Child Youth Care Forum (2015) 44:893–917
DOI 10.1007/s10566-015-9312-3

ORIGINAL PAPER

Content, Structure, and Usefulness of Juvenile Predisposition Psychological Evaluations

Samantha L. Morin · Keith R. Cruise · Holly Hinz · Evan D. Holloway · John F. Chapman

Published online: 4 March 2015

Abstract
Background There is a dearth of research regarding the content and structure of juvenile predisposition psychological evaluations. Limited research suggests that key mental health domains are insufficiently represented and judges use evaluator recommendations regarding legal outcomes more often than clinical outcomes. Studies have not addressed whether content and/or structure of evaluations influence the use of evaluator recommendations by juvenile probation officers (JPOs).
Objective This study reviewed and rated the content and structure of juvenile predisposition psychological evaluations conducted in an East Coast state to identify evaluation characteristics that informed JPOs use of evaluation recommendations in disposition planning.
Methods Juvenile predisposition psychological evaluations ($N = 150$) were reviewed and coded on key variables (e.g., legal history, sociocultural factors, Forensic Mental Health Assessment-FMHA principles, use of empirically-supported tools).
Results Multiple content areas including family history, drug and alcohol history, education history were included in the evaluations (ranging from 48 to 100 %); however, sufficient detail for content domains ranged from 10 to 76 %. Evaluator recommendations were incorporated in disposition plans 35 % of the time, regardless of evaluation content or sufficiency, with 70 % of accepted recommendations being mental health related.
Conclusions Although evaluations often included information covering multiple key content areas, there was significant variability in the amount of information provided. Inconsistent with prior research with judges, none of the variables of interest (presence and sufficiency of detail, well-justified and explained recommendations, adherence to FMHA

S. L. Morin (✉) · K. R. Cruise · H. Hinz · E. D. Holloway
Department of Psychology, Fordham University, Dealy Hall, Room 237, 441 East Fordham Road, Bronx, NY 10458, USA
e-mail: samantha.fusco7@gmail.com

J. F. Chapman
State of Connecticut, Judicial Branch, Court Support Services Division (CSSD), 936 Silas Deane Highway, Wethersfield, CT 06109, USA

principles) predicted use of evaluation recommendations by JPOs. Implications for quality assurance and training are discussed.

Keywords Juvenile justice · Predisposition evaluations · Evaluator recommendations · Empirically-supported tools

Introduction

The research literature concerning juvenile justice-involved youth has demonstrated that as many as 65 % have one or more diagnosable mental health disorders (Teplin et al. 2002, 2006; Wasserman et al. 2002). More specifically, results from a recent study demonstrated that as many as 66 % of detained males and 74 % of detained females met criteria for at least one disorder at detention intake with 46 % of males and 57 % of females having two or more disorders at intake (Teplin et al. 2013). Aside from mental health disorders, the prevalence of other psychosocial problems is also higher in justice-involved youth, including difficulties with anger control (Goldstein et al. 2013; Sarris et al. 2000), family/parenting instability (Cook and Gordon 2012; Maschi et al. 2008), negative peer group affiliation (Burt and Klump 2013; Shapiro et al. 2010), and educational problems (Cavendish 2014; Cruise et al. 2011). Therefore, evaluations for disposition planning should address more than just mental health problems.

Additionally, two themes consistently emerge in current juvenile justice (JJ) systems. Fewer females are represented in the JJ system; however, they present with higher rates of mental health symptoms and psychopathology and are greater consumers of mental health services than their male counterparts (Grisso and Barnum 2000; Vincent et al. 2008). Furthermore, a majority of youth entering the JJ system are African American or Hispanic and they present with high rates of mental health issues; however, they demonstrate lower access to mental health care and decreased service utilization (Teplin et al. 2002; Vincent et al. 2008). For example, research by Dalton et al. (2009) demonstrated that even when mental health screening information was similar between White youth and youth of color, White youth were more often identified as in need of mental health services than youth of color. Additionally, Teplin et al. (2002) found that White youth presented with higher rates of mental health disorders compared to youth of color. Despite high rates of mental health problems among youth in the JJ system, the number of youth receiving mental health care is limited. To illustrate, one study documented that among youth with psychiatric disorders, only 15 % received treatment in detention centers and 8 % received treatment in the community by the time of the case disposition or within six months after detention (Teplin et al. 2013).

Many youth with mental health issues enter the justice system without having been identified or treated. Therefore, greater attention has been paid to practice recommendations for screening and assessment of youth who come into contact with the JJ system (Grisso 2013; Grisso et al. 2005; Wasserman et al. 2003). Progress has been made with regards to enhancement of screening and follow-up mental health assessment procedures (Archer et al. 2010; Grisso 2013; Stewart and Trupin 2003; Vincent 2011; Wasserman et al. 2004). Practice recommendations indicate that youth identified at screening as in need of further assessment be provided a comprehensive evaluation (see Wasserman et al.

2004). However, little empirical research has addressed the content and quality of comprehensive assessments conducted post-screening.

There is a long history of predisposition psychological evaluations being used as part of the management and treatment planning process for youth involved in the JJ system (Grisso 2013; Hecker and Steinberg 2002). Even though this practice is common, there is very little research examining the content and quality of juvenile mental health evaluations and to what extent JJ professionals consider information in mental health evaluations informative and useful in developing disposition plans. To date, the limited research has focused on judges use of evaluations and neglected use of evaluations by juvenile probation officers (JPOs) despite JPO's significant role in developing and implementing disposition plans (Schwalbe and Maschi 2012; Wasserman et al. 2008). The purpose of this study was to address this gap by reviewing the content and structure of predisposition psychological evaluations using comparable coding schemes to past research conducted with judges and identifying the specific evaluation characteristics that impacted JPOs use of evaluation information for disposition planning. The following sections detail background information that set the stage for our study including a review of JJ screening and assessment standards and research concerning the use of evidence-based tools in evaluations with justice-involved youth. Additionally, the limited prior research addressing the content, structure, and usefulness of juvenile predisposition psychological evaluations is reviewed to support the rationale and aims of the current study.

Practice Standards for Screening and Assessment of Justice-Involved Youth

The following sections address two areas that are essential to the rationale for the current study. First, proposed recommendations for screening and assessment with juvenile justice-involved youth are presented and practice standards regarding data gathering and structural recommendations for predisposition evaluations are reviewed. Second, recommendations regarding the use of evidence-based tools are included to highlight the dearth of research evaluating the extent to which use of such instruments impacts the adoption of evaluator recommendations into disposition plans for youth.

Screening and Assessment Standards

A national group of experts in mental health screening and assessment convened in April of 2002 and derived six recommendations for JJ settings. The recommendations included the following: (a) provide an evidence-based, scientifically sound mental health screen within the first 24 hours of a youth's arrival at a facility; (b) provide an evidence–based, scientifically sound mental health screening and/or assessment for all youths as early as possible to determine need for mental health services; (c) conduct a comprehensive mental health assessment based on careful review of information from multiple sources that measure a range of mental health concerns; (d) provide an evidence-based and scientifically sound screening or assessment for all youths preparing to leave a post-adjudicatory secure facility and return to their communities; (e) provide evidence-based and scientifically sound screening/assessment on a regular basis for all youths; and, (f) ensure that mental health staff are professionally credentialed or are directly supervised by credentialed staff with regular training in evidence-based, scientifically sound mental health screening/assessment procedures (Wasserman et al. 2003). These recommendations represented a clear step in the direction toward attempting to standardize mental health assessment procedures for justice-involved youth and created basic practice standards for

different types of mental health assessments of youth at different points of contact with the JJ system.

Reflecting these recommendations, responding to a youth's mental, emotional, substance use, and behavioral problems requires accurate identification of these problems. Detecting potential mental health and substance use disorders requires reliable and valid screening and assessment tools, as well as information about how to best use and report the results of these tools in an evaluation (Grisso and Underwood 2004). More recently, Grisso (2013) outlined a structure for rehabilitation evaluations (including predisposition evaluations) that offered a framework by which evaluators and consumers of predisposition evaluations can measure the sufficiency of information included in evaluations and explanations that support evaluation recommendations. According to Grisso, a predisposition evaluation should be organized around four areas: the youth (i.e., what are the youth's important characteristics?), the objective (i.e., what needs to change?), the method (i.e., what modes of intervention could be applied toward the rehabilitation objective?), and the outcome (i.e., what is the likelihood of change, given the relevant interventions?). More specifically, when assessing the youth, Grisso recommended that the evaluator address the following areas: health and mental health history, family and social background, academic and intellectual functioning, personality description, clinical diagnostic description, delinquent behaviors and legal history, responses to past rehabilitation efforts, and risk of harm to others.

In order to provide sufficient recommendations, evaluators should provide opinions, with corresponding support, that explain the youth's delinquent behavior and provide a focus for the objectives of rehabilitative intervention. Additionally, evaluators should provide opinions and their support for a method (i.e., a placement, set of services, and rehabilitation process) that focuses on changing factors that are related to the youth's delinquency, while balancing the need for public safety. Lastly, evaluators should include their opinion and its support regarding the expected outcome of the interventions, including the likelihood that the recommended services will succeed in substantially reducing the risk of future offending (Grisso 2013).

