
University of Notre Dame

Donna Macomber and Lesley Hart
Yale University

John Chapman
State of Connecticut, Court Support Services Division

Elena L. Grigorenko
Yale University

The Delinquency Reduction Outcome Profile (DROP) is a novel situational-judgment test (SJT) designed to measure social decision making in delinquent youth. The DROP includes both a typical SJT scoring method, which captures the deviation of an individual response from an “ideal” expert-based response pattern, as well as a novel “Multiconstruct-Multisituational” (MCMS) factor-scoring method, enabling the assessment “in context” of latent dimensions reflecting stable decision-making tendencies. The authors present the development and validation of the DROP across 2 studies establishing its reliability and internal and concurrent validity using a sample of 1,922 young detainees and a sample of juveniles from the community. The authors also discuss the potential usefulness of the DROP as a prognostic tool to predict recidivism for delinquent youth and to monitor changes in intervention programs designed to improve social decision-making skills. Benefits of the MCMS scoring approach for SJT literature and psychological measurement are also discussed.

Keywords: juvenile delinquency, decision making, SJT scoring, MTMM approach

Juvenile delinquency is a serious problem in the United States. Over 2 million juveniles aged 10–17 years were arrested in 2008, with 96,000 of them having been charged with a violent crime (e.g., Puzzanchera, Adams, & Sickmund, 2010). One factor explaining the emergence and persistence of delinquent behavior is the poor social decision-making skills exercised by these juveniles (e.g., Fried & Reppucci, 2001; Steinberg, 2009). Social decision-making skills are defined by one’s ability to effectively communicate, reason, and compromise with peers, teachers, parents, and other adults: An individual must be able to think clearly, predict future outcomes, and make appropriate decisions even during times of stress. Research shows that social decision-making skills are a strong predictor of antisocial behavior, especially as children enter and progress through adolescence (Fontaine, 2008). Accordingly, numerous intervention programs aimed at preventing recidivism in juvenile delinquents tend to focus on improving these decision-making skills (e.g., Blake & Hamrin, 2007; Markowitz & Salvatore, 2005).

Given the importance of social decision-making skills for delinquency outcomes, it is surprising that there are so few measurement tools designed to assess this construct in juvenile forensic settings. Most of the measurement tools that could be used in forensic settings assess antisocial behavior, psychopathic tendencies, and antisocial attitudes rather than social decision making specifically. To date, only Fried and Reppucci (2001) have created a measure (the Criminal Decision Making Questionnaire [CDMQ]) that focuses on detecting individual differences in social decision-making skills among youth offenders. The CDMQ has proved to be an appealing measurement tool; however, it still has a number of limitations. For example, the CDMQ uses a video-
vignette format, which makes it difficult to administer to a large enough sample of delinquents (time-consuming and individual administration required), to develop benchmarks. Furthermore, the CDMQ has not been able to demonstrate the psychometric properties (e.g., test score reliability and validity) required for proper use in any high-stakes, especially forensic, assessment.

The goal of the present research was to develop and carry out preliminary validation for a measure of social decision-making skills that would be easy to administer to large samples (e.g., did not require intensive interview, parental report, or video vignettes), is easy to score (via a scoring program), is cost-effective, and that could be used in the juvenile justice system. To this end, we created the Delinquency Reduction Outcome Profile (DROP), which ultimately aims to (a) facilitate research and extend knowledge about the connection between social decision-making skills and juvenile delinquency; (b) quantify and monitor changes in intervention programs designed to improve social decision-making skills and, eventually, reduce recidivism; (c) contribute to the identification of juveniles at risk for delinquency and recidivism.

Development of the DROP

Scale Content

We developed the DROP as a Situational Judgment Test (SJT; e.g., Motowidlo, Dunnette, & Carter, 1990). A typical SJT presents participants with hypothetical or real situations for which they have to select one of multiple possible courses of action. We chose to use an SJT approach because prior research indicates that this methodology is (a) particularly well suited for assessing individual differences in traitlike constructs such as personality, judgment, decision making, and interpersonal skills (Hanson, Horgen, & Borman, 1998); (b) more likely to have realistic test content and, in turn, more reliable respondent responses (e.g., Chan & Schmitt, 2005); (c) a robust predictor of situation-specific behaviors (e.g., Christian, Edwards, & Bradley, 2010); (d) less aversive to ethnic minority groups (e.g., Nguyen & McDaniel, 2003), which are typically overrepresented in forensic facilities (Hisia, Bridges, & McHale, 2004); (e) more engaging for participants—an important feature for unmotivated individuals such as incarcerated juveniles.

Following the general guidelines for SJT development (e.g., Lievens, Peeters, & Schollaert, 2008; Motowidlo et al., 1990), the first step in creating the content of the DROP was to develop a set of social situations that juvenile offenders might encounter in their daily lives. We used an expert panel to generate a collection of situations often encountered by youth involved with the juvenile justice system. The expert panel consisted of 11 participants, including three members of the research team, three staff members working for the State of Connecticut Court Support Services Division (CSSD), three probation officers, and two other professionals working in detention centers—all of whom served this subpopulation of youth and, thus, were intimately familiar with various aspects of the experiences of this cohort both in and out of detention. The group was composed of men and women of African American, Asian, Hispanic, and Caucasian descent. When developing a new instrument, it is important to consider the ethnic, linguistic, and cultural composition of the target population (in this case, youth in the justice system; Grisso & Underwood, 2004). Thus, we sought to match the cultural composition of our experts with the cultural composition of the target population to maximize the cultural sensitivity of the test content. It was particularly important to involve experts who are themselves members of ethnic minority groups in the development of the DROP situations, because these groups are disproportionately represented in the juvenile justice system (Hisia et al., 2004).

