COA NO. 32555-5-III

IN THE COURT OF APPEALS OF THE STATE OF WASHINGTON
DIVISION THREE

IN RE DETENTION OF RONALD LOVE:

STATE OF WASHINGTON,

Respondent,   .

v.

RONALD LOVE,

Appellant.

ON APPEAL FROM THE SUPERIOR COURT OF THE
STATE OF WASHINGTON FOR FRANKLIN COUNTY

The Honorable Robert G. Swisher, Judge

BRIEF OF APPELLANT

CASEY GRANNIS
Attorney for appellant

NIELSEN, BROMAN & KOCH, PLLC
1908 East Madison
Seattle, WA 98122
(206) 623-2373

TABLE OF CONTENTS

Page

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Page

TABLE OF AUTHORITIES

TABLE OF AUTHORITIES

TABLE OF AUTHORITIES

TABLE OF AUTHORITIES

A. ASSIGNMENTS OF ERROR

1. The evidence is insufficient to find appellant continues to be a sexually violent predator, in violation of due process.

2. The "to commit" instruction violated appellant's due process right to a fair trial. CP 16 (Instruction 5).

3. Appellant received ineffective assistance of counsel, in violation of due process.

4. The court erred in admitting hearsay testimony.

5. The court erred in admitting evidence of the Structured Risk Assessment - Forensic Version (SRA-FV), a predictive tool applied to appellant, under the Frye[1] standard.

6. The court erred in entering the following findings of fact and conclusions of law:

   a. "The SRA-FV is generally accepted within the community of experts who evaluate sex offenders and assess their recidivism risk." CP 3 (FF 9).

   b. "The use of structured analysis of risk factors in sex offender evaluations is supported by a scientific theory that is generally accepted in the scientific community." CP 4 (CL 3).

---

[1] Frye v. United States, 54 App. D.C. 46, 293 F. 1013 (D.C. Cir. 1923).

c.   "The use of a split sample for validation of a risk assessment instrument is supported by a scientific theory that is generally accepted in the relevant scientific community." CP 4 (CL 5).

d.   "The SRA-FV is an instrument that is capable of producing reliable results and is generally accepted in the scientific community." CP 4 (CL 6).

Issues Pertaining to Assignments Of Error

1.   Whether the evidence was insufficient to commit appellant as a sexually violent predator under the law of the case doctrine, where the "to commit" instruction required the jury to find a mental abnormality *or* personality disorder made him likely to reoffend, but the State's expert witness opined only the combination of mental abnormality *and* personality disorder made him likely to reoffend?

2.   Whether the "to commit" instruction violated appellant's right to due process because it allowed the jury to base its verdict on speculation, lessened the State's burden of proof, and was unwarranted by substantial evidence?

3.   Whether trial counsel rendered ineffective assistance in not objecting to the flawed "to commit" instruction?

4.   Whether the court committed reversible error in admitting the former testimony of a witness under ER 804(b)(1) because the State failed to establish the witness's unavailability?

5.   Whether the court committed reversible error in failing to exclude expert testimony on the SRA-FV because the State did not show the evidence being offered was based on an established methodology generally accepted in the scientific community under the Frye standard?

B.   STATEMENT OF THE CASE

1.   Procedural Facts

In 2001, the State filed a petition seeking Ronald Love's civil commitment under chapter 71.09 RCW. CP 526-53. In 2005, a judge found Love to be a sexually violent predator (SVP) following a bench trial and ordered his indefinite commitment. CP 804. Love has been confined to the Special Commitment Center (SCC) since then. 1RP[2] 866.

In 2013, Love showed probable cause that he had a permanent physiological change making him unable to commit a sexually violent act and had changed through a positive response to continuing participation in treatment. CP 804-07. The court ordered an unconditional release trial.

---

[2] The verbatim report of proceedings is referenced as follows: 1RP - 11 consecutively paginated volumes consisting of 5/13/14, 5/14/14, 5/15/14, 5/16/14, 5/19/14, 5/20/14, 5/21/14, 5/22/14, 5/23/14, 5/27/14, 5/28/14, 5/29/14, 5/30/14, 6/2/14; 2RP - 7/29/14.

CP 807. After the presentation of evidence at trial, a jury found Love continued to be an SVP. CP 8. The court ordered Love's continued commitment. CP 7. Love appeals. CP 6.

2.   Release Trial Evidence

Love was 57 years old at the time of trial. 1RP 944. Dr. Phenix, the State's expert witness, relied on past events in forming her opinion that Love continued to meet the SVP definition. 1RP 878-91. California juvenile records reflected an attempt in 1973 to force a six-year-old boy to perform fellatio and, in 1975, sodomy of a juvenile male and attempt to rape a juvenile female.[3] 1RP 878-79, 1008-10. Records also addressed a 1978 incident during which Love and two others attempted to abduct a 16-year-old female and threatened to rape her if she did not go with them.[4] 1RP 880, 1011-12.

Love pled guilty in California to two counts of forcible rape involving two women that took place on the same night in 1978.[5] 1RP 880-86, 1503-05. The former testimony of one of these women, A.P., was admitted into evidence over defense objection. CP 32-48; 1RP 1023-27,

---

[3] Love denied committing these offenses. 1RP 1488, 1897.
[4] The charges associated with this incident were dismissed. 1RP 1013-14. Love denied the allegation. 1RP 1496-97.
[5] Love denied raping either woman. 1RP 1410-13, 1599, 1895.

1199-1201. Love also entered an Alford[6] plea to attempted first degree

rape, which involved an assaultive encounter with a man in 1992.[7] 1RP

784, 792-95, 888-90, 1534.

Love did not admit to committing any sex offenses. 1RP 1016.

Love has not participated in formal sex offense treatment at the SCC,

except for a brief period of marginal participation. 1RP 950.

Dr. Phenix, a psychologist and expert in forensic psychology,

diagnosed Love with other specified paraphilic disorder - nonconsent,

alcohol use disorder and antisocial personality disorder. 1RP 858, 869,

894, 897, 902, 907-08. Phenix believed Love's mental condition made

him likely to reoffend. 1RP 913, 960. When asked if Love would be

dangerous if released, Phenix answered there was "a very strong

contribution of his antisocial personality disorder combined with his

sexual deviance to resulting five separate sexual offenses involving child

victims, teenagers, adults, males and females, with a very wide victim

pool." 1RP 960-61.

---

[6] North Carolina v. Alford, 400 U.S. 25, 91 S. Ct. 160, 27 L. Ed. 2d 162 (1970).
[7] Love testified that he was coerced into entering the guilty plea. 1RP 811, 818, 818-19. He denied committing the offense and described reacting to the man's sexual advances by hitting him. 1RP 830-31, 1425-26, 1433-34, 1437, 1467.

Phenix testified "there's a way that these three mental disorders work together to -- to cause him to be a danger in the future to commit criminal sexual acts, and that is that he has this abnormal sexual arousal. He's drawn to do that. That is disinhibited by his alcohol dependence and alcohol intake in the community so he's more likely to act out that sexual deviance. And that his antisocial personality disorder doesn't allow him to have the stops a normal person would have. It allows him to violate the rights of others so in that way it contributes to his sexual offending." 1RP 913. Love's alcohol dependence was not enough to qualify Love as an SVP. 1RP 988. Likewise, Love's personality disorder was not enough to qualify him as an SVP. 1RP 988-99. It was the combination of the paraphilia, the alcohol dependence and the personality disorder that contributed to Love's criminal sexual behavior. 1RP 990.

Dr. Phenix evaluated risk of reoffense using two actuarial instruments, the Static-99 Revised (Static-99R) and the Static-2002 Revised (Static-2002R), as well as the Structured Risk Assessment - Forensic Version (SRA-FV). 1RP 918, 937. The Static-99R and the Static-2002R instruments addressed static risk factors. 1RP 914, 918, 934-35. Phenix described the SRA-FV as a structured method of assessing dynamic (changeable) risk factors. 1RP 934-37.

Love's score on the Static-99R and Static-2002R placed him within the high risk range compared to other sex offenders in the study sample. 1RP 918, 925, 927. The group of offenders in the sample with the same score as Love had a 45 percent risk of reoffense in five years and 55.3 percent risk of reoffense in 10 years. 1RP 926. This does not mean Love is 55 percent likely to reoffend because actuarial instruments do not predict whether a particular person will or will not reoffend. 1RP 918, 926. Love's risk could be somewhere above or below those percentage figures. 1RP 926.[8] The Static-2002R probabilities were very similar. 1RP 927-28.

For the SRA-FV, individuals receive a score based on the presence of dynamic risk factors, which is associated with low, moderate or high risk groups in the Static-99R. 1RP 935-36, 939-40. Love's score on the SRA-FV placed him in the high risk group. 1RP 940.

Dr. Halon, a licensed psychologist testifying for the defense,[9] disagreed with Phenix's assessment and opined Love did not suffer from a mental abnormality or personality disorder. 1RP 1655-58, 1664, 1680, 1725, 1852. According to Halon, there was no way to make a valid diagnosis of mental abnormality because Love showed no signs of sexual

---

[8] Risk of reoffense lowers with increased age, which is a factor incorporated into the actuarial instruments. 1RP 920, 944.
[9] 1RP 1636-49.

deviance or serious difficulty controlling himself while in prison (since 1991) and the SCC (since 2001). 1RP 1655-58, 1664. If Love had a mental abnormality, he would not be able to hide it for so long. 1RP 1664. Love did not have a current paraphilia based on the evidence. 1RP 1725. The SCC records did not indicate Love had any sexuality left in him. 1RP 1658. The personality disorder Love had when he was growing up was now in remission because there were no longer any sign of it. 1RP 1680, 1852.

Phenix scored Love as a 31 on the Hare Psychopathy Checklist, Revised, where a cut-off score of 30 and above indicates the presence of psychopathy. 1RP 928, 932-33. But Halon opined that Love no longer showed signs of being a psychopath. 1RP 1779.

Dr. Donaldson, a clinical psychologist specializing in forensic psychology, also testified on Love's behalf. 1RP 1213. According to Donaldson, there are no models in psychology capable of accurately predicting individual behavior. 1RP 1225, 1271, 1292. The Static-99 instruments badly overestimate risk of reoffense because the base rate (how many sex offenders recidivate after release) is flawed. 1RP 1236-37. The Static-99R, the best instrument available, contained a serious error in the evaluator's handbook table relating to risk for the high risk group. 1RP

1239-42, 1271. Donaldson further opined the SRA-FV should not be used because it is a flawed predictive tool. 1RP 1243-56.

Dr. Sziebert, an SCC physician, testified that Love has high blood pressure, high cholesterol, thyroid disease, hepatitis C, and a painful condition of the spine and neck. 1RP 1369-72. The spine/neck problem affects range of motion. 1RP 1372-73, 1377. Love has difficulty rotating his trunk and making quick movements. 1RP 1377. Love receives medication for the condition. 1RP 1373. The condition is degenerative, chronic and incurable. 1RP 1374. It was only going to get worse. 1RP 1378. In Phenix's opinion, Love's health problems did not substantially lower his risk of reoffense. 1RP 945.

Love is very involved in the Native American circle and its healing practices at the SCC. 1RP 1311-12, 1344, 1347-51, 1387, 1389. Love considered his participation in Native American practices to be a form of treatment. 1RP 1090. Multiple witnesses testified based on personal observation that Love had shown positive change over the years in his disposition and interaction with others, including not involving himself with those engaged in negative behavior. 1RP 1078-81, 1086, 1308-09, 1315-16, 1400. The SCC chaplain had never seen Love explode or lash out, even in negative situations. 1RP 1401-02. Love expressed grief and regret about past offenses against others. 1RP 1080-81, 1334, 1407.

According to Dr. Phenix, Love's participation in Native American spiritual activities could not substitute for intensive sex offender treatment. 1RP 951-54. Phenix opined Love had not changed in any significant way that reduced his risk of reoffense. 1RP 963.

According to Dr. Halon, the Native American healing program is a legitimate form of treatment, different than the formal sex offender treatment offered at the SCC. 1RP 1691-93, 1719-23, 1789. Love had changed as a result of his participation in the Native American healing program. 1RP 1789.