Use of Empirically-Supported Tools

Valid and reliable assessment instruments are necessary for developing and deploying an effective screening and assessment process. When conducting psychological evaluations for the courts, evaluators want to pay particular attention to the psychometric properties of instruments used and whether the screening and assessment tools are evidence-based or empirically-supported (Grisso 2013; Grisso and Underwood 2004; Vincent et al. 2007). Assessment tools with evidence of reliability and validity among youth in JJ are preferable to those that do not. Instruments that provide norms according to gender, age, and ethnic background are preferable to those that do not (Grisso and Underwood 2004). While a great detail of attention has been paid to increasing professional knowledge about the use of evidence-based or empirically-supported tools, no research exists that evaluates the extent to which use of such instruments increases the likelihood of judges and/or JPOs incorporating assessment findings into juvenile disposition plans.

Content, Structure, and Impact of Predisposition Psychological Evaluations

In conducting predisposition evaluations for the court, evaluators communicate the findings of screening and assessment to attorneys, judges, and JPOs. Even if appropriate

content standards are followed (i.e., relevant domains addressed; use of empirically-supported assessment tools), the content, structure, and quality of the written evaluation can impact the usefulness of evaluation recommendations in guiding legal decision-making and case management.

To address some of the shortcomings in mental health assessments and report writing, Heilbrun (2001) developed a set of principles for conducting Forensic Mental Health Assessments (FMHA), that included principles such as 'identify the relevant forensic issues; obtain relevant historical information; use multiple sources of information for each area being assessed; and use plain language, avoid technical jargon.' Principles such as these have been used in research to assess the content, structure, and quality of forensic evaluations. Similarly, Grisso (2010) identified common errors that evaluators make in forensic report writing including 'failure to consider alternative hypotheses for findings; improper test uses; inadequate data; irrelevant data or opinions; and opinions without sufficient explanations.' These and other report miscommunications by evaluators may have an impact on the extent to which judges and JPOs find information contained in an evaluation to be useful and influence the extent to which the evaluator's recommendations are incorporated into a youth's disposition plan.

Although the FMHA principles, Grisso's common errors, and the screening and assessment standards reviewed above set up a framework by which content and structure of evaluations can be reviewed, limited empirical research has addressed how content and structure of predisposition evaluations informs the usefulness of these evaluations in developing disposition plans for youth. Lander and Heilbrun (2009) examined criminal forensic mental health assessments in adult competence to stand trial evaluations to determine adherence to FMHA principles. They found that evaluation content demonstrated limited consistency with the principles; however, there existed a significant relationship between the number of principles applied and expert judgments (i.e., law professor, judge, attorney, psychiatrist, psychologist) of the relevance, helpfulness, and quality of the evaluations, lending some support for the benefits of evaluator adherence to FMHA principles in writing evaluations for the courts.

Specific to JJ settings, Hecker and Steinberg (2002) evaluated 172 predisposition evaluations of justice-involved youth, coding the evaluations on key psychosocial variables as well as examining the usefulness of the evaluations in terms of the use of evaluator recommendations by judges. They reported that many of the evaluators' psychological evaluations lacked information about criminal, mental health, and drug/alcohol history. Results also indicated that judges were more likely to use recommendations from evaluations when explanations for these recommendations were present, when there was some indication as to *why* the recommendations were made, and some indication of the likelihood of a positive effect of the recommended course of action (Hecker and Steinberg 2002). This finding suggests recommendations that were explained and supported impacted the use of the information by legal professionals.

Campbell and Schmidt (2000) examined the influence of legal and mental health factors on judges' decision making and the agreement between evaluators' recommendations and judges' final decisions in juvenile cases. Mental health variables played a limited role in judge disposition decisions, while seriousness of the current offense was an important variable. Overall, agreement between evaluator recommendations and judge disposition decisions was 67.5 %, with greater agreement for legal recommendations than for mental health recommendations (Campbell and Schmidt 2000). O'Donnell and Lurigio (2008) examined 248 evaluations of justice-involved youth, exploring the extent to which psychosocial factors affected evaluators' placement recommendations and judges' sentencing

decisions. Researchers reported that evaluators' recommendations accounted for over 50 % of the variance in judges' recommendations and sentencing decisions for youth in their study.

The limited research in this area suggests some observations about the usefulness of predisposition mental health evaluations: (a) key mental health domains appeared to be missing or insufficiently represented in predisposition evaluations that may limit usefulness of the evaluations; (b) judges appeared to utilize evaluator recommendations regarding legal outcomes (e.g., placement) more often than recommendations regarding clinical outcomes (i.e., specific treatment referrals); (c) studies have not addressed whether content and/or structure influence the use of evaluators' recommendations by JPOs; and (d) no studies have reported how often empirically-supported tools (ESTs) are used and if greater use of ESTs increase the likelihood of judges or JPOs incorporating evaluation findings into juvenile disposition plans. These are important gaps to remedy as JPOs often play a central role in developing and implementing recommendations and case plans for post-adjudicated youth (Perrault et al. 2012; Vincent et al. 2012a, b). Additionally, research has not addressed the presence of systematic variation in predisposition evaluations by gender or race and whether these demographic characteristics are associated with adoption of specific evaluation recommendations.

Current Study

Utilizing coding strategies from prior research (Hecker & Steinberg, Lander and Heilbrun 2009) focusing on judges, the purpose of this study was to review the content and structure of predisposition psychological evaluations conducted in an East Coast State and identify the specific evaluation characteristics that impacted JPOs' use of the information contained in the evaluations for disposition planning. A series of exploratory hypotheses were generated based on the limited research with judges. Consistent with Hecker and Steinberg (2002) and Grisso (2013), we expected that the *content* of predisposition evaluations would include key background and psychosocial factors and use of empirically-supported tools (ESTs). Second, consistent with the work of Heilbrun (2001), we examined adherence to FMHA principles and expected evaluations to incorporate FMHA principles relevant to predisposition evaluations. Based on prior research identifying differences in mental health disorders, clinical decision-making during the assessment process, and use of treatment by gender and race (Dalton et al. 2009; Grisso and Barnum 2000; Teplin et al. 2002; Vincent, et al. 2008), we examined the possible impact of gender and race on evaluation content and structure. In terms of the *usefulness* of predisposition evaluations, we hypothesized that variation in content, structure, and relevance in recommendations would predict greater use of evaluation recommendations by JPOs.

Method

Data Source

A total of 150 archived predisposition psychological evaluations were randomly selected from a larger pool of 300 evaluations conducted within a three-year period (2009–2012) across 12 juvenile court jurisdictions within an East Coast State. For the purposes of this study, a *predisposition evaluation* was defined as a mental health evaluation conducted by an evaluator for the court which usually included information about a variety of important

domains in a juvenile's life including family history, mental health history, cognitive functioning, personality functioning, substance use history, criminal history, etc. It is common for evaluators to include a number of recommendations at the conclusion of the evaluation depending on the specific referral questions including recommendations for treatment services (i.e., individual therapy, family therapy), need for placement, level of monitoring, and further evaluation (i.e., specialized substance use assessment). Neither statutory language nor case law specifically addresses the content or structure of predisposition assessments in the jurisdiction where the study was conducted. The state agency with administrative oversight over the court-based assessment process has developed and disseminated information based on best practices recommendations for screening and assessment (e.g., Wasserman et al. 2003) and report writing guidelines (Grisso 2013) to evaluators and conducts focused quality assurance reviews of assessments. However, a systematic review of a large number of evaluations conducted over a three-year time period has not been carried out in the jurisdiction.

Evaluators who conducted the 150 predisposition psychological evaluations were mostly male (91 %) and their academic degrees ranged from Ph.D. (55 %) to Psy.D. (42 %) to Other (i.e., M. A., M. Ed., etc.) (3 %). Regarding the youth being evaluated, 81 % were male with a mean age of 15.2 ($SD = 1.4$). The youth were ethnically diverse with 37.3 % African American, 33.3 % Caucasian, 24 % Latino, 0.7 % Native American, 1.3 % Pacific Islander, and 3.3 % race not given in the record. Approximately 89 % of youth were enrolled in school at the time of the evaluation. Overall, 74 % of youth had prior arrests with the mean age of first arrest being 13.4 ($SD = 1.9$) and 43 % had prior convictions with the mean age of first conviction being 14.6 ($SD = 1.5$). In terms of current charges, 38 % of youth presented with one, 24 % with two, 13 % with three, 12 % with four, 3 % with five, and 10 % with more than five current charges.

Codebook

A codebook was established for the current study based in part on prior evaluation coding schemes reported in the literature (Campbell and Schmidt 2000; Hecker and Steinberg 2002; Heilbrun 2001). The following domains were included in the codebook: legal history/current offense, sociocultural factors, empirically-supported tools (ESTs), FMHA principles (Heilbrun 2001), evaluator recommendations, and disposition/recommendation match. Each domain was operationalized and coded as follows.