The set of situations was developed on the basis of the expert panel’s experiences as well as from detention center sources (e.g., detention center violation reports and crime and probation reports). The panel generated approximately 30 situations that covered two contexts: (a) situations that could take place while in the detention center and (b) situations that could occur after release or while on probation. The DROP vignettes sampled a broad range of situations across these contexts to ensure that the instrument could provide information about whether or not a participant’s response pattern was consistent across situations (perhaps reflecting a more traitlike type of delinquency) or was situation-specific (perhaps reflecting a more transitory type of delinquency). On the basis of expert group consensus, 16 situations (eight situations within each context) were selected for the final version of the instrument. The final set of situations was chosen because the situations were deemed to be the most realistic, clearly involved decision making, and had a high frequency of occurrence for this population. The situations were edited into stems of similar length and format.

Next, each expert generated several potential responses to each situation. Responses refer to distinct strategies that young detainees could use to deal with the situations. It is worth noting that subject matter experts are especially useful for this type of task because they are able to identify “best” responses as well as generate some less optimal responses (Bergman, Drasgow, Donovan, Henning, & Juraska, 2006). Thus, the responses generated were largely based on expert knowledge and observations of actual juvenile behaviors. Three general categories of responses were generated: adaptive decision-making responses, which likely lead to positive outcomes for the proposed situation; maladaptive decision-making responses, which likely lead to negative outcomes when dealing with the situation; and, avoidance decision making, which sometimes leads to positive outcomes and sometimes leads to negative outcomes in a given situation. Seven answers per situation were selected on the basis of three criteria. First, the proposed answer had to show high consensus among the experts (i.e., being proposed multiple times and judged as highly relevant among experts). Second, response sets to each situation had to cover a range of effectiveness by representing the three decision-making strategies identified in the pool of answers (at least one answer representing each strategy in each situation). The third criterion was to select responses that clearly involved social decision-making skills, as identified by the subject matter experts. The final response options were then edited into stems of similar length and format.

Finally, we created the rating scale and participant instructions. Consistent with recent research on SJT development (e.g., McDaniel, Whetzel, Hartman, Nguyen, & Grubb, 2006), each response was accompanied by a 7-point Likert scale ranging from 1 (very bad choice) to 7 (very good choice). Instructions informed participants to rate each response on the basis of what they thought they “should do.” We chose to have participants rate how they “should” behave (i.e., knowledge-related instructions) rather than how they “would” behave (i.e., behavioral tendency instructions).
because this phrasing tends to be more discerning of stable and cross-situation constructs such as personality (McDaniel et al., 2006), which the DROP is intended to capture. Moreover, recent studies (e.g., Hooper, Cullen, & Sackett, 2006; Nguyen, Biderman, & McDaniel, 2005; Peeters & Lievens, 2005) indicate that knowledge-related instruction is more resistant to faking (i.e., deliberate response distortion) as compared with behavioral tendency instruction. Also, knowledge-related instruction focuses on the participant’s discernment of which social behavioral norms are relevant for a given SJT situation (Kanning & Ruhne, 2006). Such a component of “social desirability” (which is distinct from “faking”; see McCrae & Costa, 1983; Peeters & Lievens, 2005) appears particularly relevant to address among juveniles with a rule-breaking history.

In addition, we chose to have participants rate all of the response options instead of simply choosing a single “best” response option because prior research suggests that this strategy leads to the highest level of internal consistency of the test scores (Ployhart & Ehrhart, 2003), and it allows for an assessment of the participants’ ability to discriminate the range of effectiveness among the alternative answers (and not only to pick the most appropriate), regardless of how they would act in each of the situations. The readability of the resulting DROP survey (instructions, situations, and response options) was assessed using the Lexile Analyzer, which measures the complexity of the text and evaluates its characteristics. Results indicated a Lexile measure of 393L, a mean sentence length of 7.80, and a mean log word frequency of 3.78, which translates to a second-grade reading level. See the Appendix for an example of a DROP vignette.

Scoring System

Bergman and colleagues (2006) identified six general approaches for scoring SJTs: (a) empirical—items or options are scored according to their relationships with a criterion measure; (b) theoretical—a theory is used to identify the best and worst options in a completed test; (c) expert-based—scoring keys are developed on the basis of the responses of individuals with substantial knowledge about the topic; (d) factorial approach—construct-laden scales based on factor analysis of the item pool; (e) subgrouping—an identification of typologies of individuals based on similar patterns of response underlying different self-models that guide their behavior in a range of situations; and, (f) hybrid scoring—a combination of two independently generated scoring keys. According to Bergman et al. (2006), each of these scoring systems produces different results, and each has a number of significant limitations. For instance, it is likely that theoretical scoring may not capture the multidimensionality of the situational response inherent to SJTs (cf. Chan & Schmitt, 2005), whereas factorial approaches aiming to measure stable tendencies may be biased by the situational variance necessarily represented in the SJT’s answers (Westring et al., 2009). Taken together, research suggests that the existing SJT scoring approaches might be limited for the reliable detection of individual differences in stable and cross-situational tendencies.

To overcome the limitations of traditional SJT scoring approaches, we developed a new scoring method for the DROP that we called the Multiconstructs/Multisituational Scoring approach (MCMS). This method should improve the SJT test scores’ reliability (due to the better partitioning of SJT answer variance) and consequently contribute to the improvement of other psychometric properties such as test score criterion validity. We based our new scoring approach on work by Westring and colleagues (2009), who underlined that SJT scores have both construct- and situation variance, which could be unpacked by analyzing the items’ correlation matrix as a multitrait-multimethod matrix (MTMM; Campbell & Fiske, 1959), using a general confirmatory factor analysis (CFA) in a structural equation modeling (SEM) framework. The MTMM approach was originally designed to assess the construct validity (convergent and divergent validity) of a set of measures, which allows an examination of variance that is due to traits, and variance that is due to method and uniqueness or error variance. By extension, Westring et al. (2009) analyzed MTMM data using an SEM approach to partition variance in SJT item responses into three independent components: (a) one reflecting the particular situation to which an item response belongs (situational variance); (b) one related to the psychological construct (trait variance); and (c) a uniqueness component reflecting measurement error as well as other unmodeled variance, such as Situation × Trait interactions.