Love would be on two years of Department of Corrections supervision upon release. 1RP 957, 1197. Love knew that he would receive a life sentence if he were convicted of another serious violent offense or sex offense. 1RP 1602.

If released, Love planned to live at a Spokane apartment complex that accepted sex offenders. 1RP 1600. Love had family support in the community. 1RP 955, 1563-66, 1576, 1581. He had contacted the Choctaw tribe for support and believed he would receive it. 1RP 1600-01. Love planned to continue Native healing if released. 1RP 955, 1604. Lee and Mix, two of the Native American elders involved in the native healing practices at SCC, would support him in terms of getting Native services. 1RP 1298, 1334, 1339-40, 1351-52.

C.   ARGUMENT

1.   THE VERDICT IS NOT SUPPORTED BY SUFFICIENT EVIDENCE UNDER THE LAW OF THE CASE DOCTRINE.

Under "the law of the case" doctrine, what facts need to be proven depends on how the jury is instructed. The jury was instructed that it must find beyond a reasonable doubt that Love suffers from a "mental abnormality or personality disorder" that makes him likely to commit predatory acts of sexual violence unless confined to a secure facility. CP 16. Use of the disjunctive "or" in this instruction, instead of the conjunctive "and," requires reversal of the verdict. The State's expert testified that the combination of Love's mental abnormality and personality disorder made him likely to reoffend, not that one or the other standing alone made him likely to reoffend. The evidence is therefore insufficient to sustain Love's commitment under the jury instructions, in violation of due process. U.S. Const. amends. V, XIV; Const. art. 1, § 3.

a.   The State must prove mental illness and the likelihood of reoffense due to such illness.

Chapter 71.09 RCW authorizes the commitment of those found to meet the SVP definition. RCW 71.09.060(1). An SVP is "any person who has been convicted of or charged with a crime of sexual violence and who suffers from a mental abnormality or personality disorder which

makes the person likely to engage in predatory acts of sexual violence if not confined in a secure facility." RCW 71.09.020(18).

A person committed as an SVP has the right to an unconditional release trial if there is probable cause to believe he no longer meets the SVP definition. RCW 71.09.090(2)(c). At the unconditional release trial, "the burden of proof shall be upon the state to prove beyond a reasonable doubt that the committed person's condition remains such that the person continues to meet the definition of a sexually violent predator." RCW 71.09.090(3)(c).

A person shall not be deprived of liberty without due process of law. U.S. Const. amends. V, XIV; Const. art. 1, § 3. A person must be both mentally ill and dangerous for a civil commitment to be permissible under the due process clause of the constitution. In re Pers. Restraint of Young, 122 Wn.2d 1, 27, 857 P.2d 989, 1001 (1993) (citing Addington v. Texas, 441 U.S. 418, 99 S. Ct. 1804, 60 L. Ed. 2d 323 (1979); Foucha v. Louisiana, 504 U.S. 71, 112 S. Ct. 1780, 118 L. Ed. 2d 437 (1992)). By properly finding all the statutory elements are satisfied to commit someone as an SVP, the fact finder impliedly finds that the SVP is currently dangerous. In re Detention of Moore, 167 Wn.2d 113, 124, 216 P.3d 1015, 1021 (2009).

b. **Because the "to commit" instruction required the jury to find in the disjunctive regarding which condition made Love likely to reoffend, the verdict is not supported by sufficient evidence under the law of the case doctrine.**

"Although the commitment proceedings are civil in nature, given the standard of proof, the sufficiency of evidence is examined under the standard of beyond a reasonable doubt." In re Detention of Audett, 158 Wn.2d 712, 728 n.10, 147 P.3d 982 (2006). The commitment will be upheld only if, viewing the evidence in the light most favorable to the State, "any rational trier of fact could have found the essential elements beyond a reasonable doubt." Audett, 158 Wn.2d at 727-28.

What those elements are depends on how the jury is instructed. Whether the evidence is sufficient to sustain a verdict is measured by the jury instructions. "The law of the case is an established doctrine with roots reaching back to the earliest days of statehood." State v. Hickman, 135 Wn.2d 97, 101, 954 P.2d 900 (1998). This doctrine refers to the "rule that the instructions given to the jury by the trial court, if not objected to, shall be treated as the properly applicable law." Lutheran Day Care v. Snohomish County, 119 Wn.2d 91, 113, 829 P.2d 746 (1992) (quoting 15 L. Orland & K. Tegland, Wash. Prac., Judgments § 380, at 56 (4th ed. 1986)).

Where a party challenges the sufficiency of evidence on appeal, "[t]he sufficiency of the evidence to sustain the verdict is to be determined by the application of the instructions." Tonkovich v. Dep't of Labor & Indus., 31 Wn.2d 220, 225, 195 P.2d 638 (1948); accord Hickman, 135 Wn.2d at 102 ("to convict" instruction was law of the case); see also United States v. Spletzer, 535 F.2d 950, 954 (5th Cir. 1976) (unnecessary specific intent requirement included in jury instructions became necessary element of conviction under the "law of the case"); United States v. Nacchio, 519 F.3d 1140, 1157 (10th Cir. 2008) ("when asking what facts the jury had to find in order to convict, we look to the elements of the crime as defined by law, except that if the government did not object to jury instructions containing additional requirements, it is required to prove those too."), vacated in part on other grounds, 555 F.3d 1234 (10th Cir. 2009) (en banc).

The court gave the following "to commit" instruction to the jury:

> To establish that Ronald Love is a sexually violent predator, the State must prove each of the following elements beyond a reasonable doubt:
> (1)   That Ronald Love was previously found to be a sexually violent predator;
> (2)   That Ronald Love continues to suffer from a mental abnormality or personality disorder which causes him serious difficulty controlling his sexually violent behavior; and

> (3) *The mental abnormality or personality disorder continues to make Ronald Love likely to commit predatory acts of sexual violence unless confined to a secure facility.*
>
> If you find from the evidence that each of these elements has been proved beyond a reasonable doubt, then it will be your duty to return a verdict that Ronald Love continues to be a sexually violent predator.
>
> On the other hand, if, after weighing all of the evidence, you have a reasonable doubt as to any of one or more of these elements, then it will be your duty to return a verdict that Ronald Love is no longer a sexually violent predator.

CP 16 (Instruction 5) (emphasis added).[10]

"Mental abnormality" and "personality disorder" are alternative means for making the SVP determination. In re Detention of Halgren, 156 Wn.2d 795, 810, 132 P.3d 714 (2006). Each has its own particular statutory definition.[11] The jury was instructed on those definitions. CP 17, 28.

---

[10] The State proposed this instruction, which is taken from Washington Pattern Instruction 365.34. CP 572. Love's counsel unsuccessfully objected to the wording of the second element of the instruction, arguing the word "continued" should be replaced with the word "current." 1RP 1816-19.

[11] "Mental abnormality" means "a congenital or acquired condition affecting the emotional or volitional capacity which predisposes the person to the commission of criminal sexual acts in a degree constituting such person a menace to the health and safety of others." RCW 71.09.020(8). "Personality disorder" means "an enduring pattern of inner experience and behavior that deviates markedly from the expectations of the individual's culture, is pervasive and inflexible, has onset in adolescence or early adulthood, is stable over time and leads to distress or impairment." RCW 71.09.020(9).

Use of the disjunctive "or" in the third element of the "to commit" instruction has legal consequences under the law of the case doctrine. "Or" signifies an alternative between two things. "And" signifies connection between two things; i.e. one thing added to another. Courts recognize the difference in meaning between the disjunctive "or" and the conjunctive "and" in jury instructions. See State v. Owens, 180 Wn.2d 90, 101 n.6, 323 P.3d 1030 (2014) (conjunctive "and" rather than a disjunctive "or" in the "to convict" instruction became the law of the case in the absence of objection); State v. Stephens, 93 Wn.2d 186, 189-90, 607 P.2d 304 (1980) (where defendant was charged with one count of assault against two victims conjunctively, the jury instruction referencing the names of the victims in the disjunctive rather than conjunctive violated right to jury unanimity).

The "to commit" instruction, through use of the disjunctive, required the State to prove Love's mental abnormality continues to make Love likely to commit predatory acts of sexual violence unless confined to a secure facility *or* that Love's personality disorder continues to make Love likely to commit predatory acts of sexual violence unless confined to a secure facility. The problem is that Dr. Phenix, the State's expert, did not testify that Love's mental abnormality *or* personality disorder made Love likely to commit predatory acts of sexual violence unless confined to

a secure facility. Instead, Dr. Phenix testified the combination of the mental abnormality *and* personality disorder made Love likely to reoffend. 1RP 913, 960-61, 988-90. Dr. Phenix testified in the conjunctive. But the "to commit" instruction required the jury to find the disjunctive in order to satisfy the third element.

Dr. Phenix's expert testimony was necessary to support a jury finding that Love suffered from a mental abnormality and personality disorder that caused him to likely commit future acts of predatory sexual violence. "In general, expert testimony is required when an essential element in the case is best established by an opinion which is beyond the expertise of a layperson." Berger v. Sonneland, 144 Wn.2d 91, 110, 26 P.3d 257 (2001) (quoting Harris v. Robert C. Groth, M.D., Inc., P.S., 99 Wn.2d 438, 449, 663 P.2d 113 (1983)). "Medical facts must be proved by expert testimony unless they are observable by laypersons and describable without medical training." Berger, 144 Wn.2d at 111.

Determining whether a particular person possesses a qualifying mental condition under chapter 71.09 RCW "is based upon the complicated science of human psychology and is beyond the ken of the average juror." In Re Detention of Bedker, 134 Wn. App. 775, 779, 146 P.3d 442 (2006) (addressing "mental abnormality); see also RCW 71.09.020(9) ("Purported evidence of a personality disorder must be

supported by testimony of a licensed forensic psychologist or psychiatrist."). Similarly, a jury does not possess the specialized knowledge or medical training necessary to formulate a sound opinion on whether a mental condition makes someone likely to reoffend. Expert testimony will generally be necessary to establish most elements of causation. Berger, 144 Wn.2d at 110. Expert testimony was necessary to enable a valid jury finding that Love was likely to commit predatory acts of sexual violence if not confined in a secure facility.[12]

The existence of a fact cannot rest in guess, speculation, or conjecture." State v. Colquitt, 133 Wn. App. 789, 796, 137 P.3d 892 (2006). In the absence of expert testimony that either a mental abnormality *or* personality disorder caused Love to be at risk of reoffense, sufficient evidence is lacking to prove the proposition required by the jury instruction. The jury's verdict must be vacated and the court's commitment order reversed due to insufficient evidence.

---

[12] The State argued to the jury: "Does he have a mental abnormality or personality disorder? The State alleges that his paraphilia, that other specified paraphilic disorder, nonconsent, the paraphilia for rapists who were sexually aroused to the nonconsenting person that they are forcing sex upon is a mental abnormality. He's also diagnosed with antisocial personality disorder. In this case, because of that technical evidence, the diagnosis and the risk assessment, we had to put on expert testimony." 1RP 1953.

2. THE DISJUNCTIVE "TO COMMIT" INSTRUCTION IS FLAWED BECAUSE IT ALLOWED THE JURY TO BASE ITS VERDICT ON SPECULATION, LESSENED THE STATE'S BURDEN OF PROOF, AND WAS UNWARRANTED BY THE EVIDENCE.

Even if the evidence is sufficient to sustain the verdict, the "to commit" instruction is flawed because substantial evidence did not support use of the disjunctive "or" on the issue of whether the mental abnormality or the personality disorder made Love likely to reoffend. The instruction improperly allowed the jury to base its verdict on a finding that either the mental abnormality or the personality disorder made Love likely to reoffend, rather than requiring the jury to find both conditions made him likely to reoffend. In this manner, the instruction permitted the jury to render a verdict based on speculation and lessened the State's burden of proving both conditions, rather than one or the other, made Love likely to reoffend. It is prejudicial error to submit an issue to the jury when substantial evidence does not support it. In the alternative, Love's counsel provided ineffective assistance in failing to object to the "to commit" instruction.

