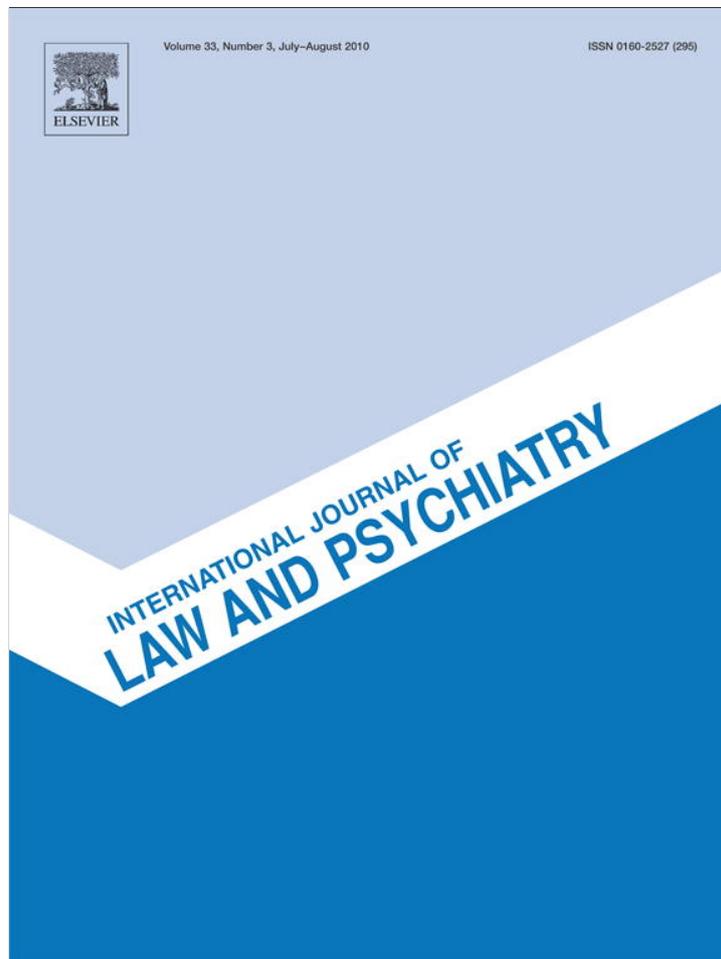


Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

## International Journal of Law and Psychiatry



## Fitness in paradise: Quality of forensic reports submitted to the Hawaii judiciary

Richard Robinson<sup>a</sup>, Marvin W. Acklin<sup>b,\*</sup><sup>a</sup> Argosy University Hawaii Campus, Honolulu, Hawaii, United States<sup>b</sup> Pacific Forensic Associates, Inc., 850 W. Hind Drive, Suite 203, Honolulu, Hawaii 96821, United States

## ARTICLE INFO

## Keywords:

Forensic reports  
Trial competency  
Forensic psychology

## ABSTRACT

This paper examined quality of forensic reports submitted to the Hawaii Judiciary. Hawaii utilizes a three panel system for assessing fitness to proceed, where two psychologists and one psychiatrist submit independent reports to the Court. Utilizing a survey instrument based on previous research and nationally-derived quality standards, 150 competency to stand trial (CST) reports were examined. Reports demonstrated pervasive mediocrity with respect to quality (Mean QC = 68.95, SD = 15.21). One quarter (N = 38) of the reports scored at or above 80% of the maximum possible score. Levels of CST agreement between evaluators and evaluators and judges were high. Report quality did not differ as a function of evaluator professional identity. Full-time employed evaluators submitted a greater number of reports above the quality criterion. For those evaluators who attended the March training, reports demonstrated significantly improved quality. Suggestions for enhancing report quality are offered with a special attention to inclusion of report elements, focus on inclusion of historical elements, and clearly described rationales supporting forensic opinions. (7664 words. Competency to stand trial, inter-rater agreement).

© 2010 Elsevier Ltd. All rights reserved.

## 1. Introduction

Forensic psychological assessment continues to be an active field within professional psychology. Competency to stand trial (CST) evaluations are the most common form of forensic evaluations. Nationally, an estimated 60,000 defendants are evaluated each year to assess their CST (Bonnie & Grisso, 2000). In 2005, there were 747 CST evaluations in Hawaii; of those evaluations, 230 (31%) were felony CST evaluations. Of 230 felony evaluations, 147 (64%) were from the First Circuit Court which covers the island of Oahu (Kennedy, 2006). From 2000 to 2005, the City and County of Honolulu, Department of the Prosecuting Attorney's, office reported that the office prosecuted an average of 2144 felony cases each year in the First Circuit Court (C. Izumoto & J. Fulton, personal communication, April 31, 2005). Based upon these numbers the question of the defendant's CST is raised in approximately 7% of felony cases. This is consistent with the national finding that between 2% and 8% of all felony defendants are referred for CST evaluations (Hoge, Bonnie, Poythress, & Monahan, 1992; LaFortune & Nicholson, 1995). In a study of felony CST evaluations from the First Circuit Court of Hawaii, Acklin et al. (2005) found that the defendant was determined to be Incompetent to Stand Trial (IST) in 16% of the cases (Waldorf, Acklin, & Kennedy, 2005). These results are similar to national findings that 16% of criminal defendants are found IST (Warren, Rosenfeld, & Fitch, 1994).

In Hawaii, when the question of CST is raised in a felony case, the court receives three independent forensic evaluation reports (Hawaii Revised Statutes [HRS], 704-404). There has never been a thorough evaluation of the quality of these reports in Hawaii. Studies in Florida, Oklahoma, Nebraska, Virginia, New Jersey, and Utah (Heilbrun & Collins, 1995; Otto, Barnes, & Jacobson, 1996; Robbins, Waters, & Herbert, 1997; Nicholson, La Fortune, Norwood, & Roach, 1995; Heilbrun, Warren, Rosenfeld, & Collins, 1994; Skeem & Golding, 1998; Skeem et al., 1998; Sanschagrin, 2005) have consistently found problems with the quality of forensic reports, so there is reason to suspect that the forensic reports in Hawaii may have similar problems. The literature illustrates that most forensic evaluations fail to incorporate many content elements deemed essential (Borum & Grisso, 1996; Crusie & Rogers, 1998; Grisso, 1996; Hecker & Steinberg, 2002; Nicholson & Norwood, 2000). Researchers and legal scholars agree certain data, clinical, and legal elements should be included in forensic mental health assessments; however, research indicates that this is not the actual practice (Cooper & Grisso, 1997; Nicholson & Norwood, 2000). This study contributes to the body of research related to the quality of forensic mental health assessments, with a special focus on the assessment of fitness in paradise.