Accordingly, we developed the MCMS model to score the DROP so that we could create a “trait” factor score for each of the three types of social decision-making styles measured (i.e., create a trait score for adaptive, maladaptive, and avoidance decision making), which would be exempt from situation-specific and uniqueness variance. The MCMS model for the DROP is depicted in Figure 1 for three situations with three responses each, but extends itself to the 16 situations with seven options each in a straightforward manner. As shown in Figure 1, each item response is loaded onto one situation factor and one decision-making factor. On the basis of our theoretical assumption that the decision-making factors could be oblique (decision-making strategies may be conceptually related), these factors are allowed to intercorrelate. Similarly, the model assumes specific variance within situations, as well as a shared variance between situations as a function of the context in which they occurred (situations occurring at home should correlate, whereas situations occurring at the detention center should correlate). We specified the parameters of the factorial model consistently by allowing intercorrelations between the situation factors within two possible contexts (home and detention center). However, situation and trait factors (decision-making factors) were assumed to be uncorrelated with one other so that the variance in each SJT answer could be partitioned into two independent components (respective decision-making factor and respective situation). Finally, the model estimates the uniqueness associated with each item response. Using this new MCMS scoring approach, we were expecting to derive “pure” decision-making strategy scores devoid of situational and uniqueness variance (i.e., the development of an empirically driven scoring key based on the estimated contribution of each DROP item into each of the “trait” factors only) that would represent stable individual tendencies in social decision-making strategies across situations. The DROP’s MCMS model was tested in Study 1 (see below), along with the

---

1 Copies of the DROP vignettes and scoring program are available upon request from Elena L. Grigorenko.
parameters used as a basis of the development of the DROP’s MCMS factor-scoring key.

We also used a traditional SJT scoring system in order to compare the scores obtained with the novel MCMS approach. Specifically, we developed an expert-based scoring system that consisted of comparing each participant’s response profile with an expert or “ideal” response profile—in the sense that experts are the proper reference to determine the degree of appropriateness of each response option or to establish an “optimal” response pattern. We developed this expert response pattern by administering the DROP to an ethnically diverse group (with African American, Asian, Hispanic, and Caucasian descent) of 13 male and 11 female CSSD staff members, who were not involved in the development of DROP vignettes. As before, it was important to recruit a diverse group of experts because the interpretation of appropriate behavioral responses to a specific situation might differ as a function of cultural values (Grisso, 2005; Lievens, 2006). This matching of cultural compositions was particularly critical because the expert response profiles would be used to evaluate the responses of the juveniles. The CSSD employees were nominated as experts due to their experience and accomplishments within the CT juvenile justice system. Each expert was asked to read the DROP vignettes and to rate the effectiveness of each related alternative answer (i.e., to assess to which degree each answer is appropriate given the proposed situation) using the 7-point Likert scales (for a total of 112 scored items). As shown in Table 1, the interexpert agreement was high across situations with an intraclass correlation coefficient ranging from .53 (Vignette 7) to .86 (Vignette 15), with a mean of .70, and Cronbach’s alpha ranging from .96 to .99, with a mean of .98. No gender differences between experts were found at the item level, F(1, 22) = 1.40, p = .25.

These analyses confirm that the experts strongly agreed on the degree of effectiveness of the alternative answers related to each DROP situation, allowing us to reliably combine expert ratings into a single “optimal answer profile” against which tested

Table 1
Interrater Agreement for Each DROP Vignette and Estimates of Variance Accounted for by Situation and Decision-Making Factors (Percentages)

<table>
<thead>
<tr>
<th>Vignette</th>
<th>ICC</th>
<th>α</th>
<th>Situation variance</th>
<th>Decision-making variance</th>
<th>Uniqueness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vignette 1</td>
<td>.67</td>
<td>.98</td>
<td>8.39</td>
<td>75.46</td>
<td>16.15</td>
</tr>
<tr>
<td>Vignette 2</td>
<td>.70</td>
<td>.98</td>
<td>23.49</td>
<td>25.46</td>
<td>51.05</td>
</tr>
<tr>
<td>Vignette 3</td>
<td>.64</td>
<td>.98</td>
<td>29.01</td>
<td>34.60</td>
<td>41.31</td>
</tr>
<tr>
<td>Vignette 4</td>
<td>.76</td>
<td>.99</td>
<td>19.91</td>
<td>21.82</td>
<td>58.27</td>
</tr>
<tr>
<td>Vignette 5</td>
<td>.68</td>
<td>.98</td>
<td>9.85</td>
<td>48.78</td>
<td>41.37</td>
</tr>
<tr>
<td>Vignette 6</td>
<td>.62</td>
<td>.97</td>
<td>27.34</td>
<td>29.08</td>
<td>43.58</td>
</tr>
<tr>
<td>Vignette 7</td>
<td>.53</td>
<td>.96</td>
<td>8.87</td>
<td>30.30</td>
<td>60.83</td>
</tr>
<tr>
<td>Vignette 8</td>
<td>.66</td>
<td>.98</td>
<td>31.91</td>
<td>26.51</td>
<td>41.59</td>
</tr>
<tr>
<td>Vignette 9</td>
<td>.62</td>
<td>.97</td>
<td>17.91</td>
<td>31.74</td>
<td>50.35</td>
</tr>
<tr>
<td>Vignette 10</td>
<td>.60</td>
<td>.97</td>
<td>24.28</td>
<td>23.85</td>
<td>51.87</td>
</tr>
<tr>
<td>Vignette 11</td>
<td>.72</td>
<td>.98</td>
<td>8.06</td>
<td>51.18</td>
<td>40.76</td>
</tr>
<tr>
<td>Vignette 12</td>
<td>.76</td>
<td>.99</td>
<td>9.40</td>
<td>24.65</td>
<td>65.95</td>
</tr>
<tr>
<td>Vignette 13</td>
<td>.80</td>
<td>.99</td>
<td>11.84</td>
<td>51.00</td>
<td>37.16</td>
</tr>
<tr>
<td>Vignette 14</td>
<td>.79</td>
<td>.99</td>
<td>15.13</td>
<td>46.96</td>
<td>37.92</td>
</tr>
<tr>
<td>Vignette 15</td>
<td>.86</td>
<td>.99</td>
<td>11.95</td>
<td>51.49</td>
<td>36.56</td>
</tr>
<tr>
<td>Vignette 16</td>
<td>.84</td>
<td>.99</td>
<td>14.25</td>
<td>33.14</td>
<td>52.60</td>
</tr>
<tr>
<td>All</td>
<td>.70</td>
<td>.98</td>
<td>15.70</td>
<td>42.20</td>
<td>42.15</td>
</tr>
</tbody>
</table>