> **a. The disjunctive language in the "to commit" instruction permitted the jury to choose between the mental abnormality and the personality disorder as the sole condition that made Love likely to reoffend, but the evidence did not support such a finding.**

The adequacy of jury instructions is reviewed de novo review. State v. Clausing, 147 Wn.2d 620, 626-27, 56 P.3d 550 (2002). "[T]he

- 19 -

chief objects contemplated in the charge of the judge are to explain the law of the case, to point out the essentials to be proved on the one side or the other, and to bring into view the relation of the particular evidence adduced to the particular issues involved." State v. Allen, 89 Wn.2d 651, 654, 574 P.2d 1182 (1978). "The instructions to be given in a particular case are governed by the facts proven in the case and instructions which are overly broad or which allow the jury to speculate as to the facts are improper." Harris, 99 Wn.2d at 447 (internal citation omitted).

The problem is that the "to commit" instruction, through use of the disjunctive, allowed the jury to find Love was an SVP if either the mental abnormality made him likely to reoffend or the personality disorder made him likely to reoffend. See Viking Automatic Sprinkler Co. v. Pac. Indem. Co., 19 Wn.2d 294, 298, 142 P.2d 394 (1943) ("Framed in the disjunctive, as it is, the instruction permitted the jury to return a verdict for respondent without regard to [one of the causes of the harm]."); State v. Bower, 28 Wn. App. 704, 708, 626 P.2d 39 (1981) ("Here 'threat' was defined to include the requisite mental state, but the disjunctive instruction was inadequate to inform the jury that the alternatives of force or violence had to be accompanied by the knowledge or intent that the conduct would prevent the performance of the guard's duties."), disapproved on other grounds by State v. Kjorsvik, 117 Wn.2d 93, 812 P.2d 86 (1991).

Substantial evidence did not support use of the disjunctive "or" in the "to commit" instruction: "The mental abnormality *or* personality disorder continues to make Ronald Love likely to commit predatory acts of sexual violence unless confined to a secure facility." CP 16. Dr. Phenix did not testify that the mental abnormality or the personality disorder made Love likely to reoffend. She testified the combination of the mental abnormality and the personality disorder made Love likely to reoffend. 1RP 913, 960-61, 988-90.

Substantial evidence does not support a finding that one or the other made him likely to reoffend. "[I]t is prejudicial error to submit an issue to the jury when there is not substantial evidence concerning it." State v. Hughes, 106 Wn.2d 176, 191, 721 P.2d 902 (1986). "[T]he giving of the instruction indicates to the jury that the court must have thought there was some evidence on the issue." Albin v. National Bank of Commerce of Seattle, 60 Wn.2d 745, 754, 375 P.2d 487 (1962).

Love had the right "to have a jury base its decision on an accurate statement of the law applied to the facts in the case." State v. Miller, 131 Wn.2d 78, 90-91, 929 P.2d 372 (1997). The jury should therefore have been instructed that it had to find "The mental abnormality *and* personality disorder continues to make Ronald Love likely to commit predatory acts of sexual violence unless confined to a secure facility." Such an

instruction would have complied with the mandate to give an instruction "governed by the facts proven in the case" and "to bring into view the relation of the particular evidence adduced to the particular issues involved." Harris, 99 Wn.2d at 447; Allen, 89 Wn.2d at 654.

Use of the disjunctive in this instruction was unwarranted by the evidence presented to the trier of fact. By giving the instruction worded in the disjunctive, the court sent a message to the jury that evidence could support a finding that one or the other made Love likely to reoffend. Albin, 60 Wn.2d at 754. The instruction may have misled the jury into believing it could find Love was an SVP based on the mental abnormality or personality disorder alone as the cause of risk of reoffense.

"[W]hen the record discloses an error in an instruction given on behalf of the party in whose favor the verdict was returned, as it does here, the error is presumed to be prejudicial and requires a new trial unless it affirmatively appears that the error was harmless." Zwink v. Burlington N., Inc., 13 Wn. App. 560, 569, 536 P.2d 13 (1975). The State cannot show the error was harmless because it cannot affirmatively show jurors found both the mental abnormality and the personality disorder, acting in combination, made Love likely to reoffend.

Jurors may have rejected Dr. Phenix's paraphilia diagnosis (the mental abnormality) and accepted the personality disorder diagnosis, in

which case it found Love to be an SVP on a basis for which there is no substantial evidence in support because the personality disorder alone did not make Love likely to reoffend.[13] Dr. Halon opined Love did not suffer from the mental abnormality of paraphilia. 1RP 1655-58, 1664, 1725. According to Halon, the paraphilia diagnosis is unreliable and should never be used in the forensic arena for decision-making. 1RP 1683-86. Even Phenix acknowledged that editors of the DSM-IV[14] opined there is no real diagnosis for rape paraphilias and that such a diagnosis is inappropriately used to civilly commit sex offenders. 1RP 990-92, 1107.[15] From such testimony, jurors could have rejected Phenix's mental abnormality diagnosis.

Conversely, jurors may have rejected Dr. Phenix's personality disorder diagnosis and accepted the mental abnormality diagnosis, in which case it found Love to be an SVP on a basis for which there is no substantial evidence in support because the mental abnormality alone did not make Love likely to reoffend. Dr. Halon opined Love did not suffer

---

[13] The jury was instructed that it was not required to accept an expert witness's opinion. CP 14 (Instruction 3).

[14] The Diagnostic and Statistical Manual of Mental Disorders (DSM) is the authoritative source commonly relied on to diagnose mental disorders. 1RP 868.

[15] Phenix did not believe this opinion to be the general consensus on the matter. 1RP 1174. The proposed inclusion of "coercive paraphilia" as an official diagnosis in the main body of the DSM-5 was rejected. 1RP 1105.

from a personality disorder. 1RP 1680, 1852. Halon also opined the antisocial personality disorder diagnosis was not a reliable diagnosis because of the low rate of agreement among evaluators about which signs verify the presence of the disorder. 1RP 1677-78. From such testimony, jurors could have rejected Phenix's personality disorder diagnosis.

The existence of a fact cannot rest in guess, speculation, or conjecture. Gardner v. Seymour, 27 Wn.2d 802, 808, 180 P.2d 564 (1947); Colquitt, 133 Wn. App. at 796. Nor can a verdict. Prentice Packing & Storage Co. v. United Pac. Ins. Co., 5 Wn.2d 144, 164, 106 P.2d 314 (1940). The issue of whether Love's mental abnormality continues to make him likely to commit predatory acts of sexual violence unless confined to a secure facility *or* that Love's personality disorder continues to make Love likely to commit predatory acts of sexual violence unless confined to a secure facility should not have been presented to the jury via the "to commit" instruction. Use of the disjunctive "or" in the instruction, instead of the conjunctive "and," allowed the jury to base its verdict on speculation rather than substantial evidence.

And it lessened the State's burden of proof. Instead of requiring the State to prove both the mental abnormality and the personality disorder made Love likely to reoffend, the instruction permitted the jury to render a verdict against Love if it found either one of those conditions made him

- 24 -

likely to reoffend. Substantial evidence did not support such a finding. "It is prejudicial error to submit an issue to the jury that is not warranted by the evidence." Clausing, 147 Wn.2d at 627.

b. The instructional error may be raised for the first time on appeal as a manifest constitutional error.

Defense counsel did not object to the "to commit" instruction on this basis. The error may be raised for the first time on appeal as a manifest error of constitutional magnitude. RAP 2.5(a)(3).

Love has the due process right not to be committed unless he is found to be dangerous — likely to reoffend — due to mental illness. Young, 122 Wn.2d at 31-32; U.S. Const. amends. V, XIV; Const. art. 1, § 3. The instruction at issue implicates Love's due process right to a fair trial because it allowed the jury to render a verdict based on insufficient evidence that he was likely to reoffend. To commit Love, the jury was required to find both the mental abnormality and the personality disorder made him likely to reoffend. The disjunctive "to commit" instruction, however, permitted the jury to find Love met the SVP definition if either one of those conditions made him likely to reoffend. See State v. Byrd, 72 Wn. App. 774, 782, 868 P.2d 158 (1994) ("Any time a requirement for conviction is not clearly stated in the instructions, a question of constitutional due process is presented."), aff'd. 125 Wn.2d 707, 887 P.2d

396 (1995). Violation of the due process right to a fair trial by a misleading and legally inapplicable instruction is an error of constitutional magnitude under RAP 2.5(a)(3). State v. O'Hara, 167 Wn.2d 91, 98-99, 105, 217 P.3d 756 (2009).

A constitutional error is manifest "if it results in a concrete detriment to the claimant's constitutional rights, and the claimed error rests upon a plausible argument that is supported by the record." State v. WWJ Corp., 138 Wn.2d 595, 603, 980 P.2d 1257 (1999). In determining whether actual prejudice is present under the manifest error analysis, the focus is on "whether the error is so obvious on the record that the error warrants appellate review." O'Hara, 167 Wn.2d 91 at 99-100. An error is manifest if the trial court could have foreseen the potential error. State v. Lamar, 180 Wn.2d 576, 583, 327 P.3d 46, 50 (2014).

The trial judge in Love's case listened to Dr. Phenix's expert testimony along with everyone else. The court was aware that she testified that both the mental abnormality and the personality disorder combined to make Love likely to reoffend, not one or the other made him likely to reoffend. From this, the disjunctive error in the "to commit" instruction was foreseeable and obvious and therefore manifest. But the court gave the instruction anyway. The flawed instruction had practical and identifiable consequences in Love's trial because, if followed, its effect was

- 26 -

to permit commitment based on less proof than required and speculation rather than substantial evidence.

      c.   **Love's counsel provided ineffective assistance in failing to object to the "to commit" instruction.**

In the event the Court declines to review the claimed error on appeal in the absence of objection below, then it will be necessary to address whether Love's counsel provided ineffective assistance.

Criminal defendants are constitutionally guaranteed the right to the effective assistance of counsel. Strickland v. Washington, 466 U.S. 668, 685-86, 104 S. Ct. 2052, 80 L. Ed. 2d 674 (1984). Those facing involuntary commitment have a statutory and due process right to counsel and courts apply the Strickland standard to determine whether counsel was ineffective. Moore, 167 Wn.2d at 122; Jenkins v. Dir. of Virginia Ctr for Behavioral Rehab., 271 Va. 4, 16, 624 S.E.2d 453 (Va. 2006) (recognizing due process right under federal constitution); U.S. Const. amend. V and XIV; RCW 71.09.050(1); RCW 10.101.005. "A claim of ineffective assistance of counsel is an issue of constitutional magnitude that may be considered for the first time on appeal." State v. Kyllo, 166 Wn.2d 856, 862, 215 P.3d 177 (2009).

To establish ineffective assistance of counsel, Love must show deficient performance and resulting prejudice. Moore, 167 Wn.2d at 122.

Deficient performance is that which falls below an objective standard of reasonableness. Strickland, 466 U.S. at 688. The strong presumption that defense counsel's conduct is not deficient is overcome where there is no conceivable legitimate tactic explaining counsel's performance. State v. Reichenbach, 153 Wn.2d 126, 130, 101 P.3d 80 (2004).

There is no legitimate reason why defense counsel failed to object to the "to commit" instruction on the basis that the use of the disjunctive "or" lessened the State's burden of proof, allowed the jury to base its verdict on speculation, and was not supported by substantial evidence. The flawed "to commit" instruction made it easier for the State to prove and the jury to find Love met the SVP definition. No competent attorney makes it easier for his client to be civilly committed.

Prejudice results from a reasonable probability that the result would have been different but for counsel's performance. Strickland, 466 U.S. at 694. A reasonable probability is a probability sufficient to undermine confidence in the outcome. Id. Love shows prejudice because, as argued above, there was a basis for a reasonable jury to reject either the mental abnormality or the personality disorder as the condition that made Love likely to reoffend. There is a reasonable probability sufficient to undermine confidence in the outcome to conclude the jury, following the "to

commit" instruction, found Love to be an SVP based on one or the other but not both conditions.

3. THE COURT COMMITTED REVERSIBLE ERROR IN ADMITTING PRIOR WITNESS TESTIMONY BECAUSE THE STATE FAILED TO ESTABLISH THE WITNESS WAS UNAVAILABLE TO TESTIFY AT THE PRESENT TRIAL.