This study examined the quality of Hawaii forensic evaluation reports based upon a model of forensic mental health assessment and report writing proposed by Melton et al. (1997), Heilbrun (2001), Nicholson and Norwood (2000), Mossman et al (2008), and Skeem and Golding (Skeem & Golding, 1998). In addition, this study incorporated the unique statutory requirements articulated in the Hawaii Revised Statutes (1979), the Hawaii Court's standard Order for Mental Examination, as well as the relevant practice guidelines of the American Psychological

\* Corresponding author. 850 W. Hind Drive, Suite 203, Honolulu, Hawaii 96821, United States.

E-mail address: acklin@hawaii.edu (M.W. Acklin).

Association (American Psychological Association, 2002) Ethical Principles of Psychologists and Code of Conduct; American Bar Association (American Bar Association, 1989) Criminal Justice Mental Health Standards; and the Specialty Guidelines for Forensic Psychologists (Committee on Ethical Guidelines, 1991).

We examined each element of forensic reports to evaluate forensic report writing practices in Hawaii. Several topical interests were examined: whether there are qualitative differences between reports prepared by evaluators employed by the State of Hawaii Department of Health (Courts and Corrections evaluators) and community-based psychologists and between community-based psychologist and psychiatrist evaluators. The role of specific training experiences on report quality was examined, by comparing the quality of reports produced by evaluators who attended the March 2005 Annual Forensic Examiners Training sponsored by the State of Hawaii Department of Health with those who did not attend. This training conference focused on legal standards, procedural and report standardization, including report formatting, to improve the consistency and overall quality of forensic reports submitted to the Judiciary. Report quality was compared for reports written prior to the training (January 2000 – March 2005) and after the training (March 2005 – December 2006). Finally, this study examined CST inter-rater agreement rates as well as agreement rates between evaluator opinions and judicial determinations.

## 2. CST evaluations in Hawaii

In Hawaii, the Dusky standard (*Dusky v. United States*, 362 U.S. 402 (1960))<sup>1</sup> has been codified into statutes that address “fitness to proceed.” Although different terminology is used in Hawaii, the intent of the Hawaii Revised Statutes (HRS) directly relates to the Dusky standard for competence to stand trial. The HRS states:

No person, who as a result of a physical or mental disease, disorder, or defect lacks capacity to understand the proceedings against the person or to assist in the person's own defense shall be tried, convicted, or sentenced for the commission of an offense so long as such incapacity endures (HRS, 704-403, 1972).

The commentary on this section connects the basis of the statute to “... the universally accepted position in Anglo-American law that a defendant cannot be proceeded against unless the defendant has the capacity to understand the proceedings against the defendant and assist in the defendant's own defense” (HRS, commentary on 704-403, 1972). As to the procedures for conducting a CST evaluation, the statute reads,

Once the issue of CST has been raised, the court shall appoint three qualified examiners in a felony case. In felony cases the court shall appoint at least one psychiatrist and at least one licensed psychologist. The third member may be either a psychiatrist, licensed psychologist, or qualified physician. One of the three shall be a psychiatrist or licensed psychologist designated by the director of health from within the Department of Health (HRS, 704-404, 1972).

The Department of Health maintains a staff of six full-time psychologists (“Courts and Corrections evaluators”) who conduct all CST evaluations in misdemeanor cases and are one of the three appointed forensic mental health evaluators (FMHE) in felony CST cases. By statute, the three FMHE must form their diagnoses and opinions independently. When the FMHE have been chosen by the court a standard Order for Mental Evaluation is issued. The order contains instructions for the FMHE and details what is required in the report. In

Hawaii, when a CST evaluation has been ordered, assessment of criminal responsibility (CR) is also typically ordered<sup>2</sup>.

The report of the examination includes the following: (a) A description of the nature of the examination; (b) A diagnosis of the physical or mental condition of the defendant; (c) An opinion as to the defendant's capacity to understand the proceedings against the defendant and to assist in the defendant's own defense ... (f) Where more than one examiner is appointed, a statement that the diagnosis and opinion rendered were arrived independently of any other examiner, unless there is a showing of a clear need for communication between or among the examiners for clarification. A description of the communication shall be included in the report (HRS, 704-404, 1979).

If, in the opinion of the FMHE, the defendant is either Incompetent to Stand Trial (IST) or not Criminally Responsible (CR) then the standard Order for Mental Evaluation adds the additional requirement of an assessment of the defendant's dangerousness and recommendations for the least restrictive care alternatives (HRS, 704-404, 1979). In Hawaii, most felony CST evaluations involve CR evaluations and may or may not include a dangerousness assessment, depending on the forensic opinion.

## 3. Assessing the quality of forensic reports

Several report elements are considered critical in forensic reports. First, these should include data elements: defendant's identifying information, demographics, and the defendant's criminal charges (Hart & Hare, 1992). Referral information should be included, identifying the referral source and the circumstances leading to the evaluation (Melton et al., 1997). Second, ethical elements should include a statement noting that the defendant was notified of the purpose of the evaluation and the limits of confidentiality, and confirmation the defendant understood the notification (Heilbrun, 2001; Nicholson & Norwood, 2000). Psychologist evaluators are required by both the Specialty Guidelines for Forensic Psychologists (Committee on Ethical Guidelines, 1991) and APA practice guidelines (APA, 2002) to provide notification of the limits of confidentiality and obtain confirmation that the defendant understood the notification, and document such notification and understanding. In reviewing seven studies in Florida, Utah, Virginia, New Jersey, Nebraska, and Oklahoma, Nicholson and Norwood (2000) found that documentation of notification of the purpose of the evaluation was rarely done.

Third, reports should include historical elements that detail the defendant's biopsychosocial history including a family history, educational and vocational history, marital history, prior medical and psychiatric history, prior criminal involvement, and substance use/abuse history (Petrella & Poythress, 1983). Heilbrun et al. (1994) suggested that an “accessibility effect” was probably at work in regards to sources of such information. The more readily obtainable information the more likely it is to be incorporated into the report.