Note. DROP = Delinquency Reduction Outcome Profile; ICC = intraclass correlation coefficient.
subject–answer profiles could be compared. To this end, we computed the squared Mahalanobis distance ($D^2$) of each participant’s ratings for each possible solution in a given vignette from the centroid of the expert sample. The resulting 16 vignette-level $D^2$ distance values per individual were then averaged (based on standard values) to determine the DROP’s total expert-based score. A high score would reflect a high degree of dissimilarity with expert “optimal” judgment and, in turn, reflect poor social decision making with higher likelihood for delinquency problems. The squared Mahalanobis distance used to compute this score is a standardized distance that adjusts differences from the mean on a particular vignette (i.e., set of alternative answers or “items”) for two criteria: (a) the variance in the expert ratings for each item and (b) the intercorrelation between the item’s ratings within the vignette (e.g., Rencher, 1995). Preliminary analyses did not reveal any gender differences among experts, $F(1, 22) = 1.57, p = .22$, on this expert-based score (i.e., total distance score).

Empirical Evaluation of the DROP

We conducted two studies to validate the DROP. In Study 1, we tested the hypothesized MCMS internal structure of the DROP using a CFA, and we derived its MCMS factor-scoring key. We examined the MCMS scores’ test–retest reliability, their overlap with expert-based scores, and their ability to distinguish single versus repeat youth offenders. In Study 2, we examined the DROP scores’ concurrent validity with constructs related to delinquency and mental health as well as social difficulty measures.

Study 1

Overview

The goal of Study 1 was to confirm the internal structure of the DROP and derive the corresponding MCMS factor-scoring key and test scores. Second, we aimed to evaluate the MCMS test scores’ reproducibility, their overlap with expert-based scores, and their ability to distinguish single versus repeat youth offenders.

Method

Participants. The sample consisted of 1,922 youths (1,452 boys, 470 girls) recruited from three juvenile detention centers in the State of Connecticut during a 3-year span, between November 2006 and November 2009. The youths’ ages ranged from 11 to 17 years old ($M = 14.35, SD = 1.02$, mode = 15). The racial-ethnic groups represented in the sample included African American ($n = 791, 40.2\%$), Caucasian ($n = 532, 27.7\%$), Hispanic ($n = 529, 27.5\%$), and “other” ($n = 21, 1.1\%$). Information regarding ethnicity was missing for 3.6% of the sample ($n = 67$). This sample was representative of juvenile delinquency statistics described in the literature, with an overrepresentation of males (Snyder, 2008), an overrepresentation of ethnic minorities (Hisia et al., 2004), and a modal age of 15 (e.g., Landsheer & van Dijkum, 2005). The reasons for incarceration of the juvenile involved in this study were not collected due to privacy concerns; however, general indicators on the juvenile delinquency population in the State of Connecticut could be used to describe our sample. First, the breakdown for the admission criteria in juvenile detention facilities over the period of data collection consisted mainly of Order of Detention (34.5\%), Take in Custody (30.3\%), Warrant (17.3\%), and Serious Juvenile Offense (14.1\%). All but the last category involved cases that required a written order of a judge to allow incarceration. It is impossible to determine the severity of the charges that these admissions represent; however, the Serious Juvenile Offense category generally involves a significant offense, often but not always the equivalent of an adult “felony” charge. Second, the most common charge for the arrest of juveniles in Connecticut is third-degree assault or “Simple Assault” (27\%), that is, assaults that are not of an aggravated type and do not involve a weapon, followed by Disorderly Conduct (19\%), Larceny (18\%), Vandalism (6\%), Drug Abuse Violation (5\%), Burglary (3\%), and first- and second-degree assaults or aggravated assaults (3\%), that is, the attack by one person on another to inflict severe bodily harm often accompanied by the use of a weapon. The average length of stay in juvenile detention facilities during the period of data collection was 14 days, with 59% of juveniles staying over a period of 1–13 days, 25.1% staying over a period of 14–27 days, 12.3% staying between 28 and 59 days, and 3.5% staying over 60 days.