Legal History/Current Offense

Evaluations were coded for the absence (0) or presence (1 or 2) of information regarding prior legal history (i.e., is the prior legal history of the juvenile given/stated in the evaluation?) and the current offense (i.e., is information on current charges given/stated in the evaluation?). Whether or not these variables received a score or "1" or "2" depended on the level of detail (e.g., sufficiency) included by the evaluator. More detailed accounts were coded "2." Each factor had coding instructions involving the necessary information to rate sufficiency. For example, information on prior legal history was coded "0" if no history was given in the evaluation, "1" if the evaluator only reported the number of prior offenses but no other legal history information, and "2" if the evaluator reported prior offenses and information regarding seriousness of offenses. Information on current offenses was coded in a similar manner.

Sociocultural Factors

Similar to past research by Hecker and Steinberg (2002), a specific coding scheme was used to assess the presence and sufficiency of detail regarding relevant sociocultural factors including family history, education, mental health history, substance use history, criminal history, cognitive functioning, personality functioning, and behavioral observations. Although criminal history overlapped with the legal history/current offense domain, all factors and coding schemes used by Hecker and Steinberg (2002) were adopted for comparison purposes. Each factor was coded as absent (0) or present (1, 2, or 3) following the same coding rationale as Hecker and Steinberg (2002). Once presence was established, evaluations were coded for sufficiency based on the amount of detail provided for each factor. For example, when coding mental health history, whether or not the evaluator included information concerning prior mental health history or prior mental health services was rated as a "1," this information plus some information about the duration and nature of the difficulty as well as the course of treatment was rated as a "2," and lastly, all of this information plus an indication of the institution or service agency the juvenile was involved with was rated a "3".

Empirically-Supported Tools

Tool(s) used by the evaluator in conducting the evaluation were rated as a dichotomous variable (present/absent) followed by coding the purpose of the tool as described in the evaluation (i.e., cognitive, personality, risk assessment, mental health, achievement, neuropsychological, etc.). Each tool was then coded for level of empirical support ("0" not empirically-supported, "1" promising, "2" empirically-supported) based on recommendations of the National Youth Screening and Assessment Project (NYSAP) and whether the tool was used properly as defined by researchers ("0" not used properly, "1" used somewhat properly, "2" used properly). Three trained graduate research assistants researched each tool used by the evaluators and rated the level of empirical support according to the NYSAP criteria. The criteria for these ratings are referenced in Table 3.

FMHA Principles

Reflecting the work of Lander and Heilbrun (2009), this domain included rating 17 of the 29 Forensic Mental Health Assessment (FMHA) principles (Heilbrun 2001) applicable to juvenile predisposition evaluations. For the current study, researchers eliminated 12 principles that either did not apply (i.e., testify effectively) or were inapplicable to juvenile disposition proceedings. The researchers developed coding criteria for each principle that reflected consistency with juvenile predisposition evaluations. Each principle was coded as absent (0) or present (1).

Evaluator Recommendations

Each recommendation was coded as absent (0) or present (1, 2, 3, or 4) based on the extent to which the evaluator synthesized information about a youth's functioning across domains to present a clear and logical explanation for disposition recommendations. Each recommendation was coded for sufficiency of the explanation or justification given. Whether the explanation for the recommendation presented was mildly sufficient ("2"), sufficient ("3"), or better than sufficient ("4") was based on the level of depth and indication of

appropriateness of the recommendation for the particular juvenile. For example, if the evaluator included a recommendation without explanation it was rated as a "1"; if the evaluator included an explanation for the specific recommendation AND/OR an indication as to why the specific placement/service recommended was appropriate for the juvenile, it was scored as a "2"; if the evaluator also included an indication of the likelihood that the recommendation would have a positive effect in light of criminal behavior and refrained from using psychological jargon, it was coded a "3"; and finally, all of this information plus a greater amount of detail concerning the likelihood of a positive effect of the recommendation for the juvenile was rated a "4" for sufficiency. Coding of "0" was reserved for an evaluation without any recommendations.

Disposition/Recommendation Match

Researchers were given access to the juvenile's initial disposition plan that detailed the specific conditions that the juvenile was to follow while on probation. As the key dependent variable in the study, researchers were interested in whether or not evaluator recommendations were incorporated into the initial disposition plan. The extent to which each element of the initial disposition plan matched evaluators' recommendations was rated categorically (0, 1, or 2). Each recommendation was coded as full rejection or no match ("0") if the disposition plan did not include the evaluator's recommendation, a partial match ("1") if the disposition plan was a consistent but scaled back or otherwise altered recommendation (e.g., evaluator recommended MST but the disposition only indicated individual therapy), and as complete match ("2") between disposition plan and recommendation. For example, if the evaluator recommended that the juvenile begin Multisystemic Therapy (MST) but the disposition did not include MST, this recommendation was coded "0". If the disposition included attend therapy, it would be coded a "1" as it represents an altered treatment recommendation. Lastly, if the disposition included that the juvenile should engage in MST, the recommendation was coded a "2".

Procedures

This study was approved by the appropriate ethics committee at Fordham University. After obtaining IRB approval, researchers were provided with a random selection of predisposition psychological evaluations conducted between 2009 and 2012 where there was clear indication via court records that the JPO reviewed the predisposition evaluation as part of the disposition process. Staff on-site identified evaluations using existing billing records of court-based assessments conducted within the three-year time period. All evaluation copies were de-identified and labeled with a research identification number before being provided to the researchers. All data were de-identified and details that might disclose the identity of youth were omitted. As the study involved use of de-identified archival records, informed consent was waived. Access to the de-identified reports were provided to the authors following approval of the study design and procedures by the administrative agency that oversees the court-based assessment process. The study was not funded by the agency and analyses were conducted independent of the agency to ensure no conflict of interest. For each evaluation, the legal history, current offense information, demographics, and probation records delineating the youth's initial disposition plan were collected from an electronic statewide probation database, extracted, and identified by the same research identification number before being provided to the researchers. Once the evaluations were received, one of three trained graduate research assistants coded each de-identified

evaluation. Training entailed each graduate assistant first coding a set of five training evaluations to demonstrate competence with the codebook and clarify any coding inconsistencies against a consensus coding. A subset of 10 evaluations was independently coded by the three graduate research assistants to assess inter-rater reliability for all coded variables. An acceptable level of agreement was established across domains (Mean ICC = 0.87, range from 0.75 to 0.95).

Analyses

The first author of this paper takes responsibility for the integrity of the data and the accuracy of the data analyses. Several variables were developed for the data analyses including aggregate and ratio scores of the domains representing *content* (i.e., legal history/current offense, sociocultural factors, empirically-supported tools) and *structure* (i.e., FMHA principles) and ratio scores of the domains representing *usefulness* (i.e., recommendation sufficiency ratio score, recommendation acceptance ratio score).

Aggregate Scores

Three separate aggregate scores were calculated. First, an aggregate score reflecting the presence and sufficiency of the Legal History/Current Offense was formed by summing the two items in this domain (possible range 0–4). Second, an aggregate score reflecting the presence and sufficiency of Sociocultural Factors was formed by summing the final rating on the eight factors rated in this domain (possible range of 0–24). Third, an aggregate score reflecting the presence of FMHA Principles was created by summing the number of principles coded as present (possible range 0–17).

Ratio Scores

Three separate ratio scores were calculated. First, a ratio score reflecting how often Empirically-Supported Tools were used by the evaluators (EST ratio score) was formed by summing the number of tools that were scored a "2" on empirical support divided by the total number of tools used by the evaluator (possible range of 0.00–1.00). Second, a ratio score reflecting the presence and sufficiency of Evaluator Recommendations (Recommendation sufficiency ratio score) was formed by summing each recommendation sufficiency score divided by the total number of recommendations (possible range of 0.00–1.00). Third, a ratio score reflecting the Acceptance of Evaluator Recommendations relative to the number of disposition requirements listed in the disposition case planning notes (Recommendation acceptance ratio score) was formed by summing the number of recommendations adopted (complete and partial match) into the disposition plan divided by the total number of recommendations (possible range of 0.00–1.00).

Descriptive Analyses

Based on the coding scheme detailed above, the first level of analyses for each domain involved calculating descriptive statistics, including frequency/percentages, for items within each domain (i.e., frequency and percentage of each FMHA principle identified across evaluations). The second level of analyses examined item level differences by gender and race via Chi square analyses.

Regression Analyses

For the third level of analyses, aggregate and ratio scores reflecting each evaluation domain (legal history/current offense and sociocultural factors aggregate scores, EST ratio score, FMHA principles aggregate score, recommendation sufficiency ratio score) were used as unique predictors in a series of multiple regressions to predict the recommendation acceptance ratio score.