Fourth, a thorough evaluation includes as many third-party sources of information as realistically possible (collateral elements). This is a key difference between forensic and traditional clinical assessments (Heilbrun et al., 1994; Melton et al., 1997). Effort to obtain this information may be an indication of motivation to get a more complete picture of the defendant and serves an essential role in corroborating or disconfirming information (Nicholson & Norwood, 2000). This is

<sup>2</sup> An anonymous reviewer raised concerns about self-incrimination in pre-trial examinations where comments made by the defendant might be used against him later in the prosecution (cf. *Estelle v. Smith*, 451 US 454 (1981)). H.R.S. section 704-416 provides: “Statements for purposes of examination or treatment [are] inadmissible except on issue of physical or mental condition. A statement made by a person subjected to examination or treatment pursuant to this chapter for the purposes of such examination or treatment shall not be admissible in evidence against the person in any penal proceeding on any issue other than that of the person's physical or mental condition....” This guarantee against self-incrimination is also supported in Hawaii case law (*State v. Wallace*, 71 Haw. 591, 801 P.2d 27 (1990), *State v. Domingo*, 69 Haw. 68, 733 P.2d 690 (1987), and *State v. Samuel*, 74 Haw. 141, 838 P.2d 1374 (1992)).

<sup>1</sup> “Whether a defendant has sufficient present ability to consult with his lawyer with a reasonable degree of rational understanding and he has a rational as well as factual understanding of the proceedings against him.”

especially important with respect to the defendant's self-report; because FMHE cannot assume that the defendant is being truthful (Christy, Douglas, Otto, & Petrila, 2004).

Fifth, clinical elements include clinical description, using a variety of resources, including mental status exams and psychological testing to assess the defendant's adjudicative competence-related skills. These include cognitive, affective, and behavioral functioning, information processing, expressive and receptive language skills, thought organization, reality testing, impulse control, mood and affective functioning.

Finally, opinion rationale elements provide information which forms the underlying basis for the forensic opinion. There has long been a debate within psychology about whether an expert should offer ultimate opinions concerning facts before the court. In Hawaii, ultimate opinion and rationale for the opinions are mandated in the statute. The rationale and reasoning underlying the CST opinion are as important as the opinion itself. The principal role in a forensic assessment is to describe the defendant's capacities and deficits, caused by a mental disorder or defect, as they relate to the defendant's criminal competencies (Grisso, 2003). Grisso (1986, 2003) proposed assessing specific characteristics which are shared by all legal competencies; he believes that "the primary components of all legal competencies are: (1) Functional, (2) Causal, (3) Interactive, (4) Judgmental, and (5) Dispositional" (p.23—the Judgmental and Dispositional elements fall within the jurisdiction of the Court). Evaluating these areas provides a comprehensive understanding of the defendant's psycholegal abilities and limitations (Grisso, 1986, 2003; Warren, Murrie, Chauhan, Dietz, & Morris, 2004; Williams & Miller, 1981). Borum and Grisso's (1996) experts viewed this information as essential to the quality of forensic reports.

### 3.1. Training and the quality of forensic reports

In a study that evaluated 100 CST reports submitted by community evaluators, Skeem and Golding (1998) found numerous deficiencies in the evaluation process and written reports. Skeem and Golding concluded that the underlying cause of the deficiencies was lack of adequate training. Despite ample research data critical of FMHE reports, improvement in report quality has been sparse and incremental. Problems persist despite implementation and development of practice guidelines and numerous journal articles and professional handbooks addressing the problem (Skeem & Golding, 1998). The impact training has on the quality of written forensic reports is still being debated. Some studies have found little or no improvement on the quality of reports after initiating training. Skeem & Golding (1998) found that an annual two day workshop for community-based evaluators did not improve the quality of forensic reports. However, Melton, Weithorn, and Slobogin (1985) found that examiners who attended a comprehensive 50-hour training that included supervised evaluations produced reports that were rated more favorably by legal personnel than examiners who did not attend. Other studies (Otto & Peters, 1990; Poythress, 1990) have reported similar results: evaluators who received forensic training were more knowledgeable about the legal and clinical issues related to forensic evaluations than evaluators who did not receive training.

### 3.2. Psychiatrist versus psychologist forensic evaluators and quality of forensic reports

In the 1970's, legal opinions and professional literature referred to psychologists' assessments as "second rate" (Perlin, 1977). In the 1980's, it was believed that psychiatrists performed better forensic evaluations than psychologists (Petrella & Poythress, 1983). Petrella and Poythress (1983) conducted a study in which reports of psychiatrists and psychologists that addressed the issues of CST and CR were compared and blindly rated by legal experts on thoroughness and quality. Results indicated no statistical difference between reports submitted by psychologists and psychiatrists but psychologists produced "more thorough" reports. By the 1990's, psychology had made great gains in

receiving recognition and parity in both the legal and medical community. Nevertheless, in a 1992 survey it was found that five states (Delaware, Iowa, New Jersey, Pennsylvania, and Vermont) continued to bar psychologist testimony on matters related to CST and CR (Farkas, DeLeon, & Newman, 1997). Concerning the issue of professional identity as factor in variation in CST opinions, Murrie and colleagues (Murrie, Boccaccini, Zapf, Warren, & Henderson, 2008), in a Virginia sample, found that rates of IST differed substantially by the evaluator's profession. Social workers had the highest rates of IST (mean = 46%), psychiatrists the lowest (5.8%), with psychologists falling in between (20%).

## 4. Method

In order to evaluate report quality, role of professional discipline in report quality, training effects, and levels of agreement among evaluators and between evaluators and the court, the following six research hypotheses were examined:

- (1) Report quality for reports submitted to the Hawaii Judiciary would score below 80% quality criterion.
- (2) Report quality for reports written by psychiatrists would be higher than reports written by community-based psychologists.
- (3) Report quality for reports written by full-time forensic psychologists employed by the State of Hawaii would be significantly higher than reports written by community-based FMHE.
- (4) Quality of reports submitted by examiners who attended a three-day forensic training conference in March 2005 would be significantly higher than reports written by evaluators who did not attend.
- (5) Agreement rates between CST opinions would reach levels recognized as acceptable in clinical psychology (Landis & Koch, 1977), that is kappa greater than or equal to .80 ("substantial" levels of agreement).
- (6) Agreement rates between the CST opinions of at least two FMHE and the judicial determination of the defendant's CST would reach levels recognized as acceptable in clinical psychology (Landis & Koch, 1977), that is kappa greater than or equal to .80.