Procedure. The DROP was administered by the detention center Classification and Program Officers to youth within 48 hr of their entry into detention, either individually or in small groups. Permission to administer the DROP was obtained from both parents and juveniles, embedded in the general permission to treat. If juveniles were detained more than once during the duration of the data collection plan, they received the DROP at each admission. The number of admissions during the study period ranged from 1 to 16, which resulted in the following distribution of DROP administrations: $n_1 = 1,413 (73.5\%), n_2 = 240 (12.5\%), n_3 = 124 (6.5\%), n_4 = 53 (2.8\%), n_5 = 30 (1.6\%), n_6 = 20 (1\%), n_7 = 10 (0.5\%), n_8$ and beyond = 32 (1.7\%). This distribution was transformed into two groups for the planned analyses: “Single Offenders” ($n = 1,413; 73.5\%$) and “Repeat Offenders” ($n = 509; 26.5\%$), referring to participants who reentered the detention center during the 3-year data collection period. Preliminary analyses revealed that the mean age for the “Repeat Offenders” group ($M = 14.64, SD = 1.09$) was slightly higher than for the “Single Offenders” group ($M = 14.38, SD = 1.08$), $F(1, 1920) = 21.48, p < .001$, which is consistent with the fact that older juveniles have a higher probability to have been previously arrested than younger juveniles. A significant difference in the distribution of ethnicity across these groups was also observed, $\chi^2(4) = 17.5, p < .01$, indicating a higher probability of rearrest (i.e., being part of the “Repeat Offenders” group) for the African American (29.7% are rearrested within this population) and Hispanic juveniles (30.2% than for the Caucasian juveniles (20.7%). This higher probability of minorities to be arrested or rearrested is consistent with previous data discussed in the literature (e.g., Feld, 1999). No significant differences in the distribution of gender were observed between the Single Offenders and the Repeat Offenders groups, $\chi^2(1) = 1.48, p = .22$. It is important to note that some juveniles included in the Single Offenders group could have been rearrested after the period of data collection, whereas the number of juveniles in the Repeat Offenders group could have an increased number of administrations (i.e., they could have been rearrested multiple times). Furthermore, it is important to keep in mind that rearrest is not
synonymous with reoffense; some “Single Offender” participants may have reoffended without being rearrested. Finally, it is also worth noting that repeated admissions within the Repeat Offenders group was time-variant, meaning that two individuals with the same number of arrests (e.g., two) could have been arrested with different delays.

**Data analytic plan.** The DROP’s MCMS factorial model was tested using CFAs in an SEM approach. Because chi-square tests of fit are highly sensitive to large sample size, practical fit indices were emphasized in assessing the model fit. Four model fit indices were used: (a) the Tucker–Lewis index (TLI; Tucker & Lewis, 1973), (b) the comparative fit index (CFI), (c) the root-mean-square error of approximation (RMSEA; Steiger, 1990), and (d) the standardized root-mean-square residual (SRMR). A subsample of 100 youth from the Multiple Offenders group, who reentered the detention center within a period of 25–45 days, was used to estimate the test–retest reliability of the DROP. With the exception of this analysis, however, the analyses presented in this study have been exclusively conducted with the data collected at first exposure to the instrument.

**Results**

We conducted a set of preliminary analyses to confirm that the data exhibited the properties needed for factor analysis. First, we examined the distributional features of the DROP’s item-level responses for each of the 16 DROP situations (total of 112 items). The means for each item ranged from 1.7 to 6 (\(M = 3.3, SD = 1.3\)), and the standard deviations ranged from 1.5 to 2.5 (\(M = 2, SD = 0.2\)); skewness statistics ranged from −1.7 to 2.4 (\(M = 0.5, SD = 1, SE = .06\)), whereas kurtosis statistics (flatness of the distributions compared with the normal distribution) ranged from −1.6 to 4.4 (\(M = 0, SD = 1.4, SE = .1\)). Taken together, these results indicated that the distributional features of the data set were acceptable for running the planned analysis. Next, we examined the properties of the correlation matrix obtained between the 112 DROP items to ensure its suitability to be used in the planned factor analyses. Results showed that the correlation matrix was not an identity matrix (Bartlett’s test of sphericity = 64011.4, \(df = 6216, p < .001\)), and the global measure of sampling adequacy was excellent (.96), reflecting individual measures of sampling adequacy that were good to excellent (ranging from .78 to .98). Correlations between the 112 DROP items ranged from −.48 to .72, with a mean correlation of .09. In sum, the correlation matrix showed the proper features to be used in a factor analyses.

Given this, we performed a principal component analysis (PCA) using Promax rotation and a hierarchical cluster analysis (HCA) of the 112 DROP items, and both supported a data reduction solution composed of three factors that presumably tap into the three social decision-making factors that the DROP aims to capture (meaningful item-factor loadings). These results also supported the specification of the MCMS factorial model by confirming the loading of each item into the corresponding social decision-making factors (correspondence supported by the item-factor saturations observed through the PCA and the organization of the items in the three clusters identified with the HCA).

**Testing the DROP’s MCMS factorial model.** We tested our MCMS factorial model (see Figure 1) using maximum likelihood estimation and showed an acceptable fit of the data to the model, taking into account its complexity (i.e., low parsimony), \(\chi^2(5379) = 23549.3, p < .001; \chi^2/df = 4.4, CFI = .81, TLI = .80, RMSEA = .042, SRMR = .073.\) With reference to general benchmarks in the literature, these results can be interpreted as adequately fitted, because models with TLI and CFI between .80 and .90 fit moderately well (e.g., Browne & Cudeck, 1993; Steiger, 1990). An RMSEA value of .05 or less is also considered to be representative of good-fitting models (Browne & Cudeck, 1993). The reasonable error of approximation of the model allowed us to examine in depth the percentage of variance estimated to account for the diverse factors. The average variance accounted for by the situational factors was about 16%, whereas the average variance accounted for by social decision-making factors (i.e., “trait” factors) was about 42% (the 42% remaining were accounted for by the uniqueness variance). These values masked a considerable variability across the situation and decision-making factors, as reflected in Table 1. Indeed, as a function of the vignettes, the situation variance ranged from 8.1% (Vignette 1) to 31.9% (Vignette 8), whereas decision-making variance accounted for by items ranged from 21.8% (Vignette 4) to 75.5% (Vignette 1). Vignettes with high levels of uniqueness (e.g., Vignette 12) suppose a high level of specificity of both the decision-making style and the situational component mobilized by the scenario (i.e., “unmodeled” or error variance; Westring et al., 2009). On the basis of this model, we derived the three Social Decision-Making factors scores (i.e., “trait” factors scores representing adaptive, maladaptive, and avoidant-type social decision making) by developing a scoring syntax computing the factor scores exempted of situational and uniqueness variance, using the regression method. This syntax allowed the automatic computation of both the expert-based score and the three MCMS factor scores to facilitate the DROP’s complex scoring in practical settings.\(^2\)Distributional features of the resultant scores obtained in this study indicated optimal variability, suggesting good sensitivity of the DROP’s test scores (see Table 2).