Results

Content of Evaluations

Frequency of Content Domains and Sufficiency Ratings

Content, as well as the sufficiency, of reported detail across key evaluation domains (legal history, sociocultural factors, use of empirically-supported tools) was examined. The mean aggregate score for presence and sufficiency of legal history/current offense was 2.17 ($SD = 1.1$), with a range of 0–4. As a reminder, these items were scored 0 when not present, 1 when present with insufficient amount of detail, and 2 when present with sufficient amount of detail. Therefore, higher numbers represent higher level of sufficiency. Information on prior legal history of the youth was present in only 48.0 % of evaluations, with only 24.0 % including a sufficient amount of detail (e.g., 0, legal history not present; 1, legal history includes only number of prior offenses but no other information, such as "youth has three prior offenses"; 2, legal history includes prior offenses and information regarding seriousness of offenses, such as "youth has two prior convictions for assault charges and one prior conviction for robbery"). In the 76 % of evaluations with insufficient information, the evaluation reported the number of prior offenses but no other legal history information. Information on current charges was present in 90.7 % of evaluations, with 54.7 % including a sufficient amount of detail. In the 45.3 % of evaluations with insufficient information, the evaluator only reported the number of current charges with no other information about the charge(s). See Table 1 for further break down of sufficiency scores for each factor.

The mean aggregate score for presence and sufficiency of sociocultural factors was 12.62 ($SD = 2.7$), with a range of 7–21, out of a possible range of 0–24 (see Table 2). As a reminder, these items were scored 0 when not present, 1 when present with insufficient

Table 1 Frequency, percentage, and aggregate score of legal domain: presence and sufficiency ratings

Item	Present in evaluation		Sufficiency rating[a]		
	Yes *n* (%)	No *n* (%)	0 *n* (%)	1 *n* (%)	2 *n* (%)
Legal history	74 (48.0)	78 (52.0)	78 (52.0)	36 (24.0)	36 (24.0)
Current offense	136 (90.7)	14 (9.3)	14 (9.3)	54 (36.0)	82 (54.7)

[a] Sufficiency Ratings are a numerical score representing the amount of detail documented for each factor in evaluation; 0 = not present, 1 = present, not sufficient, 2 = present, sufficient

Table 2 Frequency, percentage, and aggregate score of sociocultural factors: presence and sufficiency ratings

Item	Present		Sufficiency rating[a]			
	Yes *n* (%)	No *n* (%)	0 *n* (%)	1 *n* (%)	2 *n* (%)	3 *n* (%)
Family history	150 (100.0)	0 (0.0)	0 (0.0)	94 (62.7)	4 (2.7)	52 (34.7)
Educational history	149 (99.3)	1 (0.7)	1 (0.7)	51 (34.0)	9 (6.0)	89 (59.3)
Criminal history	75 (50.0)	75 (50.0)	75 (50.0)	28 (18.7)	32 (21.3)	15 (10.0)
Mental health history	142 (94.7)	8 (5.3)	8 (5.3)	52 (34.7)	51 (34.0)	39 (26.0)
Drug/alcohol history	139 (92.7)	11 (7.3)	11 (7.3)	25 (16.7)	100 (66.7)	14 (9.3)
Cognitive functioning	146 (97.3)	4 (2.7)	4 (2.7)	131 (87.3)	8 (5.3)	7 (4.7)
Personality functioning	141 (94.0)	9 (6.0)	9 (6.0)	119 (79.3)	14 (9.3)	8 (5.3)
Behavioral observations	148 (98.7)	2 (1.3)	2 (1.3)	55 (36.7)	49 (32.7)	44 (29.3)

[a] Sufficiency Ratings are a numerical score for the amount of detail included for each factor; 0 = not present, 1 = present, not sufficient, 2 = present, sufficient, 3 = present, better than sufficient. When reported in the results section, we reported a combined percentage for ratings of 2 and 3 (sufficient or better)

amount of detail, 2 when present with sufficient amount of detail, and 3 when present with better than sufficient amount of detail. Therefore, higher numbers represent higher level of sufficiency. In terms of each sociocultural factor: family history was present in 100 % of evaluations; however, this family history was rated as sufficient or better in detail only 37.4 % of the time. In the 62.6 % of evaluations with insufficient detail, family history was rated as lacking information concerning parental relationship, involvement with social services agencies, and family history of substance use/mental health issues. Educational history was present in 99.3 % of evaluations, with 65.3 % rated as sufficient or better. Criminal history was present in 48.0 % of evaluations, with 31.3 % rated as sufficient or better. In the 68.7 % of evaluations with insufficient detail, evaluations lacked information concerning nature of prior offenses and information indicating disposition or placement in previous cases. Mental health history was present in 94.7 % of evaluations, with 60.0 % rated as sufficient or better. Drug and alcohol history was present in 92.7 % of evaluations, with 76.0 % of information being sufficient or better. Information on cognitive functioning was present in 97.3 % of evaluations; however, only 10.0 % of information presented was rated sufficient or better. In the 90 % of evaluations with insufficient detail, the evaluator either did not include any information concerning cognitive functioning of the youth (rated "0") or they neglected to reference relevant scores from IQ and achievement tests *when administered in the evaluation*, failed to document intellectual strengths and weaknesses of the youth, did not integrate relevant information about the youth's cognitive functioning into summary/recommendations, or the evaluator used psychological jargon when discussing cognitive functioning (these would result in a rating of "1"). Information on personality functioning was present in 94.7 % of evaluations; however, only 14.6 % of information presented was rated as sufficient or better. In the 85.4 % of evaluations with insufficient detail on this factor, evaluators either did not clearly identify how the youth's psychological/personality functioning was or was not related to current offending behavior or the evaluator used psychological jargon when discussing personality functioning. Lastly,

behavioral observations were present in 98.7 % of evaluations, with 62.0 % of information being sufficient or better (see Table 2).

Frequency of ESTs Used in Evaluations

The mean number of assessment tools used by evaluators in the predisposition evaluations was 5.75 ($SD = 1.51$), with a range of 0–8. Breaking this down further into 'not empirically-supported,' 'promising,' and 'empirically-supported' tools, evaluators used an average of 2.02 ($SD = 1.22$; range 0–8) 'not empirically-supported', 0.49 ($SD = 0.54$; range 0–2) 'promising,' and 3.24 ($SD = 1.10$; range 0–6) 'empirically-supported' tools. The most frequently used ESTs were the Bender Gestalt ($n = 129$) and the Wechsler Intelligence Scales for Children- IV (WISC-IV; $n = 124$). These tests represent tools from the following categories: cognitive assessment and neuropsychological assessment.

Table 3 Top 10 assessment tools used in evaluations and level of empirical support

Tool	Type of tool	*n*	%	Empirical support[a]
1. Bender Gestalt	Neuropsychological	129	86.0	2
2. WISC-IV	Cognitive	124	82.7	2
3. Thematic Apperception Test	Personality	84	56.0	0
4. Personality Youth Inventory	Personality	46	30.7	2
5. Sentence Completion	Personality	46	30.7	0
6. WRAT-3/WRAT-4	Achievement	44	29.3	2
7. Symptom Assessment-45	Personality	42	28.0	1
8. Draw a Person Test	Personality	30	20.0	0
9. Child Behavior Checklist	Personality	27	18.0	2
10. Vanderbilt ADHD Scale	ADHD Symptom Specific	21	14.0	1

Level of Empirical Support, 0 = not empirically-supported; 1 = promising; 2 = empirically-supported

[a] Criteria for Empirical Support: A tool is considered an *Empirically-supported instrument* if ALL of the following are met: (1) A *manual:* A tool should have some version of a test manual that contains scoring criteria and/or detailed item descriptions to structure the administration. (2) *Contains empirically-based risk factors*: A tool should contain youth risk factors that have been empirically demonstrated to have an association with future crime and violence. (3) *The tool has demonstrated reliability in multiple studies, some of which were conducted by independent parties:* Assessment instruments should have some reported evidence for inter-examiner reliability. If the tool is self-report only (in other words, the tool does not rely on examiner ratings), then the interest is in internal consistency and test–retest reliability. However, for tools that do rely on examiner ratings, evidence for inter-rater reliability is critical to provide confidence that the tool will be completed fairly consistent across examiners. The preferred measure of reliability in this case is intra-class correlation coefficients (ICCs). ICCs should be above 0.70 at the minimum and preferably above 0.90. An instrument should have several encouraging tests of inter-rater reliability that were conducted in a JJ setting by an independent party (meaning someone other than the test developer). In addition, ideally at least one of the studies will have been conducted "in the field"—that is, in a real-world setting involving the same types of raters, the same facilities, and the same youth with whom the tool would normally be applied. (4) *The tool has demonstrated predictive validity in multiple studies, some of which were conducted by independent parties*: An assessment tool must have evidence that it predicts other measures of the same construct that are measured sometime in the past or future. When evaluating a tool, it is important to be familiar with this research, including the actual outcomes tested and the methods used (e.g., prospective versus retrospective studies). A tool is *considered a promising* instrument if at least 2 of the above criteria 1 thru 4 are met. A tool is *not considered empirically-supported* instrument if less than 2 of the above criteria are met

Table 3 contains the 10 tools most frequently used in the evaluations as well as the empirical support for each tool. The mean ratio for use of ESTs by evaluators was 0.58 ($SD = 0.19$), with a range of 0.00–1.00. The EST ratio score represents the number of empirically-supported tools used in an evaluation divided by the total number of any tools used in an evaluation. This ratio reflects that, on average, 58 % of the tools used by evaluators were considered ESTs.