### 4.1. Sampling procedures

We reviewed 150 forensic evaluation reports contained in 50 criminal case files. The evaluations were conducted on adults charged with felony offenses within Hawaii's First Circuit, which covers the island of Oahu. Only felony cases were examined in this study. It was determined a priori that the necessary number of reports to obtain acceptable statistical power of .80 was 42 cases for each FMHE group requiring a total sample size of 126 (Cohen, 1992).

Data for this study were collected from archives located at the First Circuit Court located in Honolulu. All of these records are held at the court and are publicly accessible. The 150 reports were drawn at random from more than 543 case files. Each case file included three separate forensic evaluation reports and a judicial determination of the defendant's CST. Only forensic reports addressing the initial question of CST were evaluated; however, by statute many of these reports also addressed CR and some contained a dangerousness assessment. Only reports that were filed after January 2000 were examined so as to provide a sample of the current state of forensic assessment in Hawaii.

### 4.2. Measures

A survey instrument was developed to assess forensic report quality. Formal definitions for each item were developed based on criteria proposed by Melton et al. (1997), Heilbrun (2001), Nicholson and Norwood (2000), and Skeem and Golding (Skeem & Golding, 1998), Sanschagrin (2005), the American Psychological Association's Ethical Principles of Psychologists and Code of Conduct (APA, 2002),

the American Bar Association's Criminal Justice Mental Health Standards (ABA, 1989), AAPL Practice Guideline For the Forensic Psychiatric Evaluation Of Competence to Stand Trial (2008), and the Specialty Guidelines for Forensic Psychologists (Committee on Ethical Guidelines, 1991). The instrument developed by Sanschagrin (2005) was the model for the current survey instrument. Although the Sanschagrin (2005) instrument provided an excellent model, the survey instrument was modified so that it better assessed the unique statutory requirements of the State of Hawaii, as well as the information ordered in the standard Order for Mental Examination<sup>3</sup>.

The survey instrument assessed a comprehensive list of quality elements. Information for each item was assessed for its presence or absence and for its quality. Individual items were given a score of 0 if the item was not included in the forensic report. A score of 1 was given if the item was included in the forensic report but was judged incomplete. A score of 2 was given if the item was included in the forensic report and rated as complete. There are a total of 38 items on the survey instrument; of those, eight items were statistical classification items and were not used in the calculations towards the total possible score. The remaining 30 items are equally weighed and used to calculate a quality coefficient (QC) score. Each item had a maximum score of 2 points for a total possible score of 60 points for a report that contained all of the recommended items of a forensic mental health evaluation. However, due to the "if...then" nature of the Hawaii Judiciary's Order for Mental Evaluation there are four maximum scores possible depending on the findings of the FMHE.

Once the total score for each evaluation was tabulated a Quality Coefficient (QC) was calculated. The QC was calculated by dividing the total score for each evaluation by the maximum possible score for a report in that category and converting that number into a percentage score.

The individual items were separated into different quality elements, including: data, ethical, historical, clinical, and rationale/opinion.

#### 4.3. Procedures

Investigators for this study were the first author (doctoral candidate in clinical psychology) and a research assistant (a doctoral-level graduate student in clinical psychology). The research assistant was blind to the study's hypotheses. Data collection occurred in two phases. The first phase consisted of the author and the research assistant independently rating ten reports using the survey instrument and the scoring criteria. Results from the reports were analyzed using Cohen's kappa (Cohen, 1992) to determine degree of inter-rater agreement. These reports were not included in the main study. With respect to interpretation of inter-rater agreement coefficients, this study utilized Landis and Koch's (1977) interpretation of kappa values, where kappa values of .41 to .60 indicate "moderate" levels of agreement, kappa values of .61 to .80 indicate "substantial" levels of agreement, and kappa values of .81 or greater suggest "almost perfect" levels of agreement (p. 159). Reliability analysis for the ten report training indicated that inter-rater agreement was "almost perfect" (mean  $k = .86$ ; Landis & Koch, 1977). Item level agreement ranged from .41 to 1.0 and report level kappa's ranged from .76 to 1.0. After the training was completed, the first author and research assistant reviewed the scoring of all items and further clarified their understanding of the scoring criteria.

After acceptable inter-rater agreement for scoring criteria was achieved the author and research assistant independently coded data from 150 reports (50 case files). To reduce error arising from rater drift or bias (Haynes, 1978), every tenth report was analyzed using Cohen's kappa for inter-rater agreement. The reliability analysis indicated that the inter-rater agreement was "almost perfect" (mean  $k = .93$ , Landis & Koch, 1977). Item level kappa's ranged from .41 to 1.0 and report level kappa's ranged from .90 to 1.0.

## 5. Results

One hundred and fifty forensic reports were coded. Based on Hawaii's statutory three panel system, for all felony defendants, one-third of the reports ( $n = 50$ ) were written by community-based psychiatrists, one-third of the reports ( $n = 50$ ) were written by community-based psychologists, and one-third of the reports ( $n = 50$ ) were written by psychologists employed by the State of Hawaii, Department of Health. Quality coefficient (QC) scores followed a normal distribution, Kilmogorov-Smirov  $Z = .638$ ,  $p = .811$ . The charges against the defendants were 40% violent crime ( $n = 20$ ), 34% property/other crime ( $n = 17$ ), 10% drug crime ( $n = 5$ ), 8% violent and property Crime ( $n = 4$ ), 4% violent and drug crime ( $n = 2$ ), and 4% property and drug crime ( $n = 2$ ). The court determined that 68% of defendants were CST ( $n = 34$ ) and 32% were IST ( $n = 16$ ). All of the evaluators employed by the State of Hawaii ("Courts and Corrections") attended the March 2005 training.

### Hypothesis 1. Quality Scores for Entire Sample.

Quality coefficient scores ranged from 21.7 to 100, with a mean of 68.95 ( $SD = 15.21$ ). One quarter ( $N = 38$ ) of the reports scored at or above 80% of the maximum possible score.