**Comparison of the DROP’s MCMS factor scores and expert-based scores.** Correlations between the DROP’s social decision-making factor scores and expert-based scores are presented in Table 2. As shown, these correlations were in the moderate to high range. In particular, distance to the expert profile was highly correlated with maladaptive decision making (\(r = .86, p < .001\)). This finding is consistent with our expectations, because a high expert-based score (i.e., distance to expert) reflects a high dissimilarity with expert judgment. Conversely, the expert-based score was negatively and strongly correlated with the adaptive decision-making factor score (\(r = −.65, p < .001\)), whereas the avoidance-type decision-making factor score showed a similar pattern of results as the adaptive decision-making style (a high correlation between both types of strategies: \(r = .81, p < .001\)). This suggests that, in most cases, avoidance-type decision making can be interpreted as adaptive in the context of the situations given

\(^2\) The scoring program is available, along with the DROP instrument, upon request from Elena L. Grigorenko.
by the DROP. Interestingly, the correlation between adaptive decision making and maladaptive decision making was not as high as one might have expected ($r = -0.46, p < .001$).

**DROP scores’ test–retest reliability.** We estimated the test–retest reliability of DROP scores using a subsample of 100 youths to whom the DROP was readministered within a range of 25–45 days. The test–retest reliability coefficients were satisfactory for the expert-based score ($r = .75, p < .001$), Adaptive ($r = .60, p < .001$) and Maladaptive ($r = .61, p < .001$) decision-making factor score, and lower but still acceptable for the Avoidance decision-making factor score ($r = .53, p < .001$). This relatively weaker stability could indicate that Avoidance-type decision making is more situational and sensitive to change. Conversely, the expert-based score is a broader scoring (it sums up the information contained in the 112 DROP items), resulting in a higher score stability of measurement. Overall, although the DROP was designed to be sensitive to change, there was enough consistency in the scores to capture reliably the “core” elements of the youths’ reasoning about the situations.

**Complementary analyses.** We applied a multivariate analysis of covariance on the DROP MCMS factor scores and the expert-based score in order to identify group differences in the use of the decision-making strategies and the distance to the “optimal” expert profile. We entered Principal factors (Gender, Ethnicity, and Repeat Offense group) separately and in combination to identify potential interactions between these variables. We used age as a covariate to control for possible age effects. For the purpose of this analysis, participants included in the underrepresented ethnicity group (1.1% of the total sample) were discarded. These analyses revealed significant effects of gender, $F(4, 1813) = 3.58, p = .006, \eta^2 < .04$; ethnicity, $F(8, 3626) = 6.42, p < .001, \eta^2 < .04$; and repeat offense, $F(4, 1813) = 4.47, p = .001, \eta^2 < .04$, on the DROP scores. No interactions between these factors were identified, and the effect of age was not significant, $F(4, 1813) = 1.12, p = .35$. More precisely, Tukey’s honestly significant difference tests adapted for unequal $N$ indicated that girls tend to score higher than boys on the Adaptive Decision-Making scale ($p = .025$), but do not differ significantly from boys on the other scores. Conversely, Caucasian youth tended to score lower than the other ethnic groups on the expert-based score ($p < .001$), meaning that Caucasian response profiles were closer to the expert group and tended to score higher on the Avoidance-type decision-making factor score compared with the African American group ($p < .001$). Finally, we observed larger effects for the Single Offenders group, with significantly higher scores for the Repeat Offenders group on the expert-based score ($p < .001$), and significantly lower scores for Adaptive ($p < .001$) and Avoidance-type ($p < .001$) decision-making factor scores.

**Discussion**

Study 1 established the internal validity of the DROP and provided support for a novel MCMS factor scoring system. Our hypothesized factorial model fit the data, and the derived decision-making factor scores were consistently correlated with traditional expert-based scores. Moreover, the MCMS scores showed good test–retest reliability, but were still sensitive enough to capture change even during a relatively short follow-up period. Taken together, these results provide the first empirical support for the DROP’s MCMS scoring method.

Study 1 also highlighted the additional value of the MCMS scoring method over the expert-based method. Indeed, the MCMS factor scores provide much more information than the expert-based score. For example, two individuals could show highly divergent MCMS profiles in terms of their responses on the three decision-making factors yet exhibit the same distance to the expert score. Our results also indicated a moderate correlation between adaptive and maladaptive decision-making scores. This suggests that these dimensions are not opposite poles of the same dimension but rather that both adaptive and maladaptive decision-making responses can be, for certain juveniles, considered simultaneously as appropriate (or inappropriate). It was only by using the MCMS score that we were able to identify these “indiscriminant” individuals, which might prove to be an important subtype of delinquents. Finally, the DROP decision-making profile was a predictor of risk for reoffense. Thus, it is possible that the DROP could identify profiles of individuals at greater risk for recidivism, which could assist in designing prevention and intervention programs adapted to young offenders’ social decision-making style. As a whole, our new MCMS scoring approach suggests both theoretical and practical usefulness, opening new directions for SJT scoring—a recurrent methodological issue (e.g., Bergman et al., 2006).