Gender and Race Differences

Chi square analyses were conducted to assess any gender differences in the presence/absence of each factor represented in the content domains (legal history/current offense, sociocultural factors). No significant differences were observed at the item level. Independent sample t-tests were conducted to investigate gender differences in the sufficiency scores for each content item within the legal history/current offense domain (i.e., legal history) and sociocultural factors domain (i.e., family history, educational history, etc.). No significant differences were observed by gender. Independent t-tests were conducted to assess any gender differences in total aggregate scores for the two content domains (legal history/current offense, sociocultural factors) and no significant differences emerged. Lastly, an independent t test was conducted to assess any gender differences in the EST ratio score and again, no significant differences emerged.

Race categories were collapsed into four categories to ensure adequate cell size (i.e., African American, Caucasian, Latino/a, Other) and Chi square analyses were conducted to assess any race differences in the presence/absence of each factor represented in the content domains (legal history/current offense, sociocultural factors). No significant differences were observed at the item level. A one-way Analysis of Variance (ANOVA) was conducted to assess any race differences in item-level sufficiency scores. One area of significant difference emerged for the Cognitive Functioning sufficiency score with the Other race group ($M = 1.62$, $SD = 0.92$) having a slightly higher mean sufficiency score than the Caucasian race group ($M = 1.02$, $SD = 0.32$), $F(3, 146) = 3.523$, $p = 0.02$. A one-way ANOVA was conducted to assess any race differences in total aggregate scores for the two content domains (legal history/current offense, sociocultural factors) with no significant differences observed by race. Finally, a one-way ANOVA was conducted to assess any race differences in the EST ratio score and again, no differences were observed. These results indicate that the level of detail reported across these content domains varied little by gender or race of the youth being evaluated.

Structure of Evaluations

Frequency of Adherence to FMHA Principles

Structure of the evaluations was investigated based on adherence to 17 of the original 29 FMHA principles. Consistent with prior research (Lander and Heilbrun 2009), principles were coded as absent or present. The frequency of adherence to FMHA principles ranged from 18 to 99.3 %. The most frequently present principle was "writes evaluations in sections" and the least frequent present principle was "avoids technical jargon" (see Table 4). Examples of what was coded as "jargon" included: overly convoluted/complex interpretative statements when discussing the results of personality or cognitive testing, statements that seem as if they were copy and pasted from interpretative manuals when presenting results of testing, use of statistical language without explanation (e.g., standard

Table 4 Frequency, percentage, and aggregate score of FMHA principles used in evaluations

FMHA principle	Present		Absent	
	n	%	*n*	%
Identifies the relevant juvenile forensic issue	141	94.0	9	6.0
Obtains appropriate authorization	143	95.3	7	4.7
Uses multiple sources of information	122	81.3	28	18.7
Obtains relevant historical information	139	92.7	11	7.3
Assesses clinical characteristics	94	62.7	56	37.3
Assesses legally relevant behavior	124	82.7	26	17.3
Provides notification of purpose and consent	95	63.3	55	36.7
Understands purpose and confidentiality limits	99	66.0	51	34.0
Uses third party information to assess response style	106	70.7	44	29.3
Uses testing when indicated to assess response style	37	24.7	113	75.3
Uses idiographic evidence in assessing clinical conditions	140	93.3	10	6.7
Uses nomothetic evidence in assessing clinical conditions	92	61.3	58	38.7
Uses scientific reasoning in assessing causal connections	129	86.0	21	14.0
Does not answer the ultimate legal question	117	78.0	33	22.0
Attributes information to sources	136	90.7	14	9.3
Avoids jargon	27	18.0	123	82.0
Writes report in appropriate sections	149	99.3	1	0.7

deviation, standard error), etc. The number of principles coded as present in each evaluation was summed to form a FMHA principles aggregate score. On average, evaluations reflected 12.6 ($SD = 3.1$) principles, with a range of 4–17. This score reflects adequate coverage of FMHA principles by evaluators in the predisposition evaluations.

Gender and Race Differences

Chi square analyses were conducted at the item level to investigate frequency differences in presence/absence of FMHA principles by gender. An independent t-test was conducted to assess any gender difference in the Adherence to FMHA principles aggregate score. Similar to the content domains, no gender differences emerged indicating that FMHA principles were similarly represented in evaluations of male and female youth.

Chi square analyses were conducted to investigate frequency differences in presence/absence of FMHA principles by race with no significant differences emerging. A one-way ANOVA was conducted to assess race difference in the FMHA principles aggregate score. Once again, no significant differences emerged by race indicating that FMHA principles were similarly represented in evaluations across the four race groups.

Usefulness of Predisposition Psychological Evaluations

Usefulness of the predisposition psychological evaluations was examined through a series of steps. First, we assessed how often and what type of evaluator recommendations were incorporated by JPOs into the initial disposition plans for youth. Second, we assessed if

specific evaluation characteristics (content, structure, relevance of recommendations) were associated with and predicted greater use of evaluation recommendations by JPOs.

Evaluator Recommendations

The mean number of recommendations included in the evaluations was 3.69 ($SD = 1.9$), with a range from 0 to 8. The mean recommendation sufficiency ratio for evaluator recommendations was 0.85 ($SD = 0.24$) with a range of 0.00–1.00. This ratio represents the number of recommendations with a sufficient or better explanation divided by the total number of recommendations made by the evaluator. This ratio reflects that, on average, 85 % of evaluator recommendations included sufficient detail supporting or justifying the recommendations.

The mean acceptance ratio of evaluator recommendations by JPOs was 0.35 ($SD = 0.34$) for complete and partial match, with a range of 0.00–1.00. This ratio represents the number of evaluator recommendations accepted and documented into the youth's initial disposition plan divided by the total number of recommendations made by the evaluator. This ratio reflects that, on average, 35 % of recommendations made by an evaluator were present in the youth's initial disposition plan. Of those recommendations that were accepted, 70 % were mental health recommendations (e.g., individual and family counseling, sexual offender specific treatment, psychiatric consultation), 20 % were legal recommendations (e.g., drug screening, close monitoring), 6 % were placement recommendations, 2 % were educational recommendations, and 2 % were constituted as 'Other.'

Gender and Race Differences

Independent t-tests were conducted to assess any gender differences in the number of total recommendations made by evaluators and to investigate any gender differences in the mean recommendation sufficiency and recommendation acceptance ratio scores. Similar to above, no significant differences emerged by gender on these variables. Similarly, a one-way ANOVA was conducted to investigate race differences in the recommendation variables. Similar to gender, no significant differences emerged by race. These results reflected that the total number of recommendations offered by evaluators and accepted by JPOs did not differ based on the gender or race of youth being evaluated.

Disposition/Recommendation Match

Finally, we were interested in what specific evaluation characteristics were associated with and predicted greater use of evaluation recommendations by JPOs in developing the initial disposition plans. To answer this question, a multiple regression analysis was conducted to predict the recommendation acceptance ratio score from the following variables: legal history/current offense aggregate score, sociocultural factors aggregate score, FMHA principles aggregate score, EST ratio score, and recommendation sufficiency ratio score. Regression results indicated that none of the evaluation variables significantly predicted the recommendation acceptance ratio score, $F(5,142) = 0.991$, $p = 0.43$, $R^2 = 0.03$.

Additionally, researchers were interested in the degree to which individual sociocultural factors predicted the acceptance or uptake of evaluation recommendations. A multiple regression analysis was conducted to predict the recommendation acceptance ratio score from individual sociocultural factor ratings. Regression results indicated that none of the

sociocultural factors scores significantly predicted the recommendation acceptance ratio score, $F(8, 140) = 1.055$, $p = 0.40$, $R^2 = 0.06$. Collectively, these results indicated that none of the variation in legal history/current offense, sociocultural factors, use of ESTs, or sufficient support for evaluator recommendations predicted the use of evaluator recommendations by JPOs in developing initial disposition plans.

Discussion

The broad aims of this study were to review the content and structure of juvenile predisposition psychological evaluations (i.e., legal history, sociocultural factors, presence of FMHA principles, use of empirically-supported tools, justified recommendations) conducted in an East Coast state and identify the specific evaluation characteristics that impacted juvenile probation officers' (JPOs) use of evaluator recommendations in initial disposition plans for youth. Although a majority of evaluations included information from multiple key content areas, there was significant variability in the amount of information included for each key component. Past research has demonstrated gender and race differences with regards to the presence of mental health symptoms and psychopathology, as well as mental health service utilization (Grisso and Barnum 2000; Teplin et al. 2002; Vincent et al. 2008); however, the current study did not identify differences by gender and race in terms of the content and structure of evaluations. Additionally, gender and race differences were not reflected in the number of recommendations made by evaluators, the sufficiency of the explanations for recommendations given by evaluators, or the rate at which JPOs accepted evaluator's recommendations into youth's initial disposition plans. Evaluators generally adhered to a large number of FMHA principles and used several empirically-supported tools in their evaluations of youth. However, none of the evaluation variables of interest in the study that predicted judges use of evaluation recommendations in prior research by Hecker and Steinberg (2002), Campbell and Schmidt (2000), and Lander and Heilbrun (2009) (e.g., presence and sufficiency of detail of sociocultural factors, well-justified and explained recommendations, adherence to FMHA principles) predicted use of evaluation recommendations by JPOs in the current study.