### Hypothesis 2. Quality Comparisons: Psychologist vs. Psychiatrist Evaluators.

Report quality was compared between community-based psychologists ( $n = 50$ ) and psychiatrists ( $n = 50$ ). Reports prepared by psychologists had QC scores that ranged from 35.2 to 96.3, with a mean of 65.82 ( $SD = 15.52$ ). Twenty percent ( $n = 10$ ) of the psychologists' reports scored at or above 80% of the maximum possible score. Reports prepared by psychiatrists had QC scores that ranged from 53.7 to 100, with a mean of 70.46 ( $SD = 10.96$ ). Twenty percent ( $n = 10$ ) of the psychiatrists' reports scored at or above 80% of the maximum possible score. No significant difference in report quality was found between reports submitted by psychologists and psychiatrists,  $t(100) = -1.73$ ,  $p = .088$ ,  $d = .35$ . Percentage of reports at or above the 80% quality criterion did not vary significantly between psychologists and psychiatrists,  $\chi^2(1, n = 100) = .00$ ,  $p = 1.00$ .

### Hypothesis 3. Quality comparison: community-based vs. courts and corrections evaluators.

This hypothesis compared quality of reports prepared by community-based evaluators ( $N = 100$ ) and full-time evaluators employed by Courts and Corrections ( $n = 50$ ). Reports prepared by community-based evaluators had QC scores that ranged from 35.2 to 100, with a mean of 68.14 ( $SD = 13.57$ ). Twenty percent ( $n = 20$ ) of reports scored at or above 80% of the maximum possible score. Reports prepared by Courts and Corrections evaluators had QC scores that ranged from 21.7 to 100, with a mean of 70.55 ( $SD = 18.08$ ). Thirty six percent ( $n = 18$ ) of the reports scored at or above 80% quality criterion score. Results revealed no significant QC difference between the community-based and Courts and Corrections reports,  $t(150) = -.83$ ,  $p = .407$ ,  $d = .15$ . Comparison of community-based and Courts and Corrections evaluation reports revealed that the percentage of reports above the 80% criterion varied significantly between the two groups,  $\chi^2(1, N = 150) = 4.51$ ,  $p = .034$  with a significantly greater percentage of Courts and Corrections reports above the 80% criterion (36% vs. 20%).

### Hypothesis 4. Quality comparison of attendees' report quality after annual training.

A three way ANOVA was utilized to evaluate the effects of type of evaluator; attendance at annual training; and before/after March training on report quality. The analysis yielded significant interaction effects, including Type FMHE-by-Attend Training,  $F(1, 130) = 5.37$ ,

<sup>3</sup> A copy of the instrument is available upon request from the corresponding author.

$p = .022$ ; Type FMHE-by-Before/After March,  $F(1, 130) = 10.07$ ,  $p = .002$ ; and Attend Training-by-Before/After March  $F(1, 130) = 7.37$ ,  $p = .008$ . Results revealed that there was a significant improvement in the report quality by community-based evaluation written after March 2005 by evaluators who attended the annual training (Fig. 1).

#### Hypothesis 5. CST Agreement Between Evaluators.

Quantification of CST opinion agreement revealed that in 70% of cases ( $N = 35$ ) all three evaluators agreed on the defendant's CST. In 24% of cases ( $n = 12$ ) at least two of three evaluators agreed on the defendant's CST. Thus, in 94% of cases ( $n = 47$ ) at least two of the three evaluators agreed on the defendant's CST. In 4% of the 50 cases ( $n = 2$ ) two of the three evaluators agreed on the defendant's CST and one had no opinion. In 2% of the 50 cases ( $N = 1$ ) none of the three evaluators agreed on the defendant's CST, (1-CST, 1-IST, 1-No opinion). Overall, evaluators demonstrated a "moderate" level of agreement ( $k = .56$ ; Landis & Koch, 1977). Evaluation of agreement between community-based psychologist and Courts and Corrections evaluators ( $N = 47$ , three cases no opinion) yielded CST agreements rates of 87% ( $N = 41$ ) and had a "substantial" level of agreement ( $k = .68$ ; Landis & Koch, 1977). When community-based psychologist evaluators were compared to community-based psychiatrists ( $n = 50$ ) rate of CST agreement was 82% ( $n = 41$ ) with a "moderate" level of agreement ( $k = .59$ ; Landis & Koch, 1977). Finally, when community-based psychiatrist and Courts and Corrections evaluators were compared rate of CST agreement was 74% ( $N = 41$ ) with a "fair" level of agreement ( $k = .40$ ; Landis & Koch, 1977).

#### Hypothesis 6. CST Agreement between: Forensic Evaluators and Judicial Determination of CST.

Judicial determination and all three evaluators agreed on CST in 66% of cases ( $n = 33$ ). Judicial determination and two evaluators agreed on CST in 24% of cases ( $n = 12$ ). Thus, in 90% of cases ( $n = 45$ ) judicial determination of CST agreed with two or more evaluators. In 6% of cases ( $n = 3$ ) the judicial determination of CST agreed with one of the three evaluators on CST. In 2% of cases ( $n = 1$ ) judicial determination of CST agreed with none of the evaluators on CST. In 6% of cases ( $n = 3$ ) one of the evaluators had no CST opinion. Overall, FMHE and judicial CST determination demonstrated a "substantial" level of agreement ( $k = .67$ ; Landis & Koch, 1977). Judges agreed with

Courts and Corrections evaluators ( $n = 47$ , three cases no opinion) on CST in 82% of cases ( $N = 41$ ) with a "substantial" level of agreement ( $k = .68$ ) (Landis & Koch, 1977). Comparison of judicial CST determinations of CST with community-based psychologists ( $n = 50$ ) indicated agreement in 88% of cases ( $n = 44$ ) with "substantial" agreement ( $k = .72$ ; Landis & Koch, 1977). When the judicial CST determinations were compared with community-based psychiatrists ( $n = 50$ ) agreement was obtained in 82% of cases ( $N = 41$ ) with a "moderate" level of agreement ( $k = .59$ ; Landis & Koch, 1977).

#### 5.1. Quality elements included in reports

Examining the frequency of quality elements in reports permits a better understanding of which elements account for quality problems.

#### 5.2. Data elements included in forensic reports

Nearly every report included the defendant's name; case caption; date of evaluation; date of report; and evaluator degree. However, only 39% of the reports ( $n = 58$ ) included the defendant's age, only 40% ( $n = 60$ ) included the charges against the defendant; and only 36% ( $N = 54$ ) reported an explanation of the evaluation procedure.