To admit a witness's prior testimony as an exception to the hearsay rule under ER 804, the proponent must establish unavailability, which means the use of reasonable means to secure the witness's attendance. The State made no effort to obtain the voluntary attendance of A.P. as a witness at trial. Having failed to establish A.P. was unavailable within the meaning of the rule, her former testimony was inadmissible. The court abused its discretion and committed reversible error in ruling otherwise.

The State sought the admission of A.P.'s former testimony under ER 804(b)(1). 1RP 1023. The defense objected, arguing the State had not shown A.P. was unavailable under ER 804 and that the testimony was cumulative and unnecessarily prejudicial. 1RP 1023.

The State contended A.P. was not amenable to subpoena because she lived in Puerto Rico and was therefore "unavailable" under ER 804(b)(1). 1RP 1024. The State further argued the former testimony was not cumulative because Dr. Phoenix's testimony regarding the A.P. rape was admitted for the limited purpose of explaining the basis for her expert

opinion. 1RP 1024. A.P.'s former testimony, in contrast, was offered for substantive purposes. 1RP 1024. The assistant attorney general pleaded "I need to be able to argue in the end as part of my theory of the case that part and parcel of Mr. Love's mental state is his refusal to acknowledge acts that I think somebody at some point in the trial said his version doesn't survive even a cursory examination." 1RP 1024-25. A.P.'s former testimony should be allowed for substantive purposes because "there's room for difference of opinion about the -- about what happened during those incidents." 1RP 1025.

The defense pointed out this was the first time it had heard the State's claim that A.P. was unavailable and "it appears they haven't even tried to get her here." 1RP 1025. The trial court admitted A.P.'s former testimony because it was part of the State's case-in-chief, accepting the State's representation that A.P. was unavailable. 1RP 1026-27. A.P.'s former testimony was read to the jury. 1RP 1199-1201; CP 32-48.

A trial court's finding of unavailability is reviewed under the abuse of discretion standard. Kinsman v. Englander, 140 Wn. App. 835, 840, 167 P.3d 622 (2007). A trial court abuses its discretion when it applies the wrong legal standard, bases its ruling on an erroneous view of the law, or otherwise fails to adhere to the requirements of an evidentiary rule. State

v. Quismundo, 164 Wn.2d 499, 504, 192 P.3d 342 (2008); State v. Foxhoven, 161 Wn.2d 168, 174, 163 P.3d 786 (2007).

ER 804(a)(5) defines a witness as unavailable if the witness "[i]s absent from the hearing and the proponent of his statement has been unable to procure his attendance . . . by process or other reasonable means." "If a witness is found unavailable under this test, the witness' former testimony may be admitted into evidence under ER 804(b)(1)." Rice v. Janovich, 109 Wn.2d 48, 57, 742 P.2d 1230 (1987).

The State argued A.P. was unavailable because she could not be subpoenaed. 1RP 1024. But the inability to reach a witness by subpoena power is insufficient to establish unavailability. Rice, 109 Wn.2d at 57. The party calling the witness must also establish an inability to reach the witness by "other reasonable means." Young v. Key Pharm., Inc., 63 Wn. App. 427, 432, 819 P.2d 814, review denied, 118 Wn.2d 1023, 827 P.2d 1392 (1991). "The party offering the out-of-court statement of a witness beyond the legal reach of a subpoena should at least be required to represent to the court that it made an effort to secure the voluntary attendance of the witnesses at trial." Rice, 109 Wn.2d at 57.

The State, by representing A.P. lived in Puerto Rico, demonstrated that she was not subject to a subpoena to secure her presence. But the State

made no effort to show A.P. refused to voluntarily provide testimony.[16] The record gives no indication that the State made *any* effort to obtain the voluntary attendance of A.P. as a witness at trial. Since no showing of unavailability was made, the trial court erred in admitting A.P.'s former testimony. Rice, 109 Wn.2d at 58. The court abused its discretion in failing to adhere to the "unavailability" requirement of ER 804(a)(5).

The Supreme Court has flatly stated "[t]he admission of evidence without a proper showing of unavailability of the witness is reversible error." Rice, 109 Wn.2d at 58. It has also been stated that an evidentiary error is prejudicial where, within reasonable probabilities, the outcome of the trial would have been materially affected had the error not occurred. State v. Sanchez, 42 Wn. App. 225, 231, 711 P.2d 1029 (1985); State v. Tharp, 96 Wn.2d 591, 599, 637 P.2d 961, 965 (1981).

The admission of A.P.'s former testimony is reversible error. The admission of A.P.'s testimony as substantive evidence bolstered Dr. Phenix's expert opinion. Dr. Phenix's testimony on the A.P. rape was not substantive evidence; it was admitted as the basis for her expert opinion. 1RP 875-76; CP 14. The admission of facts forming the basis for an expert's opinion is not proof of them. Group Health Co-op. of Puget Sound,

---

[16] A.P. testified in person at Love's 2005 commitment trial, and she was flown from Puerto Rico to do so. 1RP 1026-27.

Inc. v. Dep't of Revenue, 106 Wn.2d 391, 399-400, 722 P.2d 787 (1986).

Under ER 703,[17] an otherwise inadmissible fact underlying an expert's opinion is admissible for the limited purpose of explaining the basis for an expert's opinion, but that underlying fact is not substantive evidence. Allen v. Asbestos Corp., Ltd., 138 Wn. App. 564, 579, 157 P.3d 406 (2007), review denied, 162 Wn.2d 1022, 178 P.3d 1033 (2008).

Dr. Phenix relied on the A.P. rape as an important basis of her opinion that Love harbored deviant sexual arousal. 1RP 885-86. The probative force of her expert testimony hinges on the accuracy of the bases for her opinion: if a basis is false or mistaken, then the expert's opinion has diminished value. Without substantive evidence regarding A.P. to back up Dr. Phenix's opinion, the jury may have been more inclined to discount the persuasive force of her expert testimony.

Further, the State argued A.P.'s former testimony, offered as substantive evidence, was important to rebut Love's version of events. 1RP 1024-25. In doing so, it implicitly acknowledged that Dr. Phenix's rendition was not up to the task precisely because her testimony was not substantive evidence. The State needed A.P.'s testimony admitted as

[17] ER 703 provides: "The facts or data in the particular case upon which an expert bases an opinion or inference may be those perceived by or made known to the expert at or before the hearing. If of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject, the facts or data need not be admissible in evidence."

substantive evidence to counter Love's denial of the rape and his account

of his relationship with the woman. 1RP 1470-85, 1505, 1599. The jury

was erroneously allowed to consider A.P.'s former testimony as

substantive evidence, i.e., as factual proof of the event. Her testimony, in

all its troubling detail, was sure to leave a lasting impression on jurors as it

considered Love's dangerousness. CP 32-48. The jury should never have

been allowed to hear it and consider it as substantive proof of the event.

4. THE COURT WRONGLY ADMITTED EXPERT TESTIMONY ON RISK ASSESSMENT UNDER THE FRYE STANDARD.

Expert testimony on the SRA-FV was inadmissible under the Frye

standard because the State failed to prove the method used to assess risk

based on the presence of dynamic risk factors was accepted in the

scientific community. Reversal is required because there is a reasonable

probability that, but for the error, the outcome would have been different.

a. Summary of the novel dynamic risk assessment known as the SRA-FV.

The SRA-FV is a "novel dynamic risk assessment instrument." In

re Detention of Ritter, 177 Wn. App. 519, 525, 312 P.3d 723 (2013),

review denied, 180 Wn.2d 1028 (2014). "[W]here an expert witness

derives a prediction of future dangerousness in whole or part from a novel

dynamic risk assessment instrument like the SRA-FV, the trial court must

hold a Frye hearing on the instrument before the expert may use it at trial."

Ritter, 177 Wn. App. at 525.

The defense moved to exclude expert testimony on the SRA-FV under Frye. CP 425-87. At the Frye hearing, Dr. Phenix testified for the State. 1RP 519-612. Dr. Donaldson testified for the defense. 1RP 612-658. At the conclusion of the hearing, the court ruled the Frye standard was satisfied. 1RP 672-74; CP 2-5.

To address that ruling, a summary of risk assessment is in order. "In greatly simplified terms, there are two broad approaches to conducting risk assessments: clinical judgment or actuarial assessment." In re Detention of Thorell, 149 Wn.2d 724, 753, 72 P.3d 708, 720 (2003). Risk factors are either static, which are unchangeable, or dynamic, which are changeable; dynamic risk factors are either stable, which can change slowly, or acute, which can change quickly. Ritter, 177 Wn. App. at 523 n.4.

An actuarial instrument like the Static-99R measures the presence of static risk factors. 1RP 566-68. The SRA-FV, on the other hand, is a structured clinical judgment tool for evaluating "stable dynamic risk factors" and integrating them with "static risk factors" considered by actuarial instruments. Id. at 523. "Thus, a prediction of future

- 35 -

dangerousness based on the SRA-FV is neither purely actuarial nor purely clinical." Id. at 523.

The SRA-FV considers three domains of stable dynamic risk factors: "Sexual Interests," "Relational Style," and "Self-Management." The sexual interests domain includes "Sexual preferences for children," "Sexualized violence," and "Sexual preoccupation." The relational style domain includes "Emotional congruence with children," "Lack of emotionally intimate relationships [with adults]," "Callousness," and "Grievance thinking." The self-management domain includes "Lifestyle impulsivity," "Resistance to rules [and] supervision," and "Dysfunctional coping." Id. at 523 n.4.

According to Dr. Phenix, "the items [in the SRA-FV] are not statistically weighted for their contribution to risk" and "there can be no probabilities of sexual re-offense derived from this particular instrument." 1RP 535. The instrument helps evaluators determine the presence of dynamic risk factors. 1RP 536. The SRA-FV score places the subject in a low, medium or high risk category, which "guides you to the appropriate probabilities of sexual re-offense on Static-99R. So basically, the two instruments work together to determine those overall probabilities of sexual reconviction." 1RP 538-39.

- 36 -

Total scores on the SRA-FV range from zero to six. The total

score on the SRA-FV is "associated with a guide that tells you which of

the three groups to pick. So there are statistical calculations that looked at

[the] Static-99 score, it looked at recidivism rates, and it looked at the

presence of dynamic needs." 1RP 543. The total score on the SRA-FV is

used to select one Static-99R "reference group" among three available

options: routine, preselected for treatment, and high risk. 1RP 542, 570,

600-01. From those, the SRA-FV developers created a Level Of Needs

Index (LONI), which is a table that directs the evaluator to the appropriate

Static-99R norm. 1RP 543, 582-83. In that manner, the risk of reoffense

is quantified. 1RP 598.

Phenix scored Love as a 4.45 on the SRA-FV. 1RP 582. She

plugged Love's score into the LONI to place him in the Static-99R high

risk needs norm. 1RP 582-83, 585-86.

        b.   **The scientific evidence was inadmissible under <u>Frye</u> because the method employed by the State's expert to conduct the risk assessment had not achieved consensus in the relevant scientific community.**

The trial court determined (1) "The SRA-FV is generally accepted

within the community of experts who evaluate sex offenders and assess

their recidivism risk" (CP 3 (FF 9)); (2) "The use of structured analysis of

risk factors in sex offender evaluations is supported by a scientific theory

that is generally accepted in the scientific community" (CP 4 (CL 3)); and "The SRA-FV is an instrument that is capable of producing reliable results and is generally accepted in the scientific community." CP 4 (CL 6).

As set forth below, Love challenges these determinations, none of which are entitled to deference on review. Frye determinations are reviewed de novo. Anderson v. Akzo Nobel Coatings, Inc., 172 Wn.2d 593, 600, 260 P.3d 857 (2011). A reviewing court will undertake a searching review that is not confined to the trial record. State v. Copeland, 130 Wn.2d 244, 255-56, 922 P.2d 1304 (1996).