With regard to the content of evaluations, a majority of evaluations contained information from several key content areas with presence ranging from 48 % (criminal history) to 100 % (family history). This is consistent with Grisso's (2013) recommendation that evaluators address multiple areas including mental health history, family and social background, intellectual functioning, personality, clinical diagnosis, legal history, treatment response, and risk of harm to others. Although the full range of these factors was rated as present across the evaluations, the rate of sufficient or better information included to support these content areas was highly variable, ranging from 10 % (cognitive functioning) to 76 % (drug and alcohol history). This variability is similar to findings by Hecker and Steinberg (2002) where quality of content areas rated as sufficient or better ranged from 10 to 63 %. This calls into question the depth of the information reported in the evaluations about key background, family, and psychosocial factors that may affect the usefulness of evaluations in informing disposition plans. Hecker and Steinberg (2002) also found that juvenile evaluations were lacking information in key content areas linked to recidivism such as criminal history, mental health history, and drug and alcohol history. Criminal history was referenced in only 48 % of the evaluations in the current study; however, mental health and drug and alcohol history were both highly present (94.7 and 92.7 % respectively). This finding may reflect an increased emphasis on the relevance of

these factors in recent years (see Grisso et al. 2005; Wasserman et al. 2003) or reflect evaluators' perspective in this jurisdiction that documenting prior mental health and substance use history are key components of a juvenile predisposition evaluation.

Little prior research exists regarding the use of empirically-supported tools (ESTs) in juvenile predisposition evaluations and whether or not the use of such tools influences JPOs' decision-making in developing disposition plans. The current study was the first to examine the use of ESTs by evaluators when conducting juvenile predisposition evaluations. On average, evaluators used between five and six instruments when conducting predisposition evaluations with approximately three of these instruments being rated as empirically-supported tools. The most frequently used ESTs were the Bender Gestalt (Bender 1938; Brannigan and Decker 2003), a test of neuropsychological functioning, and the Wechsler Intelligence Scale for Children-IV (WISC-IV; Wechsler 2003), a test of intellectual functioning. Other ESTs used in the evaluations included the Personality Inventory for Youth (PIY; Lachar and Gruber 1995), the Child Behavior Checklist (CBCL; Achenbach and Rescorla 2001), and the Wide Range Achievement Test (WRAT-4; Wilkinson and Robertson 2006). Overall, 58 % of the tools used by evaluators in their evaluations were considered ESTs. The remaining tools used were either considered 'promising' (e.g., Symptom Assessment-45; Strategic Advantages, Inc. 1998) or 'not empirically-supported' (e.g., Thematic Apperception Test, TAT; Murray 1943). These results imply that evaluators are making an effort to use tools supported by research in their evaluations, which follows recent efforts encouraging evaluators to use tools supported by research specific to JJ and/or with adolescents in their evaluations. Although this may be the case, the use of ESTs by evaluators did not influence JPO decision making in regards to disposition planning. While the use of ESTs was encouraging, the EST ratio was skewed by the common use of the Bender Gestalt and WISC-IV (e.g., used in over 80 % of all evaluations). Review of this result with the administrative agency overseeing the court-based assessments confirmed that evaluators were requested to screen youth for visual-motor processing and intellectual deficits in order to potentially identify unidentified learning problems and as appropriate, refer youth for special education evaluation/services. Evaluators appeared to have complied with this mandate and included these tests in evaluations. While the evaluators were appropriately responding to an agency-specific mandate, recommendations for educational services were found in only a small number of reports. JPOs may not have clearly understood why these tests were commonly reported or viewed intellectual deficits as extending beyond their role in addressing behavioral needs and supervising youth in the community. It is also possible that information gleaned from assessments was not being explained in a way that is easily understood by evaluation consumers. As discussed below, evaluators had a tendency to use psychological jargon in the evaluations and although ESTs were being used, the language used to communicate the results may have made it difficult to discern what information was important or relevant. As such, rationale for test selection and reporting of results may have limited the impact of the EST ratio on predicting JPOs use of evaluation recommendations.

With regard to the structure of evaluations, researchers coded 17 of the original 29 FMHA principles and found, on average, 12 principles were reflected in evaluations. Similar to past research using the FMHA principles by Lander and Heilbrun (2009), the frequency of adherence to FMHA principles ranged from low to high with the current study finding adherence ranging from 18 to 99.3 % compared to Lander and Heilbrun's range from 3.2 to 92.8 %. Interestingly, in the current study the least common principle adhered to was "avoid technical jargon"; however, Lander and Heilbrun (2009) found this principle to be present in 92 % of the evaluations in their sample. Additionally, in the

present study, the principle identified most often was "write report in sections;" however, in Lander and Heilbrun (2009), this principle was only present in 12 % of evaluations. This may reflect a difference in study methodology as Lander and Heilbrun (2009) evaluated the content of adult criminal mental health assessment reports, while the current study assessed juvenile predisposition evaluations. Utilizing 70 % as a benchmark for consistent use of the FMHA principles, the vast majority of evaluations reflected a thorough and comprehensive approach (i.e., obtains/documents consent/assent, gathers history, uses multiple sources of information, etc.), which represents good practice (Grisso 2013; Lander and Heilbrun 2009); however, many of the evaluations were also identified to contain technical jargon. This may reduce the ease at which JPOs can incorporate and utilize information from the evaluations when making decisions regarding disposition planning.

According to Grisso (2013), evaluators should include recommendations in their evaluations that detail what interventions could be applied toward established rehabilitative objectives. In order to provide sufficient recommendations, evaluators should provide support for their opinions that justifies and explains why a certain service is appropriate for the youth and the likelihood that the intervention will have a positive effect in reducing the youth's risk of future delinquency. In the current study, evaluators offered an average of four recommendations in their evaluations and provided sufficient explanations for these recommendations 85 % of the time. Although this was the case, only 35 % of recommendations were incorporated by JPOs into disposition plans. This is inconsistent with past research in which judges were more likely to use evaluators' recommendations when sufficient explanations for these recommendations were present and there was some indication as to the likelihood of a positive effect from the recommended course of action (Hecker and Steinberg 2002). Similarly, Campbell and Schmidt (2000) found an agreement rate of 67.5 % between evaluator recommendations and judges' disposition decisions in juvenile cases, while O'Donnell and Lurigio (2008) reported that evaluator recommendations accounted for 50 % of the variance in judges' recommendations and sentencing decisions for youth. One reason for the discrepancy between the current results and past research may be that the current study was the first to examine the agreement between evaluator recommendations and JPO decision-making. JPOs may take different factors into account than judges when considering evaluator recommendations such as the availability of recommended services or perceived feasibility of securing the recommended service for the youth.

Of those recommendations that were used by JPOs and incorporated into youth disposition plans, a majority were mental health recommendations such as individual or family counseling, sexual offender specific treatment, and psychiatric consultation, among others. This was followed by a smaller percentage of legal recommendations such as drug screening or close supervision and placement recommendations. A small percentage of accepted recommendations were educational. This contrasts with prior research in which mental health variables played a limited role in judge disposition decisions, while legal factors, such as seriousness of the offense, was an important variable (Campbell and Schmidt 2000). Additionally, past research demonstrated that judges were more likely to agree with evaluator recommendations that were legal in nature rather than mental health (Campbell and Schmidt 2000). One possible reason for this inconsistency may be that in the current study, youth were referred to evaluators for predisposition psychological evaluations based on prior mental health histories or when either the judge or attorneys had specific concerns about mental health functioning. Results of the evaluations were available to the attorneys and likely informed the eventual disposition and court recommendations that were then provided to probation staff. As such, the evaluations tended to focus

more on mental health variables and recommendations rather than legal recommendations which would also explain the greater use of mental health recommendations by JPOs.

Although past research demonstrated that the content, structure, and sufficiency of recommendations predicted the rate at which judges adopted recommendations into judicial decisions (e.g., Campbell and Schmidt 2000; Hecker and Steinberg 2002), none of the variables capturing variation in these same characteristics (i.e., content, use of ESTs, structure, sufficiency of recommendation explanations) predicted the degree to which JPOs accepted evaluators' recommendations and incorporated recommendations into youth disposition plans. One reason for this inconsistent finding may be that the restricted range of accepted recommendations by JPOs limited the ability to identify significant predictors based on the content and structure of the evaluations. As mentioned above, another reason for this discrepancy could be that JPOs are taking different factors into account when considering evaluator recommendations such as the availability of recommended services or perceived feasibility of securing the recommended service for the youth. Additionally, JPOs may rely on their own decision-making process to guide supervision or management of youth rather than information and recommendations from the evaluations. For example, JPOs in the jurisdiction utilized a locally developed risk/needs tool to help guide decisions about level of supervision and supervision strategies to be employed. JPOs may simply have found the results of the evaluation that primarily focused on mental health information as less relevant to their primary mandate of supervising youth in the community. Given the limited amount of time that youth are on probation, JPOs may be prioritizing recommendations that they see as most important and leaving out those that they feel would not be as beneficial. Lastly, the recommendations being offered by evaluators, even if well-supported, may simply be out of the scope of service or management strategies used by JPOs such that is was unlikely certain recommendations would be incorporated into the disposition plan. It appears that further research is needed to better understand the different explanations for the restricted range of accepted recommendations by JPOs in the current study.