#### 5.3. Sources of information included in forensic reports

Evaluators reported utilizing four or more separate sources of information in 74% of reports ( $n = 111$ ). Evaluators reported utilizing between two and four sources of information in 24% of the reports ( $n = 37$ ). Thus, in 98% of cases ( $n = 148$ ) evaluators used two or more sources of information. The most common source of information utilized in reports was information obtained from the Adult Probation Division (APD) file (92%) followed closely by the clinical interview (91%), medical records (81%), and arrest reports (74%). Conversely, the least utilized sources of information were forensic assessment instruments (18%), psychological tests (13%), structured clinical or forensic interviews (9%), other sources (9%), and information from defense attorney (5%). Reported frequency of evaluators utilizing arrest reports may be deceiving because the arrest report is commonly part of the APD file; however if the evaluator did not specifically state that the arrest report was reviewed it was not scored as utilized. Psychiatric and medical records presented a similar problem as they are often combined in one record. If the evaluator did not clearly state that they had reviewed both records only the record they specifically mentioned was scored as utilized.

#### 5.4. Ethical elements included in forensic reports

Twenty four percent of the reports ( $n = 37$ ) included a complete statement that the limits of confidentiality were explained to the defendant. Complying with ethical requirements for disclosure does not appear to be a common practice amongst Hawaii forensic evaluators.

Historical Elements Included in Forensic Reports—Nearly all of the historical background items had low "complete" ratings. Information related to Medical History was scored as "complete" in 14% of reports ( $n = 21$ ), Psychiatric History was scored as "complete" in 37% of reports ( $n = 56$ ), Educational History scored as "complete" in 15% of reports ( $n = 22$ ), Vocational History scored as "complete" in 11% of reports ( $n = 17$ ), Criminal History scored as "complete" in 27% of reports ( $n = 41$ ), History of Violence scored as "complete" in 13% of reports ( $n = 19$ ), and Substance Use History scored as "complete" in 28% of reports ( $n = 42$ ).

#### 5.5. Clinical elements included in forensic reports

Thirty five percent ( $n = 53$ ) of the reports included a formal mental status examination (MSE). It should be noted that in many of the reports elements of the MSE were embedded in the CST



Fig. 1. Mean Quality Coefficient scores for all forensic assessment reports submitted prior to and following training for forensic examiners in March 2005.

rationale. Many reports did not have a formal MSE section, so they did not receive a rating of “complete.” Some commentators believe that the MSE deserves separate consideration. Since psychological testing—which provides information similar to mental status examinations—is underutilized, it seems plausible that the clinical description of the defendant is likely lacking in a majority of reports.

#### 5.6. CST elements included in forensic reports

One of the most important parts of the CST evaluation is the rationale utilized to explain and justify the forensic opinion. In 67% of the reports ( $n = 101$ ), evaluators provided a complete diagnosis of the defendant's condition. In 82% of reports ( $n = 122$ ), where a diagnosis was given evaluators included a complete causal or functional explanation of the defendant's impairment (Grisso, 2003). In 74% of the reports ( $n = 111$ ) where a psychiatric diagnosis was given the evaluator included a complete rationale for their opinion. The findings suggest that a substantial minority of reports failed to include these rationale elements supporting the forensic opinion.

#### 5.7. Criminal responsibility elements included in forensic reports

In 70% of reports ( $n = 87$ ), where a CR opinion was ordered, the evaluator included a complete opinion. Of those reports where an opinion of CR was given, 62% of reports ( $n = 73$ ) included a complete rationale for the opinion. Similar to CST rationale elements, a significant minority, 30 to 40%, of reports lacked a complete opinion providing linkage between clinical and functional impairment in criminal responsibility.

#### 5.8. Dangerousness elements included in forensic reports

In 76% of reports ( $n = 70$ ) where an opinion of the defendant's dangerousness was expected, evaluators included a complete opinion, and of those reports where an opinion of dangerousness was given, 77% ( $n = 66$ ) included a complete rationale for the opinion. Eighty percent ( $n = 71$ ) followed up with complete recommendations for the least restrictive care alternatives. An opinion concerning the least restrictive care standard is mandated by the statute.

### 6. Discussion

Overall, the quality of forensic reports in Hawaii is comparable to those in other states. That does not mean to say that there is not room for improvement. With only 25% of reports scoring above the 80% criterion, 66% of reports failing to document the ethically mandated notice of limits of confidentiality, and a general lack of attention to quality elements, especially historical and opinion rationale elements, there is clearly room for improvement. Although this study may be perceived as overly critical or too stringent, the goal was to assess the quality in Hawaii's forensic reports, make comparisons to other jurisdictions, and propose improvements, in much the same way that other studies have provoked improvements in report quality since the 1970s and 1980s. In looking at the past research on quality of reports in Hawaii it is clear that evaluators are making a concerted and successful effort to improve the quality of their forensic reports, especially when training is focused on formatting and content.

Feedback is necessary to improve report quality. Consistent with common knowledge in the observational methodologies (Haynes, 1978), the authors recommends that the Court provide evaluators with timely feedback on the quality and usefulness of their reports including the results of the CST judicial determination. One of the most important findings of the study was that reports of community-based evaluators who attended the March 2005 annual training, which focused on legal criteria, and standardization of procedures and reporting, showed a significant improvement in report quality. Overall,

however, report quality after the March 2005 training did not show an improvement because low quality reports continued to be submitted by those community-based evaluators who did not attend the training. These low quality reports attenuated the overall improvement in report quality made by those FMHE who attended the training. In contrast to the findings of Murrie et al. (2008), the study found that reports submitted by community-based psychiatrists and psychologists did not significantly differ with respect to report quality.

The issue of weighting of quality items deserves consideration. It is questionable whether it makes sense to assess the quality of forensic reports using an overall score on a survey instrument in which the items are equally weighted. One could convincingly argue that some items are more important than others when it comes to assessing report quality, and that using equally weighted items may therefore be misleading. For example, items relating to whether the evaluator included a rationale for his/her conclusion, or items pertaining to whether the evaluator documented the ethically mandated notice of limits of confidentiality, may ultimately be more important than items relating to other elements (such as identifying data, which should certainly be included in reports, but could be derived from other sources). We acknowledge that reliance on equally-weighted items is a limitation of the current study. The utility of using equally-weighted items in calculating report quality is an open empirical question.