Under Frye, novel scientific evidence is admissible only where (1) the scientific theory or principle upon which the evidence is based has gained general acceptance in the relevant scientific community of which it is a part; and (2) there are generally accepted methods of applying the theory or principle in a manner capable of producing reliable results. State v. Riker, 123 Wn.2d 351, 359, 869 P.2d 43 (1994). Both the scientific theory underlying the evidence and the technique or methodology used to implement it must be generally accepted in the scientific community for evidence to be admissible under Frye. State v. Gore, 143 Wn.2d 288, 302, 21 P.3d 262 (2001). While unanimity is not required, scientific evidence is inadmissible "[i]f there is a significant dispute among qualified scientists in the relevant scientific community." Gore, 143 Wn.2d at 302.

The State, as proponent of the challenged expert testimony, bore the burden of showing the Frye standard is met. In re Marriage of Parker, 91 Wn. App. 219, 226, 957 P.2d 256 (1998). The State failed to establish expert consensus that the SRA-FV method of risk assessment is reliable.

The SRA-FV was published in a peer reviewed professional journal on December 30, 2013. David Thornton & Raymond A. Knight, Construction and Validation of the SRA-FV Need Assessment, Sexual Abuse: A Journal of Research and Treatment (December 30, 2013); see CP 701-17 (article in clerk's papers). The developers of the SRA-FV authored this publication. They claimed the SRA-FV scores are statistically correlated with sexual recidivism, and that the SRA-FV has shown significant incremental validity in improving risk assessment relative to the Static-99R. Thornton & Knight (2013) at 1, 9-12; CP 702, 710-13. Phenix testified to the same effect. 1RP 546-47. Donaldson testified the instrument adds some predictive validity to the Static-99R, but is used in a way that says nothing about the accuracy of the predictions it purports to make. 1RP 632-33.

Phenix acknowledged the 2013 Thornton article does not address the use of the SRA-FV to determine which recidivism tables from the Static-99R to use. 1RP 589-90. It addressed the development of the instrument, "but not the use of the total score to choose the norms for Static-99R."

1RP 587. There is nothing published or peer reviewed on using the SRA-FV score to choose a Static-99R norm as a means to measure risk of reoffense.[18] 1RP 587-88.

Phenix said Thornton and Hanson presented data at the 2012 Association for Treatment of Sexual Abusers (ATSA) conference and validated the SRA-FV as a tool to select a Static-99R reference group based on the SRA-FV score. 1RP 584, 587. Neither the presentation nor the data has been published. 1RP 584-85.

Phenix claimed the SRA-FV cut off scores were a legitimate way to choose the Static-99R norm because research established the risk of reoffense increases when offenders have more dynamic risk factors. 1RP 602. She said there was an "independent replication" of the appropriate cut-offs at the 2012 ATSA conference. 1RP 602. But the presenters at this conference were Thornton and Hanson, one of whom was the developer of the SRA-FV and author of the 2013 article. 1RP 591. Thornton and Hanson's work on the SRA-FV has not been peer reviewed, published or replicated. 1RP 591-92.

There is a peer-reviewed publication, authored by someone who did not develop the SRA-FV, which addresses the validity of using the SRA-FV to choose Static-99R recidivism estimates: Brian Abbott, The

---

[18] The cut off scores are found in the Evaluator's Handbook. 1RP 604.

Utility of Assessing "External Risk Factors" When Selecting Static 99R Reference Groups, Open Access Journal of Forensic Psychology 5, 58-118 (2013) (attached as app. A). Abbott concluded such a use is scientifically unjustified and leads to erroneous results. Id. at 104. Abbott discovered "clinicians cannot rely upon the evaluee's total dynamic risk score to select a single Static-99R reference group." Id. at 99. Donaldson agreed with Abbott that the groups could not be separated in the way envisioned by the SRA-FV and that the instrument was not ready for forensic use in the courtroom. 1RP 635-36, 640.

According to Abbott, those who use cut-off scores on the SRA-FV to choose which recidivism estimates to use for the Static-99R assume that the members of the different Static-99R recidivism groups (high risk group, pre-selected treatment group, routine group) have a distinct and exclusive range of scores on the SRA-FV. Id. at 97, 102. For example, Thornton, in unpublished material, teaches evaluators to do the following: if the evaluee scores a 3.3 or higher on the SRA-FV, then use the high risk recidivism rates for the Static 99R; if the evaluee scores between 2.4 and 3.2 on the SRA-FV, then use the preselected for treatment group recidivism rates on the Static-99R; and if the evaluee scores a 2.3 or below on the SRA-FV, then use the routine recidivism rates on the Static-99R. Id. at 93-94 (Table 2) and 99-100 (Table 4).

The validity of this procedure assumes all of the sex offenders in the Static-99R high risk reference group would have scored a 3.3 or higher on the SRA-FV; that all members of the preselected for treatment group would have scored between a 2.4 and a 3.2, and that all members of the routine sample would have scored a 2.3 or lower on the SRA-FV. Id. But Thornton recommended using cut-off scores on the SRA-FV to choose the Static-99R reference group without ever actually scoring the SRA-FV on each member of the respective Static-99R recidivism groups. Id. at 95. Instead, Thornton only scored a single sample of the preselected high risk need group. He then used that date to "statistically contrive" a Static-99R reference group selection model. Id.

According to Abbott, Thornton's research suffers from a fatal flaw. If evaluators are to use a risk assessment instrument to select Static-99R reference groups, there must be three ranges of scores that are mutually exclusive, one for each Static-99R reference group. Id. at 93-94. This is not the case when evaluators use the SRA-FV to choose Static-99R reference groups. Abbott analyzed the raw data from the Static-99R developers and found that the members of the different Static-99R recidivism groups — high risk, preselected, and routine — had a variation of SRA-FV scores and those SRA-FV scores overlapped all three of Thornton's proposed cut scores. Id. at 97-100 (Tables 3 and 4).

Phenix believed the overlap, if it exists at all, was negligent or small. 1RP 585. So there is a disagreement among the experts. But the State did not meet its burden of showing a lack of significant dispute among experts that the SRA-FV was a reliable method of doing what it claims to do. The court's task is not to determine whether a scientific method is correct because such determination is beyond the expertise of judges. State v. Sipin, 130 Wn. App. 403, 419, 123 P.3d 862 (2005). Instead, its task is to determine whether the appropriate scientific community has generally reached consensus that the method is reliable. Sipin, 130 Wn. App. at 419-20.

Thornton describes the SRA-FV as a "newly designed instrument." Thornton & Knight (2013) at 1; CP 702. Thornton could only hypothesize that their results would generalize to other sex offenders: "it seems reasonable to hypothesize that the present results will generalize to a similar range of settings. Definitive evidence about this will, however, depend on new studies carried out with other samples." Thornton & Knight (2013) at 12; CP 713.

The SRA-FV was validated on a split sample of offenders (a sample taken from the same original Bridgewater population on which the SRA-FV was originally developed). 1RP 544-46, 589. Phenix testified a split sample validation is a common and accepted way to validate an

instrument. 1RP 528. Donaldson disagreed. 1RP 619-22, 652-53. The State did not meet its burden of showing a lack of significant dispute among experts that use of a single split sample is sufficient to validate a risk assessment instrument. The trial court erred in concluding otherwise. CP 4 (CL 5).

The SRA-FV has not been cross-validated on an independent sample (a sample of offenders taken from a different population). This is significant. Thornton, the developer of the SRA-FV, recognized "the present study has a number of limitations that must be addressed in future research. First, as we have noted, because the present results are limited to a particular population, cross validation of the scale on other populations is essential." Thornton & Knight (2013) at 14; CP 715.

If the split sample is sufficient to show the reliability of the method, as Phenix contends, then why do the developers of the SRA-FV concede that cross validation on new samples is essential? Thornton does not spell it out. But "[i]t is well known that predictive validity tends to be stronger in initial validation studies than in cross validation studies, a pattern often referred to as shrinkage. Shrinkage occurs because prediction equations capitalize on chance characteristics of the validation sample to achieve optimal prediction, and these same characteristics are not likely to be present to the same degree in a new sample." Blair,

Marcus & Boccaccini, Is There Allegiance Effect for Assessment Instruments? Actuarial Risk Assessment as an Exemplar, Clinical Psychology: Science and Practice, Vol. 15 Issue 4 at 349 (Dec. 2008).[19] Blair studied three actuarial tools used in SVP proceedings (SORAG, VRAG, and Static 99) and found the predictive value for each instrument was highest in the initial validation studies (conducted by the developer of the instrument). Id. The value decreased in cross validations studies by the developers of the instruments, and further decreased in cross validation studies by independent researchers. Id.

One reason for this bias is that instrument authors may be unwilling to publish studies showing poor performance of their instruments. Id. Other researchers have discovered similar results. Singh, Grann and Fazel found evidence of a significant authorship bias specifically to risk assessment studies published in peer reviewed journals. Singh, Grann & Fazel, Authorship Bias in Violence Risk Assessment? A Systematic Review and Meta Analysis, PLOS ONE, Vol. 8 Issue 9 (Sept. 2013). Such concerns illustrate the problem of treating the SRA-FV assessment as a reliable method accepted in the scientific community when it is still so new.

---

[19] Phenix knows when an instrument is developed on a certain sample of sex offenders, its predictive accuracy is maximized for that group, and lessens when later applied to other sample groups. 1RP 525. Phenix recognized it is important to test an instrument on a different group of sex offenders — a new sample — to see if the initial predictive accuracy holds up. 1RP 525.

The inter rater reliability of the SRA-FV is another concern in the scientific community. Inter rater reliability is the degree to which experts will arrive at the same score when they apply the same instrument to the same offender based on the same available information. When an instrument lacks inter rater reliability, it is an unreliable measure of risk because one cannot be sure of the subject's actual score on the instrument. "[T]he lower the reliability of a given test, the lower the limit on the validity of the construct being measured. It should thus be no surprise that tests with reliability coefficients below .80 have been criticized for containing excessive error variance and, hence, poorer validity." Kirk Heilbrun, The Role of Psychological Testing in Forensic Assessment, Law and Human Behavior, vol. 16 No. 3 at 265 (1992).

There have been limited studies on the SRA-FV's inter rater reliability. 1RP 551-52. There are no published and peer-review studies examining the inter rate reliability for each of the scored items in the SRA-FV. 1RP 592. One of the authors of the 2013 Thornton & Knight article trained, supervised and consulted with individuals who scored the SRA-FV. Thornton & Knight (2013) at 8; CP 709. Even with these added safeguards to ensure reliability, the SRA-FV had low reliability: a .64 rating for a single rater working alone and .78 for two raters working together. Thornton & Knight (2013) at 9; CP 710. .80 is the standard for use in forensic

evaluations. Abbott (2013) at 96 (citing Heilbrun (1992)). Thornton acknowledged "The results of the study do raise a particular concern about the SRA-FV. The observed inter rater reliability was lower than desirable." Thornton & Knight (2013) at 13; CP 714. In unpublished research, the SRA-FV could only muster a .55. rating. Abbott (2013) at 96.

Phenix said people differ on what is a minimal level of inter rater reliability but acknowledged the inter rater reliability was lower than she would like to see. 1RP 552, 593. She expected reliability to improve in time and opined that, even with low reliability, predictive accuracy was still acceptable. 1RP 1RP 552-53, 593. According to Phenix, the flaw is in the raters, not the instrument. 1RP 553-54. Abbott believes the flaw is in the instrument because the rating criteria are inherently subjective. Abbott (2013) at 95. Donaldson opined a .64 rating is acceptable to begin developing an instrument but unacceptable for scoring accuracy. 1RP 630, 648-49.

"The core concern of Frye is whether the evidence being offered is based on an established scientific methodology." State v. Russell, 125 Wn.2d 24, 41, 882 P.2d 747 (1994). For the reasons stated, the State failed to show the SRA-FV method of risk assessment meets that standard.

In February 2011, California adopted the SRA-FV as its official dynamic risk assessment instrument for evaluating sex offenders' future

dangerousness. Ritter, 177 Wn. App. at 524. But in September 2013, California switched to the Stable-2007/Acute-2007 instrument. Id.[20] Phenix said the SRA-FV was used by all of the federal Adam Walsh evaluators in California.[21] 1RP 588. She said many colleagues used the SRA-FV. 1RP 588-89.

Donaldson testified there is a difference between the science being used and the science being accepted. 1RP 628. Donaldson knew there was "some significant group of people" using the SRA-FV based on his review of other SVP evaluations by state evaluators. 1RP 629. These evaluators use it because they are instructed to use it, but that did not mean the instrument met the scientific standards for acceptability. 1RP 629.