Strengths and Limitations

The current study was strengthened by the large sample size of predisposition psychological evaluations and the availability of a large database of information on the youths' legal history, current offense information, case notes, and disposition plan. Additionally, the use of three trained coders served to lessen any bias that may have occurred if only one coder was used to rate evaluations. The inter-rater reliability between coders was also quite high lending credence to the codebook used in the current study. Finally, by using a random sampling of evaluations from several court jurisdictions statewide, the current results are not likely impacted by either evaluator characteristics, JPO decision-making, or service availability within any one single jurisdiction.

In terms of limitations, the current sample was drawn from one East Coast state; thus, results may not generalize to other jurisdictions. Additionally, the evaluators used in this state do not necessarily have specialized forensic training and as a result, the evaluators were more likely to focus on mental health issues rather than the legal history and factors more directly related to the supervision and management of youth while on community supervision probation. The documentation in the youth records made it difficult at times to discern the exact probation requirements in the disposition plans (i.e., use of 'standard probation' versus listing out the separate conditions of the plan) and this may have affected the validity of the match rate between evaluator recommendations and the youth's

disposition plan developed by the JPO. Lastly, only one indicator of "usefulness" of the evaluations was used (i.e., the extent to which JPOs adopted the recommendations into the disposition plan). JPOs may find other aspects of the evaluations useful in their day-to-day management and supervision (e.g., thorough descriptions of the youth and family functioning, educational functioning, and substance use history) that were not captured given the focus on the disposition plan.

Implications and Future Directions

The current results have important implications for evaluators with respect to report writing and JPOs in terms of using evaluation recommendations. The results call into question the depth of information reported on key background and psychosocial factors that may affect the usefulness of the evaluations in preparing disposition plans. Evaluators should be thorough when assessing various psychosocial factors that pertain to a youth's case and include pertinent details in their evaluations about history and current functioning across the full range of legal and sociocultural domains. For example, focusing on consistent and thorough documentation of domains that are known correlates of future delinquent behavior (i.e., prior/current offenses, family/parenting, peer associations, history and current use of substances, antisocial attitudes) (Cottle et al. 2001; Olver et al. 2009) should enhance the overall utility of the evaluations in developing effective disposition plans. Additionally, Viljoen et al. (2010) documented increasing use of structured risk tools by forensic evaluators in juvenile risk assessments such as the Structured Assessment of Violence Risk (SAVRY; Borum et al. 2006) and the Youth Level of Service Case Management Inventory (YLS/CMI 2.0; Hoge et al. 2011) and discussed how juvenile risk assessment results are linked to discussion of protective factors and treatment. Use of such tools is one way to ensure comprehensive coverage of known correlates of delinquent behavior in predisposition evaluations that could then inform recommendations about interventions that are more directly linked to delinquency risk and more closely aligned with the JPOs role of developing a case plan to address delinquency reduction and public safety. Additionally, evaluators must avoid technical jargon and present findings in a way that is understandable to the consumers (i.e., judges, JPOs, lawyers). The JPOs in this study incorporated approximately 1 out of 3 recommendations; however, variation in content and structure of the evaluations did not predict the rate at which JPOs used recommendations. This finding suggests that JPOs may need further training in how to review and consider the breadth and depth of the evaluation content, as well as how to identify sufficiently justified recommendations in the evaluation.

One approach to training JPOs is to focus on the Risk-Needs-Responsivity (RNR) model, which emphasizes the process of matching treatment services to the risk level of a given youth and that youth's particular need areas (Andrews and Bonta 2006). Briefly, there are three principles in the model: risk principle, match the level of service to the offender's risk to re-offend; need principle, assess criminogenic needs and target them in treatment; and responsivity principle, maximize the offender's ability to learn from interventions by providing evidence-based treatment and tailoring the intervention to the learning style, motivation, abilities, and strengths of the offender (Andrews and Bonta 2006). Recent studies demonstrate reductions in recidivism when there is a good match between justice-involved youths' needs and the recommended treatment services (Vincent et al. 2012a, b). Therefore, training JPOs to recognize the particular needs of a youth identified in an evaluation and matching the recommended services to the youth's particular needs may prove to be beneficial. Consistent with the RNR model, cognitive

strengths/deficits and/or specific identified mental health concerns that resulted in recommendations for mental health treatment (i.e., anxiety, depression, trauma reactions) should be understood as responsivity factors that could impact the youth's response to other delinquency reduction interventions. Training in case planning consistent with RNR model could facilitate JPOs integrating information across multiple sources of information (e.g., risk/needs tools included in the evaluation and/or conducted by JPO staff and predisposition psychological evaluations that primarily addressed mental health functioning as found in this study). Similar training could occur with evaluators to ensure that evaluations include a well-developed formulation that addresses how key needs are linked to problem behaviors which can also reinforce prioritizing and justifying the need/intervention match in recommendations (see Hart et al. 2011).

There may be other factors beyond content and structure of the evaluation used by JPOs when deciding what recommendations to include. This may include factors such as whether or not the JPO agrees with the recommendation or the availability of the recommended services. Even though this may be the case, evaluators should still recommend the best approach possible that is matched to the particular needs of a given youth, whether or not they are aware of the service being available in the community. Rather than tailor recommendations to what services are available, some form of quality assurance and/or feedback system could be put into place for JPOs to follow-up on how to prioritize recommendations made by evaluators. This may help JPOs better match what services they do have available in the community to the recommendations made in the evaluations and the particular needs of the youth.

Future studies in this area should conduct a similar examination of predisposition psychological evaluations in other state jurisdictions. Researchers in the current study did not look at differences in content, structure, and usefulness based on specific court jurisdiction; thus, this could also be analyzed in future studies as there may be differences in the frequency of use of evaluator recommendations depending on court jurisdiction. Training practices may differ based on the specific jurisdiction and/or the specific evaluators used within jurisdictions as well. Additionally, expanding the usefulness outcome by further exploring how JPOs are reviewing, understanding, and utilizing evaluations and evaluator's recommendations by conducting focus groups and/or interviews with the JPOs is recommended. Researchers could inquire into possible reasons why JPOs may not use certain recommendations, what information they find important in evaluations, how often they use these evaluations for other job responsibilities beyond developing the disposition plan, what components they pay the most attention to, and/or how these evaluations may or may not influence the way they supervise youth. For example, if the evaluation details extensive substance abuse by the youth, is the JPO more likely to supervise that youth more closely? Or if the youth's parents are more involved in their care, is the JPO likely to place less emphasis on supervision for that youth? Detailed interviews with JPOs may help to elucidate the reasons behind some of the findings in the current study and set the stage for future research in this area. It is critically important to understand how JJ personnel can best use comprehensive evaluations in developing case plans for justice-involved youth.

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families. ISBN 0-938565-73-7.

Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Newark, NJ: LexisNexis/Matthew Bender.

Archer, R. P., Simonds-Bisbee, E. C., Spiegel, D. R., Handel, R. W., & Elkins, D. E. (2010). Validity of the Massachusetts Youth Screening Instrument-2 (MAYSI-2) scales in JJ settings. *Journal of Personality Assessment, 92*(4), 337–348. doi:10.1080/00223891.2010.482009.

Bender, L. (1938). *A visual-motor Gestalt test and its clinical use*. American Orthopsychiatric Association Monograph Series Number 3. NY: American Orthopsychiatric Association.

Borum, R., Bartel, P., & Forth, A. (2006). *Structured assessment of violence risk in youth (SAVRY)*. Lutz, FL: Psychological Assessment Resources.

Brannigan, G. G., & Decker, S. L. (2003). *Bender visual–motor gestalt test* (2nd ed.). Itasca, IL: Riverside Publishing.

Burt, S. A., & Klump, K. L. (2013). Delinquent peer affiliation as an etiological moderator of childhood delinquency. *Psychological Medicine, 43*(6), 1269–1278.

Campbell, M., & Schmidt, F. (2000). Comparison of mental health and legal factors in the disposition outcome of young offenders. *Criminal Justice and Behavior, 27*(6), 688–715. doi:10.1177/0093854800027006002.

Cavendish, W. (2014). Academic attainment during commitment and postrelease education-related outcomes of juvenile justice-involved youth with and without disabilities. *Journal of Emotional And Behavioral Disorders, 22*(1), 41–52. doi:10.1177/1063426612470516.

Cook, A., & Gordon, J. A. (2012). Get him out of my house: Parental competencies of juvenile probationers. *Youth Violence And Juvenile Justice, 10*(2), 205–223. doi:10.1177/1541204011418352.