One might argue that we set the quality standard for inter-rater agreement too high ( $kappa \geq .80$ ). There is no consensus on classifying level of agreement by magnitude of kappa. With respect to interpretation of inter-rater agreement coefficients, this study utilized Landis and Koch's (1977) interpretation of kappa values, where kappa values of .41 to .60 indicate “moderate” levels of agreement, kappa values of .61 to .80 indicate “substantial” levels of agreement, and kappa values of .81 or greater suggest “almost perfect” levels of agreement (p. 159). The psychiatric community considers  $k > .60$  or even  $k > .50$  as acceptable (Grove, Andreasen, McDonald-Scott, Keller, & Shapiro, 1981)<sup>4</sup>. Kappa greater than .80, that is, “almost perfect” agreement, may be an overly stringent standard given that the accuracy and agreement of forensic evaluator judgments are affected by numerous procedural, environmental, temporal, target, personal, and social factors (Acklin & Robinson, 2007). Consistent with recommendations in observational methodologies (Haynes, 1978), specification and clarity of statutory language; operationalization of legal definitions; standardization of information sources and evaluation process, including use of structured forensic assessment instruments; sensitivity to settings, time sequence, and variability in examiners; examination of bias; training of evaluators; and feedback about rater's psycholegal opinions are all likely to enhance the accuracy of opinions and quality of forensic reports submitted to the Courts (Acklin & Robinson, 2007).

The findings suggest that training on definitional criteria, procedures, and standardization of report format to promote thoroughness and consistency, enhanced report quality. Since the quality of reports for those evaluators who did not attend training showed no improvement, mandatory attendance at trainings may be required to insure consistent report quality. The use of checklists for quality elements and report formatting may help to standardize and improve report quality. The issue of definitional criteria appears to be a critical issue (Morris, Haroun, & Naimark, 2004). Of special focus is the commonly heard statement of judges that “we are interested in the thinking underlying your opinion.” Linking clinical data, functional legal impairment, and forensic opinion is likely to provide the most explicit type of rationale for forensic opinions (Grisso, 2003). Finally, the use of CST-related forensic assessment instruments (FAI; Otto &

<sup>4</sup> Shrout & Fleiss (1979) provide further guidelines for interpretation of kappa:  $<.40$  = poor;  $.40-.59$  = fair;  $.60-.74$  = good;  $>.74$  = excellent. Further recommendations for interpreting kappa (Cicchetti, 1994; Cicchetti & Sparrow, 1981) are:  $<.40$  = poor,  $.40-.59$  = fair,  $.60$  to  $.74/.79$  = good,  $>.75/.80$  = excellent, and  $>.80$  as nearly perfect.

Heilbrun, 2002), for example, the MacArthur Competence Assessment Tool—Criminal Adjudication (Mac-CAT-CA, Evaluation of Competency to Stand Trial-Revised (ECST-R), and the Fitness Interview Test-Revised (FIT-R) has been shown to improve report quality and to enhance inter-rater reliability (e.g., Rogers, Jackson, Sewell, Tillbrook, & Martin, 2003; Rogers, Tillbrook, & Sewell, 2004).

Attendance at the March 2005 training yielded a strong effect on quality of reports. This finding stands in contrast to some commentators (e.g., Skeem & Golding, 1998) who express skepticism that short-term training is effective in improving report quality. Through regular training, feedback, and continuing education, practitioners should develop a thorough understanding of the clinical and ethical conduct of forensic mental health evaluations and become increasingly proficient in offering their services to the Court.