**Study 2**

**Overview**

The primary goal of Study 2 was to evaluate the concurrent validity of the DROP scores by comparing them with measures of constructs related to delinquency, mental health, and social diffi-
A secondary objective of Study 2 was to administer the DROP in a sample of juveniles from the community (rather than a sample of detainees), based on the assumption that (a) the structure of the DROP is consistent across samples of juveniles, (b) the MCMS scoring model applies to community samples, and (c) it can help to identify juveniles at risk for delinquency.

Method

Participants. Seventy-five youths from the community participated in this study. Participants attended schools in the State of Connecticut and were sampled from a variety of settings to represent the diversity of the school-age population in the state. There were 25 boys (33.3%) and 45 girls (60%); five (6.7%) did not indicate gender. The age range was 10–19, with a mean of 14.49 and standard deviation of 1.8. The sample included 61.3% Caucasian (n = 46), 22.7% African American (n = 17), and 10.6% “other” (n = 8) ethnicities. Ethnicity data were missing for four individuals (5.3%). The majority of youths (n = 51, 68%) reported that they had never been arrested, whereas some reported being arrested once (n = 3, 4%) or twice (n = 3, 4%). Eighteen youths (24%) did not respond to the question.

Measures. A revision/extension of the Social and Health Assessment (SAHA; Schwab-Stone et al., 1995; Weissberg, Voyce, Kasprów, Arthur, & Shriver, 1991) served as the primary external criterion for this study. The SAHA is a multiscale assessment of the risk and protective factors in the social and health development of school-age youth, which includes various validated scales such as the Strengths and Difficulties Questionnaire (SDQ; Goodman, Meltzer, & Bailey, 1998), a brief behavioral screening questionnaire (peer relationship problems, and prosocial behavior); a somatic symptoms scale, the Child Post-traumatic Stress-Reaction Index (Nader, 1996), the Disapproval of Deviancy Scale, and the Antisocial Behavior Scale (further details on the SAHA survey, including evidence of its strong psychometric properties across several countries, are available in the literature; e.g., Schwab-Stone et al., 1995; Vermeiren, Schwab-Stone, Deboutte, Leckman, & Ruchkin, 2003).

Procedure. Participants agreed to take part in the study and signed a minor assent form, while parents signed a parental consent form to let their child participate. The DROP was embedded within the SAHA survey. Data were collected by members of the research team in a variety of settings (e.g., schools and after-school programs), in either small groups or individually. The youths were monetarily compensated for their participation. Expert-based scores and MCMS decision-making factor scores were computed using the scoring program developed in Study 1. The data of incarcerated juveniles obtained in Study 1 (which included only DROP scores) were also used in this study to compare the delinquents’ DROP scores with the scores obtained from the community sample.

Results

Preliminary analyses indicated sample differences (i.e., community vs. incarcerated juveniles) on the DROP scores, F(4, 1971) = 5.3, p < .001, even after controlling for gender, ethnicity, and age differences among these samples (note that ethnicity was recombined for the purpose of this analysis into three groups: African American, Caucasian, “other”). Compared with the detained juveniles involved in Study 1, community participants scored significantly lower on the Maladaptive decision-making factor score (p < .01) and higher on the Adaptive (p < .03) and Avoidance decision-making factor scores (p < .04). On the one hand, no differences between the groups were observed for the expert-based score. On the other hand, distributional features of the DROP scores obtained in Study 2 indicated a greater variability, with MCMS factor scores ranging from −1.53 to 1.51 (SDs ranging from .64 to .69) and the expert-based score ranging from −0.93 to 2.94 (SD = 0.64).

The internal consistency of the SAHA scores was high, with Cronbach’s alphas ranging from .65 to .92 (median α = .83). Table 3 presents the internal consistency coefficient of the SAHA scores obtained for this study along with their main distributional features, and the correlation coefficient obtained with the DROP measure. Results revealed moderate to strong correlations between the DROP decision-making factor score and most of the SAHA scores. In particular, the Adaptive decision-making factor score was correlated with the SDQ Prosocial scale (r = .45, p < .001) and intrinsic motivational orientation for school (r = .45, p < .001). The DROP Maladaptive decision-making factor score was associated with the SAHA Readiness to Fight scale (r = .56, p < .001), negative school climate (r = .52, p < .001), as well as somatic anxiety (r = .51, p < .001) and conduct problems (r = .51, p < .001). Finally, the Avoidance-type decision-making factor score was essentially correlated with the intrinsic motivational orientation for school (r = .38, p < .01), the SDQ Prosocial scale (r = .29, p < .05), and school negative climate (r = −.34, p < .001). In sum, associations were quite high between DROP MCMS decision-making factor scores and numerous behavioral and psychological constructs measured by the SAHA, such as readiness to fight (r = .56, p < .001), posttraumatic stress disorder (r = .40, p < .001), conduct problems (r = .51, p < .001), affiliation with deviant peers (r = .36, p < .001), and somatic anxiety (r = .51, p < .001).

The magnitude of the correlations between the DROP expert-based scores and the SAHA variables was slightly stronger than those observed with the MCMS decision-making factor scores taken independently. In particular, the expert-based measure was strongly associated with the SAHA Conduct Problems scale (r = .59, p < .001) and readiness to fight (r = .60, p < .001). However, the shared variance between the MCMS decision-making factor scores and the SAHA scores (total of 21.7%; with 5%, 13.2%, and 3.5% for adaptive, maladaptive, and avoidance decision making, respectively) was greater than the shared variance between the expert-based score and the SAHA scores (15.2%). Although this larger association was obtained in a shared contribution between the DROP’s decision-making factor scores, this suggests an additional value of the MCMS scoring method over the expert-based method.