Cottle, C. C., Lee, R. J., & Heilbrun, K. (2001). The prediction of criminal recidivism in juveniles: A meta-analysis. *Criminal Justice and Behavior, 28*(3), 367–394. doi:10.1177/0093854801028003005.

Cruise, K. R., Evans, L. J., & Pickens, I. B. (2011). Integrating mental health and special education needs into comprehensive service planning for juvenile offenders in long-term custody settings. *Learning and Individual Differences, 21*, 30–40. doi:10.1016/j.lindif.2010.11.004.

Dalton, R. F., Evans, L. J., Cruise, K. R., Feinstein, R. A., & Kendrick, R. F. (2009). Race differences in mental health service access in a secure male juvenile justice facility. *Journal of Offender Rehabilitation, 48*(3), 194–209. doi:10.1080/10509670902766570.

Goldstein, N. S., Serico, J. M., Romaine, C., Zelechoski, A. D., Kalbeitzer, R., Kemp, K., & Lane, C. (2013). Development of the juvenile justice anger management treatment for girls. *Cognitive and Behavioral Practice, 20*(2), 171–188.

Grisso, T. (2010). Guidance for improving forensic reports: A review of common errors. *Open Access Journal of Forensic Psychology, 2*, 102–115.

Grisso, T. (2013). Chapter 5: Rehabilitation Evaluations of Juveniles. In T. Grisso (Ed.), *Forensic evaluation of juveniles* (2nd ed.). Sarasota, FL: Professional Resource Press.

Grisso, T., & Barnum, R. (2000, 2006). *Massachusetts youth screening instrument-second version*. Sarasota, FL: Professional Resource Press.

Grisso, T., & Underwood, L. A. (2004). Screening and assessing mental health and substance use disorders among youth in the juvenile justice system: A resource guide for practitioners. *Report.* Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Grisso, T., Vincent, G. M., & Seagrave, D. (2005). *Mental health screening and assessment in juvenile justice*. New York: Guilford Press.

Hart, S., Sturmey, P., Logan, C., & McMurran, M. (2011). Forensic case formulation. *The International Journal Of Forensic Mental Health, 10*(2), 118–126. doi:10.1080/14999013.2011.577137.

Hecker, T., & Steinberg, L. (2002). Psychological evaluation at juvenile court disposition. *Professional Psychology: Research and Practice, 33*(3), 300–306. doi:10.1037/0735-7028.33.3.300.

Heilbrun, K. (2001). *Principles of forensic mental health assessment*. New York: Kluwer/Plenum.

Hoge, R. D., Andrews, D. A., & Leschied, A. W. (2011). *Youth Level of Service/Case Management Inventory 2.0 (YLS/CMI 2.0)*. North Tonawanda, NY: Multi-Health Systems Inc.

Lachar, D., & Gruber, C. P. (1995). *Personality Inventory for Youth (PIY)*. Los Angeles: Western Psychological Services.

Lander, T. D., & Heilbrun, K. (2009). The content and quality of forensic mental health assessment: Validation of a principles-based approach. *International Journal of Forensic Mental Health, 8*(2), 115–121. doi:10.1080/14999010903199324.

Maschi, T., Hatcher, S., Schwalbe, C. S., & Rosato, N. (2008). Mapping the social service pathways of youth to and through the juvenile justice system: A comprehensive review. *Children and Youth Services Review, 30*(12), 1376–1385. doi:10.1016/j.childyouth.2008.04.006.

Murray, H. A. (1943). *Thematic Apperception Test manual*. Cambridge, MA: Harvard University Press.

O'Donnell, P. C., & Lurigio, A. J. (2008). Psychosocial predictors of clinicians' recommendations and judges' placement orders in a juvenile court. *Criminal Justice & Behavior, 35*(11), 1429–1448. doi:10.1177/0093854808324061.

Olver, M. E., Stockdale, K. C., & Wormith, J. (2009). Risk assessment with young offenders: A meta-analysis of three assessment measures. *Criminal Justice And Behavior, 36*(4), 329–353. doi:10.1177/0093854809331457.

Perrault, R. T., Paiva, M., & Vincent, G. M. (2012). Probation officers' perceptions of youths' risk of reoffending and use of risk assessment in case management. *Behavioral Sciences & the Law, 30*(4), 487–505. doi:10.1002/bsl.2015.

Sarris, A., Winefield, H. R., & Cooper, C. (2000). Behaviour problems in adolescence: A comparison of juvenile offenders and adolescents referred to a mental health service. *Australian Journal of Psychology, 52*(1), 17–22. doi:10.1080/00049530008255362.

Schwalbe, C. S., & Maschi, T. M. (2012). Probation officers' perspectives on interagency collaboration for juvenile offenders with mental health problems. *Psychiatric Services, 63*, 830–833. doi:10.1176/appi.ps.201100166.

Shapiro, C. J., Smith, B. H., Malone, P. S., & Collaro, A. L. (2010). Natural experiment in deviant peer exposure and youth recidivism. *Journal of Clinical Child and Adolescent Psychology, 39*(2), 242–251. doi:10.1080/15374410903532635.

Stewart, D. G., & Trupin, E. W. (2003). Clinical utility and policy implications of a statewide mental health screening process for juvenile offenders. *Psychiatric Services, 54*(3), 377–382. doi:10.1176/appi.ps.54.3.377.

Strategic Advantages, Inc. (1998). *Symptom Assessment-45 Questionnaire*. North Tonawanda, NY: Multi-Health Systems Inc.

Teplin, L. A., Abram, K. M., McClelland, G. M., Dulcan, M. K., & Mericle, A. A. (2002). Psychiatric disorders in youth in detention. *Archives of General Psychiatry, 59*(12), 1133–1143. doi:10.1001/archpsyc.59.12.1133.

Teplin, L. A., Abram, K. M., McClelland, G. M., Mericle, A. A., Dulcan, M. K., & Washburn, J. J. (2006, April). Psychiatric disorders of youth in detention. *Juvenile Justice Bulletin.* Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Teplin, L. A., Abram, K. M., Washburn, J. J., Welty, L. J., Hershfield, J. A., & Dulcan, M. K. (2013, February). The northwestern juvenile project: Overview. *Juvenile Justice Bulletin.* Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Viljoen, J. L., McLachlan, K., & Vincent, G. M. (2010). Assessing violence risk and psychopathy in juvenile and adult offenders: A survey of clinical practices. *Assessment, 17*(3), 377–395. doi:10.1177/1073191109359587.

Vincent, G. M. (2011). *Screening and assessment in juvenile justice systems: Identifying mental health needs and risk of reoffending*. Washington, DC: Technical Assistance Partnership for Child and Family Mental Health.

Vincent, G. M., Grisso, T., & Terry, A. (2007). Mental health screening and assessment in juvenile justice. In C. L. Kessler & L. Kraus (Eds.), *Mental health needs of young offenders: Forging paths toward reintegration and rehabilitation* (pp. 270–287). Cambridge: Cambridge University Press.

Vincent, G. M., Grisso, T., Terry, A., & Banks, S. (2008). Sex and race difference in mental health symptoms in juvenile justice: The MAYSI-2 national meta-analysis. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*(3), 282–290. doi:10.1097/CHI.0b013e318160d516.

Vincent, G. M., Guy, L. S., Gershenson, B. G., & McCabe, P. (2012a). Does risk assessment make a difference? Results of implementing the SAVRY in juvenile probation. *Behavioral Sciences and the Law, 30*(4), 384–405. doi:10.1002/bsl.2014.

Vincent, G. M., Paiva-Salisbury, M. L., Cook, N. E., Guy, L. S., & Perrault, R. T. (2012b). Impact of risk/needs assessment on juvenile probation officers' decision making: Importance of implementation. *Psychology, Public Policy, and Law, 18*(4), 549–576. doi:10.1037/a0027186.

Wasserman, G. A., Jensen, P. S., Ko, S. J., Cocozza, J., Trupin, E., Angold, A., et al. (2003). Mental health assessments in juvenile justice: Report on the consensus conference. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*(7), 752–761. doi:10.1097/01.CHI.0000046873.56865.4B.

Wasserman, G. A., McReynolds, L. S., Ko, S. J., Katz, L. M., Cauffman, E., Haxton, W., & Lucas, C. P. (2004). Screening for emergent risk and service needs among incarcerated youth: Comparing MAYSI-2 and voice DISC-IV. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*(5), 629–639. doi:10.1097/00004583-200405000-00017.

Wasserman, G. A., McReynolds, L. S., Lucas, C. P., Fisher, P., & Santos, L. (2002). The voice DISC-IV with incarcerated male youths: Prevalence of disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*(3), 314–321. doi:10.1097/00004583-200203000-00011.

Wasserman, G. A., McReynolds, L. S., Whited, A. L., Keating, J. M., Musabegovic, H., & Huo, Y. (2008). Juvenile probation officers' mental health decision making. *Administration and Policy in Mental Health and Mental Health Services Research, 35*, 477–478. doi:10.1007/s10488-008-0191-x.

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide range achievement Test 4 professional manual*. Lutz, FL: Psychological Assessment Resources.