## References

- Acklin, M. W., & Robinson, R. (2007). Quantifying consensus: The trial competency examination as interrater agreement trial. Manuscript submitted for publication.
- Acklin, M., Kennedy, R., Robinson, R., Dunkin, B., Dwire, J., & Lees, B. (2005). Quality of forensic reports: An empirical investigation of three panel reports. *Paper presented at the Annual State of Hawaii Forensic Mental Health Examiner Training Conference, Honolulu, HI* (March).
- American Bar Association. (1989). *Criminal justice mental health standards*. Washington D.C: Author.
- American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, 57, 1060–1073.
- Bonnie, R., & Grisso, T. (2000). Adjudicative competence and youthful offenders. In T. Grisso & R. Schwartz (Eds.), *Youth on trial* (pp. 73–103). Chicago, IL: University of Chicago Press.
- Borum, R., & Grisso, T. (1996). Establishing standards for criminal forensic reports: An empirical analysis. *Bulletin of the American Academy of Psychiatry and Law*, 24, 297–317.
- Christy, A., Douglas, K., Otto, R., & Petrla, J. (2004). Juveniles evaluated incompetent to proceed: Characteristics and quality of mental health professionals' evaluations. *Professional Psychology: Research and Practice*, 35, 380–388.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290.
- Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, 86(2), 127–137.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.
- Committee on Ethical Guidelines for Forensic Psychologists (1991). Specialty guidelines for forensic psychologists. *Law and Human Behavior*, 15, 655–665.
- Cooper, D. K., & Grisso, T. (1997). Five year research update (1991–1995): Evaluations for competence to stand trial. *Behavioral Sciences and the Law*, 15, 347–364.
- Cruise, K. R., & Rogers, R. (1998). An analysis of competency to stand to stand trial: Integration of case law and clinical knowledge. *Behavioral Sciences and the Law*, 16, 35–50.
- Dusky v. United States, 362 U.S. 402 (1960).
- Farkas, G., DeLeon, P., & Newman, R. (1997). Sanity examiner certification: An evolving national agenda. *Professional Psychology: Research and Practice*, 28, 73–76.
- Grisso, T. (1986). *Evaluating competencies: Forensic assessment instruments*. New York: Plenum.
- Grisso, T. (2003). *Evaluating competencies: Forensic assessment instruments*, 2nd ed. New York: Plenum.
- Grisso, T. (1996). Pretrial clinical evaluations in criminal cases: Past trends and future directions. *Criminal Justice and Behavior*, 23, 90–106.
- Grove, W. M., Andreasen, N., McDonald-Scott, P., Keller, M. B., & Shapiro, R. W. (1981). Reliability Studies of Psychiatric Diagnosis. Theory and Practice. *Archives of General Psychiatry*, 38, 408–413.
- Hart, S., & Hare, R. (1992). Predicting fitness for trial: The relative power of demographic, criminal and clinical variables. *Forensic Reports*, 5, 53–54.
- Hawaii Revised Statutes. (1979). Penal responsibility and fitness to proceed. *Chapter 704, Vol. 114*.
- Haynes, S. N. (1978). *Principals of Behavioral Assessment*. New York: Gardner Press.
- Hecker, T., & Steinberg, L. (2002). Psychological evaluation at juvenile court disposition. *Professional Psychology: Research and Practice*, 33, 300–306.
- Heilbrun, K. (2001). *Principles of forensic mental health assessment*. New York: Kluwer Academic/Plenum Publishers.
- Heilbrun, K., & Collins, S. (1995). Evaluations of trial competency and mental state at the time of the offense: Report characteristics. *Professional Psychology: Research and Practice*, 26, 61–67.
- Heilbrun, K., Warren, J., Rosenfeld, B., & Collins, S. (1994). The use of third party in forensic assessments. A two-state comparison. *Bulletin of the American Academy of Psychiatry and Law*, 22, 399–406.
- Hoge, S. K., Bonnie, R. J., Poythress, N., & Monahan, J. (1992). Attorney–client decision making in criminal cases: Client competence and participation as perceived by their attorneys. *Behavioral Sciences and the Law*, 10, 385–394.
- Kennedy, R. (2006). AMHD 2005/2006 Forensic Services update. *Paper presented at the Annual State of Hawaii Forensic Mental Health Examiner Training Conference, Honolulu, HI* (April).
- Lafortune, K., & Nicholson, R. (1995). How adequate are Oklahoma's mental health evaluations for determining competency in criminal proceedings? The bench and the bar respond. *Journal of Psychiatry and Law*, 23, 231–262.
- Landis, J., & Koch, G. G. (1977). The measurement of observed agreement for categorical data. *Biometrics*, 33, 159–174.
- Melton, G. B., Weithorn, L. A., & Slobogin, C. (1985). *Community-based mental health centers and the courts: An evaluation of community-based forensic services*. Lincoln, NE: University of Nebraska Press.
- Melton, G., Petrla, J., Poythress, N., & Slobogin, C. (1997). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers*, Second edition New York: Guilford Press.
- Morris, G. H., Haroun, A. M., & Naimark, D. (2004). Assessing competency competently: Toward a rational standard for competency to stand trial assessments. *Journal of the American Academy of Psychiatry and Law*, 32, 231–245.
- Mossman, D., Noffsinger, S., Ash, P., Frierson, R., Gerbasi, J., & Hackett, M. (2008). AAPL Practice Guideline for the Psychiatric Evaluation of Competence to Stand Trial. *Supplement to the American Academy of Psychiatry and the Law*, 35(4), S59–S67.
- Murrie, D. C., Boccaccini, M. T., Zapf, P. A., Warren, J. T., & Henderson, C. E. (2008). Clinician variation in findings of competence to stand trial. *Psychology, Public Policy & Law*, 14(3), 177–193.
- Nicholson, R., & Norwood, S. (2000). The quality of forensic psychological assessments, reports, and testimony: Acknowledging the gaps between promise and practice. *Law and Human Behavior*, 24, 9–44.
- Nicholson, R., LaFortune, K. A., Norwood, S., & Roach, R. L. (1995). Quality of pretrial competency evaluations in Oklahoma: Report Content and consumer satisfaction. *Paper presented as part of a symposium, Assessing the content and quality of forensic evaluations*. New York: R.K. Otto, Chair, at the annual convention of the American Psychological Association.
- Otto, R. K., Barnes, G., & Jacobson, K. (1996). The content and quality of criminal forensic evaluations in Florida. *Paper presented at the biennial meeting of the American Psychology Law Society, Hilton Head, SC*.
- Otto, R., & Peters, R. (1990). The Florida forensic evaluator training program. *Paper presented at the Biennial Meeting of the American Psychology–Law Society, Williamsburg, VA* (March).
- Otto, R., & Heilbrun, K. (2002). The practice of forensic psychology: A look toward the future in light of the past. *American Psychologist*, 57(1), 5–18.
- Petrella, R. C., & Poythress, N. G. (1983). The quality of forensic evaluations: An interdisciplinary study. *Journal of Consulting and Clinical Psychology*, 51, 76–85.
- Perlin, M. (1977). The legal status of the psychologist in the courtroom. *Journal of Psychiatry & Law*, 41, 41–54.
- Poythress, N. (1990). The Alabama forensic evaluator training program. *Paper presented at the Biennial Meeting of the American Psychology–Law Society, Williamsburg, VA* (March).
- Robbins, E., Waters, J., & Herbert, P. (1997). Competency to stand trial evaluations: A study of actual practice in two states. *Journal of the American Academy of Psychiatry and Law*, 25, 469–483.
- Rogers, R., Jackson, R., Sewell, K., Tillbrook, C., & Martin, M. (2003). Assessing dimensions of competency to stand trial: Construct validation of the ECST-R. *Assessment*, 10, 344–351.
- Rogers, R., Tillbrook, C. E., & Sewell, K. W. (2004). *Evaluation of competency to stand trial-revised*. Lutz, FL: Psychological Assessment Resources, Inc.
- Sanschagrinn, K. (2005). The quality of forensic mental health assessments of juvenile offenders: An empirical investigation. Unpublished doctoral dissertation, Drexel University, Philadelphia.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428.
- Skeem, J. L., & Golding, S. L. (1998). Community examiner's evaluations of competence to stand trial: Common problems and suggestions for improvement. *Professional Psychology: Research and Practice*, 29, 357–367.
- Skeem, J. L., Golding, S. L., Cohn, N. B., & Berge, G. (1998). The logic and reliability of evaluations of competence to stand trial. *Law and Human Behavior*, 22, 519–547.
- Waldorf, M., Acklin, M., & Kennedy, R. (2005). Certification update & report standards for mental examinations under HRS 704–404. *Paper presented at the Annual State of Hawaii Forensic Mental Health Examiner Training Conference, Honolulu, HI* (March).
- Warren, J., Murrie, D., Chauhan, P., Dietz, P., & Morris, J. (2004). Opinion formation in evaluating sanity at the time of the offense: An examination of 5, 175 pre-trial evaluations. *Behavioral Sciences and the Law*, 22, 171–186.
- Warren, J., Rosenfeld, B., & Fitch, W. F. (1994). Beyond competence and sanity: The influence of pretrial evaluation on case disposition. *Bulletin of the American Academy of Psychiatry & Law*, 22(3), 379–388.
- Williams, W., & Miller, K. (1981). The processing and disposition of incompetent mentally ill offenders. *Law and Human Behavior*, 5, 245–261.