As argued, there is still a significant debate that this new instrument employs a reliable methodology to predict risk of reoffense. Scientific evidence is inadmissible "[i]f there is a significant dispute among qualified scientists in the relevant scientific community." Gore, 143 Wn.2d at 302. The SRA-FV is still in its experimental phase. It is not ready for use in the courtroom.

_____

[20] Phenix claimed the Stable-2007 instrument replaced the SRA-FV because it was targeted to the California population at issue (parolees/probationers), had a larger validation, and showed incremental validity. 1RP 556-57.
[21] Under the Adam Walsh Child Safety and Protection Act, the federal government may seek the civil commitment of certain individuals determined to be a "sexually dangerous person." 18 U.S.C. § 4248.

c.   **The error is prejudicial because it impacted a material and disputed issue in the case.**

Reversal is required when there is a reasonable probability that, but for the Frye error, the outcome of his trial would have been different. Sipin, 130 Wn. App. at 421. Improper admission of evidence constitutes harmless error only if the evidence is of minor significance in reference to the evidence as a whole. State v. Neal, 144 Wn.2d 600, 611, 30 P.3d 1255 (2001). Expert testimony on the SRA-FV cannot be considered of minor significance in Love's case.

The two sides presented dueling expert opinion on whether Love was likely to reoffend. Dr. Phenix relied on the SRA-FV as a major part of her risk assessment involving dynamic risk factors for the jury. 1RP 935-940. Instead of a pure clinical evaluation of dynamic risk factors, which Phenix described as less reliable (1RP 916-17), the State was able to impress the jury with a mathematical calculation of risk involving those factors. Indeed, the State argued to the jury that the SRA-FV was the "state of the art" in risk assessment. 1RP 1962-65. The danger is that the jury took the same view and placed particular weight on this risk assessment tool when the jury should not have been allowed to consider it as evidence.

The Frye standard prevents "pseudoscience" from entering the courtroom. Copeland, 130 Wn.2d at 259. It prevents jurors from being

misled by unreliable evidence and "awed by the apparent infallibility of scientific experts and their techniques." State v. Brewczynski, 173 Wn. App. 541, 556, 294 P.3d 825, review denied. 177 Wn.2d 1026, 309 P.3d 505 (2013). Reversal is required because the outcome of the trial might reasonably have been different if the trial court had excluded the challenged evidence. Sipin, 130 Wn. App. at 421.

D.   CONCLUSION

For the reasons stated, Love requests that this Court vacate the jury's verdict and reverse the court's commitment order.

DATED this 26th day of January 2015.

Respectfully submitted

NIELSEN, BROMAN & KOCH, PLLC
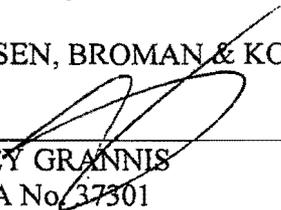
CASEY GRANNIS
WSBA No. 37301
Office ID No. 91051
Attorneys for Appellant

- 50 -

APPENDIX A

The Utility of Assessing "External Risk Factors" When Selecting Static-99R Reference Groups

Author: Brian R. Abbott, Ph.D., Independent Practice, San Jose, CA, USA.
Email: brian@dr-abbott.net

Abstract

The Static-99 has been one of the most widely used sexual recidivism actuarial instruments. It has been nearly four years since the revised instrument, the Static-99R, has been released for use. Peer-reviewed literature has been published regarding the basis for changing the scoring system for the age-at-release item, the utility of relative risk data, and variability of sexual recidivism rates across samples. Thus far, the peer-reviewed literature about the Static-99R has not adequately addressed the reliability and validity of the system to select among four possible actuarial samples (reference groups) from which to obtain score-wise observed and predicted sexual recidivism rates to apply to the individual being assessed. Rather, users have been relying upon the Static-99R developers to obtain this information through a website and workshops. This article provides a critical analysis of the reliability and validity of using the level of density of risk factors external to the Static-99R to select a single reference group among three options and discusses its implications in clinical and forensic practice. The use of alternate methods to select Static-99R reference groups is explored.

Keywords: Static-99R; risk assessment; sex offenders; prediction; recidivism

The initial release of the Static-99R (Hanson & Thornton, 2000) provided a single sexual recidivism experience table for users to consult to obtain score-wise risk estimates to compare to the individual being assessed. The developers of the Static-99 ("developers") released multiple experience tables in 2008, ostensibly in an effort to better account for the significant variability in base rates observed across different samples of sexual offenders (Harris, Helmus, Hanson, & Thornton, 2008). The developers instructed users to select one among three experience tables ("reference groups") but this procedure was later modified to report a range of risk bounded by the sexual recidivism rates from two reference groups (Abbott, 2009). These procedures were abandoned one year later with the release of the Static-99R (Hanson, Phenix, & Helmus, 2009) and users were advised to follow a specific method by which to select a single reference group among four options (Phenix, Helmus, & Hanson, 2009).

In 2009, Phenix et al. (2009) instructed Static-99R users to select reference groups employing a method referred to as *cohort matching* where clinicians consider broadbrushed descriptions of the four Static-99R reference groups, devised by the developers without empirical validation, and attempt to match the characteristics from a single reference group that is most similar to the group representing the individual being assessed. The reliability and validity of the cohort-matching process has not been established and Wollert (2010) reported how classification error (the probability of

selecting an erroneous reference group to compare to the individual being assessed) reduces the accuracy of the observed sexual recidivism rates. In a July 2012 revision of the Static-99R Evaluators' Handbook, Phenix, Helmus, & Hanson (2012) state the preferred method for selecting reference groups is to consider the density of risk factors external to the Static-99R. Studies of this procedure have not been conducted to test its reliability and validity. Phenix et al. (2012) constructed the reference-group-selection procedures based on two major untested hypotheses.

One, the developers assume, through post-hoc logical inference, that risk factors external to the Static-99R account for the differences in sexual recidivism base rates between reference groups (Phenix et al., 2012; Thornton, Hanson, & Helmus, 2010; Helmus, 2009). This assertion has never been tested empirically with the 20 samples comprising the three primary Static-99R reference groups; nor have the external risk factors believed to cause this differentiation ever been identified. In other words, it is unknown whether the density of risk factors external to the Static-99R is a valid conceptualization for explaining the differences in sexual recidivism base rates between the reference groups. Moreover, the developers have failed to operationalize the term "density," so it is unknown how to define, quantify, or measure it. In essence, the developers have introduced reference-group-selection procedures devised on a premise with unknown validity and unproven reliability. For the purposes of the following analysis, however, the author presumes that the density of risk factors external to the Static-99R account for discrimination of the reference groups.

Two, the current selection procedures (Phenix et al., 2012) are premised on the idea that the sexual recidivism base rate contained in the experience table that a user chooses is similar to that of the group representing the evaluee. The developers have not produced data supporting this assumption. The fit in sexual recidivism base rate between the two groups is essential in producing accurate score-wise risk estimates to compare to the individual being assessed (Donaldson & Wollert, 2008). Consequently, users who employ the current reference-group-selection procedures may unwittingly select an experience table that results in the inaccurate estimation of risk at the score assigned to the evaluee.

Since the inception of the Static-99R reference-group-selection procedure (Phenix et al., 2009), the developers have published peer-reviewed articles documenting the change in scoring the age-at-release item (Helmus, Thornton, Hanson, & Babchishin, 2012), the reporting of relative risk data (Hanson, Lloyd, Helmus, & Thornton, 2012), and the variability in base rates and score-wise risk estimates across the various samples comprising the Static-99R data set (Helmus, Hanson, Thornton, Babchishin, & Harris, 2012). In July 2012, the developers released a revised version of the Evaluators' Handbook (Phenix et al., 2012) that provides the current recommended procedures to select Static-99R reference groups. To date, the developers have not produced data informally or through peer-reviewed publication establishing the reliability and validity of the Static-99R reference-group-selection system, yet many clinicians have been using it for nearly four years as if its reliability and validity have been proven.

As readers will see, the following analysis of the Static-99R reference-group-selection procedure is based largely on unpublished material produced by the developers. Without peer-reviewed publication, Static-99R users must obtain information to implement the reference-group-selection procedures from disparate sources, including the Static-99 website and trainings conducted by the developers. In justifying the application of the reference-group-selection system in forensic and clinical practice, users are expected to rely on the developers' assurances that the hypothetical, empirically untested reference-group-selection system is reliable, works as intended, and produces accurate risk estimates for the individual being assessed.

The overarching aim of this article is to fill the gap in peer-reviewed literature about the Static-99R reference-group-selection method so that practitioners have a frame of reference to assess the reliability and validity of the procedures as applied in sexual recidivism risk assessments. Since Wollert (2010) has addressed the reliability and validity issues associated with the cohort-matching system, this article will focus on the preferred reference-group-selection procedure that assesses the density of risk factors external to the Static-99R. First, this article briefly describes background information about the statistical methods the developers employed to discriminate among the Static-99R reference groups. Second, the author examines the reliability and validity of using risk factors external to the Static-99R to select reference groups. Finally, the Discussion and Conclusions section explores alternative methods to select Static-99R reference groups.

Evolution of the Static-99R Reference Groups

Based on concerns raised by Abbott (2009), Helmus (2009) analyzed the possible influence of moderator variables (e.g., sample type, offender type, age at release, country, or era of study) effecting the discrimination of Static-99R reference groups. The meta-analyses of sample type (e.g., routine vs. preselected) involved eight iterations that produced a final solution (Preselected Version 3), as reported by Helmus (2009), Table 15, where observed differences in base rates of sexual recidivism distinguished three reference groups designated as Routine Corrections ("RC"), Preselected Treatment Need ("PTN"), and Preselected High Risk Needs ("PHRN"). The Non-Routine Corrections ("NRC") reference group is a compilation of the samples comprising the PTN and PHRN reference groups and three other studies (Phenix et al., 2012) and is considered only when applying the cohort-matching selection system.

After observing the statistically significant base rates of sexual recidivism across the three reference groups, the developers engaged in post-hoc, logical inference to explain the characteristics they hypothesized as distinguishing the three reference groups (Helmus, 2009). The developers assert that certain judicial, correctional, or administrative decisions were made that led offenders to be placed in different settings such as general prison populations, community supervision, outpatient or correctional treatment programs, and secure facilities for high-risk sex offenders. The developers posit that this decision-making process relied on the consideration of the density of risk factors external to the Static-99R and they characterize this as the "preselection effect."

Phenix et al. (2012; p. 35) state that the magnitude of the preselection effect (i.e., low, moderate, or high) is the basis by which to discriminate the RC reference group from the PTN and PHRN reference groups, as well as distinguishing differences across the two non-routine reference groups (PTN and PHRN). This is considered the preferred reference-group-selection method with the cohort-matching procedure serving as an alternative. The astute reader may recognize that the Evaluators' Handbook contains internally inconsistent instructions on reference-group-selection methods. Phenix et al. (2012: p. 32) also advise users who are considering using routine samples to select the RC reference group when a Static-99R experience table devised on a local sexual offender group is not available. This recommendation presumes that the RC reference group is representative of the typical sexual offender from the routine corrections population. If the eight studies comprising the RC reference group were sampled representatively from a larger predefined routine corrections population, then it would be assumed that the resulting sexual-recidivism base rate and score-wise risk estimates would generalize to other samples drawn from this population within a certain margin of error. Data from Helmus (2009) appear to contradict this assumption.