Discussion

This study established the concurrent validity of the DROP scores and further highlighted the usefulness of the MCMS scoring approach. MCMS factor scores were highly associated with measures of behavioral problems (e.g., delinquency and readiness to fight), prosocial behavior, parental and environ-
mental variables associated with delinquency (e.g., violence exposure, school climate), and other psychological problems such as alcohol use, risk for substance use, and somatic anxiety. The pattern of correlations provided both convergent and divergent validity for the DROP MCMS and expert-based scores. In addition, our results indicated that the MCMS factor scores, used in combination (i.e., profile of decision making), might be a better predictor of behavioral and psychological difficulties than the expert-based score. Finally, our analyses confirmed the applicability of the DROP to juveniles from the community and revealed its capability to differentiate these juveniles from incarcerated juveniles.

**General Discussion**

The DROP was designed to capture a key feature of juvenile delinquency: maladaptive social decision making. We created the DROP using an SJT approach and developed an innovative MCMS scoring technique that overcomes several limitations of the traditional SJT scoring approaches. Special effort and attention was taken in generating the content of the DROP, including using a highly experienced expert focus group to ensure the content validity of the scenarios involving decision-making strategies. We conducted two studies to examine the DROP’s psychometric properties; impressively, we were able to administer the DROP to nearly 2,000 juvenile detainees as well as a sample of juveniles in the community. The results of our studies indicated that the DROP is empirically sound with good psychometric properties. Specifically, the DROP exhibited (a) internal validity with strong support for the hypothesized three-factor decision-making structure; (b) good test–retest reliability while exhibiting enough variation to indicate that it is sensitive to change; (c) the additional value of the MCMS scores over an expert-based scoring system; (d) the ability to distinguish single from repeat offenders; (e) convergent and divergent validity with violent and nonviolent delinquency measures (e.g., readiness to fight, violence exposure, school climate, and other psychological problems such as alcohol use, risk for substance use, and somatic anxiety measures); and (f) applicability to community samples with the potential for identifying juveniles at risk for delinquency.

In addition to filling a critical need in the area of juvenile delinquency, the development of the DROP provided a unique opportunity to design and apply a new scoring approach for SJTs—MCMS factor scoring. This approach should prove useful for a number of reasons. First, MCMS scoring offers a way to measure constructs in simulated situations while capturing the core elements of the constructs across contextual specificities. This represents an advance not only in the situational judgment test literature but also for psychological testing in general. Indeed, the MCMS approach takes advantage of the situational judgment-testing format (e.g., increased predictive power and reduced ethnic biases), while providing the same features as the classic psychometric approach (i.e., involving empirical scoring). The results of our two studies underscore the usefulness of the MCMS scoring approach as compared with the traditional expert-based scoring on both an empirical (e.g., meaningful association with criterion measures) and a practical level (MCMS profile more informative than a unitary “distance” score). Thus, the MCMS approach appears to be useful beyond the separation of situational and construct variance and represents new directions in psychological assessment.
that can allow the measurement of constructs in context (simulated) to infer stable predispositions across situations.

Although our results provide strong support for the DROP, it is important to remember that these were the first two studies to investigate this measure. Cross-validation studies with diverse samples from across the nation are needed to further validate the DROP and examine whether it is structurally comparable among more diverse populations. Our study was also limited by a relatively short follow-up period. It is necessary to conduct further studies with longer term follow-up intervals in order to create cutoff zones for DROP scores that may help to identify juveniles (both community and incarcerated adolescents) who are at the greatest risk for delinquency.

In addition, the gender differences observed in Study 1 suggest that this factor should be taken into account when measuring the decision making of juvenile delinquents as well as the need to develop gender-specific norms for the DROP. Finally, a limitation of our study is that we did not have access to the reason for incarceration (i.e., specific offense committed) in our delinquent sample. Examining this factor in future research could provide new perspectives for scoring and improving the predictive validity of the DROP. For instance, it is possible that the type of offense might result in unique and distinguishable decision-making profile subtypes. These subtypes could increase the predictive power for identifying one’s risk for particular law violations and recidivism.

Moreover, intervention programs could be tailored to the adolescent’s subtype profile. Such data, used in combination with recidivism outcomes, will support the study of the development of maladaptive decision making in relation to the emergence of a criminal career by situating this development with trajectories of crime. Finally, further exploration of “situational variance” unpacked with the new MCMS approach could be useful for future studies. Specifically, to what extent might this variance help to understand an individual’s “sensitivity” to situational factors that may push her or him toward crime (cf. Piquero & Tibbetts, 1996)? Moreover, could this situational component of the SJT help to identify specific situations associated with risk for maladaptive decision making?

In conclusion, the DROP appears to possess psychometric qualities required for best-practice assessment in forensic settings. It is effective in differentiating youth who are at risk for repeated admissions to detention, while reliably measuring the social decision making that may contribute to violence, rearrest, or behavioral problems. We are presently pursuing the next logical step for the DROP, which is to develop norms for a juvenile delinquent sample (with specific benchmarks as a function of age, gender, and race). Complemented by reoffense data and basis for incarceration data, DROP norms could have the potential to become a powerful source of information to determine whether maladaptive social decision making and antisocial behaviors are due to transient immaturity or contextual disadvantages (Moffitt, 1993), as opposed to deep-seated predispositions to criminality (e.g., Steinberg, 2009). Such a measurement tool is needed as decisions made within juvenile justice systems have important consequences for youth and society, and the quality of these decisions will depend directly on the quality of the information provided about these youth (Hoge, 2012).

References


trajectories from pre through middle adolescence and their continuation in late adolescence. Adolescence, 40, 729–748.


(Appendix follows)
Appendix

Example of Vignette

Situation 1-
Imagine that you are at the detention center. You are at recreation and another person starts calling you names and making fun of you.

What should you do? Put a number from 1 to 7 on each line.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very bad choice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average choice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very good choice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- ____ a) Call them names and make fun of them.
- ____ b) Ask them to stop.
- ____ c) Ask staff to deal with it.
- ____ d) Beat them up.
- ____ e) Ask one of your friends to help.
- ____ f) Walk away.
- ____ g) Ignore them.