Table 1
Static-99R RC Reference Group:
Descriptive Data for 8 Studies

Routine Corrections	5-Year Base Rate %	n	n_r^+
Bartosh et al. (2003)	13.3	90	12
Bigras (2007)	9.2	207	19
Boer (2003)	3.7	299	11
Craissati et al. (2008)	7.5	200	15
Eher et al. (2009)	2.0	151	3
Epperson (2003)	10.6	151	16
Hanson et al. (2007)	0	31	0
Langström (2004)	5.4	1,278	69
Total	6.0	2,407	145

n_r = number of recidivists

The RC reference group is comprised of eight studies as listed in Table 1. The Bartosh, Garby, Lewis, & Gray (2003) and Epperson (2003) studies are from the United States with the remaining samples coming from Canada, England, Austria, and Sweden. Using meta-analytic techniques, Helmus (2009) determined the sexual recidivism base rates among the eight samples varied more than would be expected by chance. This finding reflected that the samples comprising the RC reference group unlikely represented the same population of sexual offenders. After removing the two United States samples, Helmus (2009) determined the base rates for the remaining six samples appeared to vary only by chance. For reasons not clearly explained, Helmus (2009) decided to include the two United States

samples in the RC reference group experience table. These circumstances raise practical concerns when clinicians decide to select the RC reference group based on the assumption that the evaluee best matches this group.

The disparity in base rates among samples from the United States and other countries means that clinicians lack confidence that the 6.0% base rate for the RC reference group accurately represents the recidivism potential for sexual offenders from routine corrections populations. The 6.0% sexual recidivism base rate and corresponding score-wise recidivism rates from the RC reference group may underestimate the risk potential of offenders from the United States and inflate the recidivism potential of sexual offenders from other countries. The variability in sexual reoffense rates observed among the RC reference group studies in Table 1 suggests that the base rates among routine correctional samples may vary in meaningful ways from the aggregate base rate of 6.0%. Clinicians should be careful in accepting the 6.0% base rate and the resulting score-wise risk estimates as being generalizable to offenders who are presumed members of the routine corrections population. As demonstrated by Donaldson and Wollert (2008), the score-wise risk estimates are a direct function of the base rate of sexual recidivism and score-wise likelihood ratios. Consequently, clinicians who rely on the score-wise risk estimates calculated from the 6.0% sexual recidivism base rate would report inaccurate score-wise risk estimates if the group representing the evaluee had a sexual recidivism base rate of 2.0% or 13.3%, as seen, for example, in two studies reported in Table 1. This issue is further explored in the Discussion and Conclusions section.

Using a Dynamic Risk-Assessment Instrument
For Selecting a Static-99R Reference Group

Phenix et al. (2012; p. 35) conceptualize the preselection effect as the density of risk factors external to the Static-99R and instruct users to quantify the density of external risk factors by applying a "dynamic risk-assessment scale." Phenix et al. (2012) further describe three levels of preselection corresponding to the Static-99R reference groups (as represented in parenthesis following each reference group): RC (low), PTN (moderate), and PHRN (high). Recommendations for selecting a specific dynamic-risk instrument to assess the preselection effect are not contained in the Evaluators' Handbook (Phenix et al., 2012); however, instructions for selecting a measure are found in other sources. Thornton, Hanson, and Helmus (2010) endorse the Stable-2007 (Hanson, Harris, Scott, & Helmus, 2007), the Forensic Structured Risk Assessment ("Forensic SRA;" Thornton & Knight, 2009), and the Violence Risk Scale- Sexual Offender ("VRS-SO;" Olver et al., 2007) to quantify the preselection effect, but they provide no guidance to users as to what scores from the instruments correspond to each preselection level. Thornton (2010) proposes the Structured Risk Assessment- Forensic Version ("SRA-FV"), a revision of the Forensic SRA, as a means to quantify the preselection effect for selecting Static-99R routine and non-routine reference groups.

If evaluators are to use a dynamic risk-assessment instrument to select a Static-99R reference group, there must be three ranges of scores that are mutually exclusive ("cut-

scores"), one for each Static-99R reference group (RC, PTN, and PHRN). Table 2 reports two Static-99R reference-group-selection models using cut-scores from the SRA-FV and Stable-2007.

In unpublished instructions, Thornton (2011) proposes a specific system for using an evaluee's score on the SRA-FV to select a Static-99R reference group. Although Thornton's (2011) proposal produces seven levels of SRA-FV scores, he instructs evaluators to use the evaluee's score to select one of three Static-99R reference groups: RC, PTN, or PHRN. Table 2, column 2, shows Thornton's (2011) recommendations for using an evaluee's SRA-FV score to choose a Static-99R reference group. The specific ranges of SRA-FV scores in Table 2 were devised by Thornton (2011).

While the developers have not proposed a reference-group-selection model for the Stable-2007, Phenix et al. (2012) describe three levels of preselection effect (low, moderate, and high) that can be quantified using dynamic-risk measures. Hanson and Harris (2008) provide Stable-2007 cut-scores designated as low, moderate, and high dynamic risk or need, as reported in Table 2, column 3. The reader should note that no one has proposed that the Stable-2007 mutually exclusive score ranges (0-3, 4-11, and $\geq$ 12) represent the ideal cut-off scores for choosing a Static-99R reference group. For the purposes of this analysis, however, the author assumes the Stable-2007 cut-off scores shown in Table 2, column 3 correspond to the three Static-99R reference groups: RC, PTN, and PHRN.

Table 2
Descriptive Statistics for Three Instruments Used to
Measure Static-99R Preselection Effects

	SRA-FV	Stable-2007	VRS-SO[c]
N of items	10	13	17
N of risk domains	3	5	3
Possible item points	0 - 2	0 - 2	0 - 3
Range of total score	0 - 6	0 - 26	0 - 52
Scores associated with preselection effect (Static-99R reference group):			
Low (RC)	≤ 2.3[a]	$0 - 3$[b]	N.R.[d]
Moderate (PTN)	$2.4 - 3.2$[a]	$4 - 11$[b]	N.R.[d]
High (PHRN)	≥ 3.3[a]	≥ 12[b]	N.R.[d]

[a] From Thornton (2011); [b] From Hanson & Harris (2008); [c] Pretreatment dynamic risk factor; [d] Not reported by instrument developer

For the VRS-SO pretreatment dynamic-risk factor, Olver, Wong, Nicholaichuk, & Gordon (2013) report scores in four mutually exclusive score bands. No one has proposed a procedure by which to segregate the VRS-SO pretreatment dynamic risk total scores

into three mutually exclusive score bands to select among the three Static-99R reference groups.

It is not surprising that the Evaluators' Handbook (Phenix et al., 2012) lacks specificity regarding the selection of instruments to assess risk factors external to the Static-99R because research in this area is limited. At the time of the release of the Evaluators' Handbook (Phenix et al., 2012), the VRS-SO had not been administered to any of the 20 studies comprising the Static-99R reference groups. The Stable-2007 had been administered to 31 participants from a single study within the RC reference group (Hanson et al., 2007). The SRA-FV was devised on the Bridgewater sample (Knight & Thornton, 2007), which is one among six studies comprising the PHRN reference group. Later, Thornton (2011, 2010) used the Bridgewater data to statistically contrive a Static-99R reference-group-selection model as reflected in Table 2, column 2. Readers should note that Thornton (2010) did not administer the SRA-FV to individuals from any of the other 19 samples comprising the three Static-99R reference groups.

Reliability of Quantifying the Preselection Effect

In completing the Stable-2007, SRA-FV, and VRS-SO, raters assign a specific numerical value to each item according to the standardized rating or coding guidelines provided by each instrument developer (Cf., **Stable-2007**: Hanson & Harris, 2008; **SRA-FV**: Thornton, 2012 & 2011; **VRS-SO**: Wong, Olver, Nicholaichuk, & Gordon, 2003; Olver et al., 2013). The item scores are summed to obtain a total score. Determining the selection of an appropriate Static-99R reference group is therefore dependent on raters reliably ascertaining total instrument scores that fall within the critical region defining each level of the preselection effect.

It is apparent from reading the three instrument manuals that the rating criteria for items are inherently subjective which, in turn, is likely to result in users inconsistently applying the rating instructions. The score for each risk factor is determined by the fit of the evaluee's behavior to the scoring guidelines. In arriving at this decision, users must consider a wide array of complex behavioral patterns displayed by an evaluee that occur at specific rates over the duration of the specified assessment time frames (e.g., within one year of assessment on the Stable-2007 or a lifetime for the SRA-FV). The reliability of the assessment procedure is obviously affected by such factors as the ambiguity of terms, users assigning differing meaning to the rating criteria, the extent and quality of available information to rate factors, and rating biases (e.g., anchoring bias, negative and positive halo effects, or recency bias). Moreover, the subjectivity of the scoring methods opens the door to an allegiance effect diminishing reliability (Murrie, Boccacinni, Caperton, & Rufino, 2012; Boccaccini, Murrie, Caperton, & Hawes, 2009; Murrie, Boccaccini, Turner, Meeks et al., 2009). No matter what the source, variability in scorer judgment is likely the largest threat to the reliability of the results obtained from the Stable-2007, SRA-FV, and VRS-SO.

The Specialty Guidelines for Forensic Psychology includes, "Forensic practitioners seek to provide opinions and testimony that are sufficiently based upon adequate scientific

foundation, and reliable and valid principles and methods that have been applied appropriately to the facts of the case" (Guideline 2.05, American Psychological Association, 2011). How reliable is reliable enough? Qualitative labels provided by Cicchetti (1994) for inter-rater reliability are "poor" for Intra-class Correlation (ICC) values less than .40, "fair" for values between .40 and .59, "good" for values between .65 and .74, and "excellent" for values between .75 and 1.0. More relevant to the developers' proposals (Phenix et al., 2012; Thornton, 2011, 2010) to use dynamic risk-assessment instruments to guide the choice of a Static-99R reference group, Heilbrun (1992) recommends a minimum floor of .80 in the reliability coefficient when selecting which measures to use in forensic applications. Marshall (2006) argues for a higher inter-rater agreement of not less than .90.

One study examines the inter-rater agreement for the SRA-FV total score. In unpublished research, Sachsenmaier, Thornton, and Olson (2011) report an ICC of .55 among 19 psychologists who completed the SRA-FV on 69 individuals committed to a sexually violent predator civil confinement center in Wisconsin. Cicchetti (1994) considers this ICC value as a fair level of agreement among the raters. The ICC value of .55 represents the proportion of true-score variance (e.g., the extent to which the SRA-FV measures long term vulnerabilities). Subtracting this value from 1.0 reveals the proportion of error variance,[1] which is .45. When the error variance approximates the true-score variance, it becomes obvious that the SRA-FV total score will inform as often as it will mislead.

The Stable-2007 was developed and validated on samples of sexual offenders under community supervision (Hanson et al., 2007). Inter-rater reliability data was not provided for the developmental sample but it is interesting that two studies of sexual offenders in custody report it. Eher, Matthes, Schilling, Hauber-MacLean, and Rettenberger (2012) randomly selected 15 Stable-2007 protocols obtained from male prisoners in Austria and found excellent inter-rater reliability for the total score (ICC = .90). Fernandez (2008) studied 55 Stable-2007 ratings completed by correctional program officers for incarcerated Canadian sexual offenders. The ICC for the total score was .92.

The VRS-SO validation research (Olver, Wong, Nicholaichuk, & Gordon, 2007) studied participants from a high-intensity sexual offender treatment program operated in a Canadian prison. Thirty-five randomly selected cases were studied for inter-rater reliability. The ICC (single measure) was .74 for pretreatment dynamic risk total score. Beggs and Grace (2010) reported VRS-SO inter-rater reliability derived from a study of child molesters incarcerated at a New Zealand special treatment unit. ICC coefficients were computed on 23 cases scored by two raters. The pretreatment dynamic risk total score achieved an ICC of .90.

---

[1] In the true-score model, error variance is attributable to random sources irrelevant to the measurement of the trait or ability the instrument purports to measure in an observed score or distribution of scores. Common sources of error variance include those related to test construction (including item or content sampling), test administration, and test scoring and interpretation (Cohen & Swerdlik, 2001).

Available research reveals a level of inter-rater reliability for the SRA-FV that is inadequate for any forensic application. In contrast, the studies mentioned in this section suggest that inter-rater reliability is not an inherent barrier to the use of the Stable 2007 or the VRS-SO. The reliability of the procedure, however, is not sufficient to determine whether the proposed reference-group-selection methods work as intended. The validity of the preselection effect models is therefore explored in the next section.

Do Instrument Scores Accurately Quantify the Preselection Effect?

Validity determines the extent to which the Static-99R reference-group-selection procedures perform as intended. The validity of the Static-99R reference-group-selection system is premised on three untested assumptions. One, the preselection effect can be categorized into low, moderate, and high levels. Two, a single Static-99R reference group is associated with only one level of preselection effect. Three, each level of preselection effect can be quantitatively determined based on non-overlapping cut-scores from the Stable-2007, SRA-FV, and VRS-SO. For example, these assumptions would lead to the hypothesis that members from the PHRN population only evidence high levels of preselection as measured by SRA-FV scores equal to or greater than 3.3. The assumptions underlying the validity of the preselection effect model have not previously been tested, despite its widespread use. This section describes the methods used to test the three assumptions undergirding the Static-99R reference-group-selection system and reports the results of the analysis.

Methods and data analysis. In a conference workshop, Hanson and Thornton (2012) reported results from the administration of the SRA-FV, Stable-2007, and VRS-SO among samples that were considered representative of RC, PTN, and PHRN populations. The study examined 15 samples, two of which were part of the 20 samples comprising the three Static-99R reference groups (Hanson et al., 2007; Knight & Thornton, 2007). Appendix A provides the references for the studies. The research by Hanson and Thornton (2012) explored whether the level of preselection, as measured by the mean scores for the three measures, accounted for the discrimination of the base rate differences for the three Static-99R reference groups. The researchers found that the mean scores from the three instruments suggested that certain levels of preselection corresponded with each Static-99R reference group. The results further indicated that it might be possible to test whether Static-99R reference groups could be selected using mutually exclusive cut-off scores from the SRA-FV, Stable-2007, and VRS-SO.

To obtain the data necessary to test the reference-group-selection system, the author contacted the developers to request the following information for each reference group by each dynamic-risk measure: mean score, standard deviation, range of scores, number of participants, and number of studies. Dr. Hanson graciously provided the data with the exception of the range of scores, which was not available. The data was used for three purposes, including to test Thornton's (2011) SRA-FV selection model, to determine whether the Stable-2007 levels of dynamic risk (low, moderate, and high) correspond to the selection of Static-99R reference groups as reflected in Table 2, and

to explore the feasibility of constructing a selection model based on the VRS-SO pre-treatment dynamic total score.

Table 3
Stable-2007, SRA-FV, and VRS-SO: Number of Participants (N) and
Studies (k) and Descriptive Statistics by Static-99R Reference Groups

Reference Group (Preselection Effect)	N (k)	Mean Score	95% Confidence Interval	Standard Deviation	Expected Range of Scores[a]
Stable-2007 **(maximum score = 26)**					
RC (low)	1,198 (2)	7.06	6.7 – 7.4	5.15	0 – 22.51
PTN (moderate)	646 (6)	10.99	10.1 – 11.8	3.92	0 – 22.75
PHRN (high)	189 (1)	14.70	14.2 – 15.2	3.40	4.5 – 24.9
SRA-FV **(maximum score = 6)**					
PTN (moderate)	439 (3)	2.22	2.1 – 2.3	.922	0 – 4.99
PHRN (high)	513 (2)	3.26	3.2 – 3.3	.761	0.98 – 5.54
VRS-SO **(maximum score = 52)**					
PTN (moderate)	481 (3)	20.74	20.1 – 21.4	7.23	0 – 42.43
PHRN (high)	510 (2)	27.23	26.7 – 27.8	6.02	9.17 – 45.29

[a] Computed as ± 3 standard deviations from the mean score

Table 3 reports the descriptive data for each measure in three panels, with each panel representing a specific instrument. The number of studies and sample sizes for each reference group are reported in column 2. The mean scores and standard deviations provided by Dr. Hanson are reported in columns 3 and 5, respectively. The 95% confidence intervals for the mean values for each instrument are presented in column 4 and the author computed these data using the following formula:

$$\bar{x} \pm z\alpha_{/2}\left(\frac{\hat{s}}{\sqrt{n}}\right)$$
(1)

Where,

$\bar{x}$ = Estimate of population mean
$\hat{s}$ = Estimate of population standard deviation
α = 1 – (Confidence Level/100)
$z\alpha_{/2}$ = z-table value
n = Sample Size

Since Hanson and Thornton (2012) did not report actual ranges of scores, the author computed expected ranges of instrument scores by adding and subtracting three stan-

dard deviation units from the mean score. These data are shown in the last column of Table 3. Based on the statistical properties of a normal distribution of scores around the mean, it is expected that 99% of all scores in a population fall within three standard deviations below and above the mean. The lower limits for some of the ranges of scores fell below zero. In such instances, the lower limit was truncated at zero since the three measures do not report negative score values. Inspection of the last column in Table 3 reveals that the upper limit of the score range does not exceed the maximum score for each measure. This suggests that the standard deviations for scores corresponding to each reference group adequately estimate the range of scores for each instrument.

Results. Column 4 in Table 3 demonstrates that the 95% confidence interval about the mean instrument scores does not overlap between preselection levels. Since Helmus (2009) has shown that each reference group is statistically independent, as reflected by sexual recidivism base rates, a rule-of-thumb method can be applied to determine whether the Stable-2007, SRA-FV, and VRS-SO mean scores are statistically different across the three levels of preselection effect. Cumming and Finch (2005) indicate that non-intersecting confidence intervals among independent groups reflect statistically significant differences at $p < .01$. This indicates that mean instrument scores distinguish one reference group from another and this finding is consistent with the developers' hypothesis that higher base rate groups exhibit a greater density of risk factors external to the Static-99R. As discussed in the concluding section of the article, significant differences in Stable-2007, SRA-FV, and VRS-SO mean scores are insufficient to devise a valid system to select Static-99R reference groups.

The results summarized in Table 3 further reveal a wide dispersion of Stable-2007, SRA-FV, and VRS-SO scores around the mean values at each level of preselection. In fact, scores span nearly the entire range of values for each instrument. These data reflect that the mean dynamic-risk scores may be statistically different but sexual offenders within each reference group show remarkably similar variation in their dynamic-risk scores. Because the variation in dynamic-risk scores for each instrument is homogeneous across reference groups, clinicians cannot expect sexual offenders who are assessed on the Stable-2007, SRA-FV, or VRS-SO to demonstrate restricted ranges of dynamic-risk scores corresponding to low, moderate, or high levels of preselection. Taken together, these findings suggest that clinicians cannot rely upon the evaluee's total dynamic-risk score to select a single Static-99R reference group. This hypothesis is explored further next.

Table 4 reports the data used to assess the validity of the Stable-2007 and SRA-FV selection models. For the purposes of this comparison, the author rounded the mean SRA-FV scores reported in Table 4, column 3, to one decimal place, consistent with Thornton's (2011) recommendation for reporting the level-of-need scores. The following table summarizes the results from this analysis.

The results reported in Table 4, columns 2 and 3 reveal that Thornton's (2011) model for selecting Static-99R reference groups failed to replicate in the five studies where the

SRA-FV was administered. The results contradict that a single Static-99R reference group can be chosen based on a restricted range of SRA-FV total scores. This is apparent in two ways when inspecting the results from Table 4.

Table 4
Testing SRA-FV and Stable-2007 Selection Models

Reference Group (Preselection Effect)	SRA-FV Selection Model[a]	SRA-FV Results Mean (Range)	Stable-2007 Selection Model[b]	Stable-2007 Results Mean (Range)
RC (low)	≤ 2.3	-	0 - 3	7.06 (0 – 22.51)
PTN (moderate)	2.4 – 3.2	2.2[c] (0.00 – 4.99)	4 – 11	10.98 (0 – 22.74)
PHRN (high)	≥ 3.3	3.3[c] (0.98 – 5.54)	≥ 12	14.70 (4.5 – 24.9)

[a] Thornton (2011); b Hanson & Harris (2008); [c] mean scores from Hanson and Thornton (2012) rounded to single decimal place according to SRA-FV scoring instructions (Thornton, 2011)

One, the mean SRA-FV score of 2.2 for the PTN reference group (moderate preselection) falls in the critical region of ≤ 2.3 that defines the RC reference group (low preselection). Assuming a normal distribution of total scores, it is reasonable to conclude that more than one-half of the SRA-FV scores from the PTN reference group are likely consistent with the scores defining samples thought to have less preselection effect. Even though the mean SRA-FV score of 3.3 is at the lower limit for the cut-off associated with the PHRN reference group, it is likely that nearly half the scores below the mean fall within the lower risk PTN group.

Two, the SRA-FV cut-scores for the PTN and PHRN reference groups overlap substantially. As a result, there is a high probability that the SRA-FV score assigned to an evaluee will be consistent with two Static-99R reference groups instead of one reference group as Thornton's (2011) model posits. While the SRA-FV was not tested on samples representing routine correctional settings, it is reasonable to assume that the SRA-FV scores from the RC reference group would likely span a large portion of the range of possible SRA-FV values, thus overlapping the other two reference groups by wide margins.

The reader should note that Hanson and Harris (2008) designed the Stable-2007 cut-scores to be associated with low, moderate, and high dynamic risk, as shown in Table 4, column 4. For the purposes of this analysis, the author assumed that the three Stable-2007 cut-score ranges correspond to the respective Static-99R reference groups. The results in the last column of Table 4 reveal that the dispersion of Stable-2007 scores around the mean values for each Static-99R reference group is remarkably similar. Like the SRA-FV results, it is apparent that sexual offenders at any level of preselection demonstrate wide variation in Stable-2007 scores that run the gamut of dynamic-risk levels. As a result, clinicians will encounter the quandary of an evaluee's Stable-2007 score being representative of two, if not all three, Static-99R reference groups. Moreover, the homogenous variation in Stable-2007 scores across reference

groups bodes poorly for developing a reference-group-selection model that relies on mutually exclusive cut-off scores.

A goal of this study was to explore whether the VRS-SO pretreatment dynamic-risk cut-scores could be developed to select a single Static-99R reference group. Inspection of the last column in Table 3, panel 3, casts substantial doubt that such an endeavor would be successful. Like the two other instruments, the dispersion of VRS-SO pre-treatment dynamic-risk total scores around the mean values for the reference groups are so homogeneous that it would be virtually impossible to establish mutually exclusive cut-off scores to select a single Static-99R reference group.

Discussion and Conclusions

In developing the Static-99R, Helmus (2009) found that twenty convenience samples sorted into three groups, based on statistically significant differences in sexual recidivism base rates. The three groups were designated as RC, PTN, and PHRN, and the developers issued separate sexual recidivism experience tables for each reference group (Phenix et al., 2012). The developers used post-hoc logical inference to hypothesize that the density of risk factors external to the Static-99R explain the statistically significant differences in sexual recidivism base rates among all the samples and the three reference groups. Based on this conceptualization, the developers recommend using SRA-FV, Stable-2007, or VRS-SO to quantify risk factors external to the Static-99R. The reference-group-selection procedures have been advanced in clinical and forensic practice without virtually any empirical examination as to the reliability or validity of the methods.

In their initial effort to capture the preselection effect, the developers devised a qualitative system for selecting reference groups based on cohort matching. This procedure introduces a fourth reference group (non-routine corrections) into the selection mix (Phenix et al., 2012; p. 19). Wollert (2010) points out that the reliability for this procedure has not been established and it is fraught with misclassification error that degrades the accuracy of score-wise risk estimates. To ameliorate misclassification error, Wollert (2010) recommends that users consider a single-experience table representing the aggregate of the four Static-99R reference groups, but the developers have not provided this data.

More recently, the developers appear to have moved in a direction to quantify the pre-selection effect using instruments that purport to measure risk factors external to the Static-99R. This selection model hinges on the conceptualization that the total scores derived from measures of risk factors external to the Static-99R quantify the preselection effect. Moreover, it is speculated that the total instrument scores can be grouped into mutually exclusive cut-scores that discriminate each reference group. Phenix et al. (2012) describe this as the preferred method for reference group selection and the cohort-matching procedure is applied when users do not complete an instrument that measures other risk factors. Phenix et al. (2012) do not recommend a specific instrument to assess the preselection effect but other publications by members of the Static-

99R team suggest the use of the Stable-2007, SRA-FV, and VRS-SO (Thornton et al., 2010; Thornton, 2010 & 2011). This study appears to be the first attempt to critically analyze the reliability and validity of applying the three instruments to quantify the hypothesized preselection effect when choosing Static-99R reference groups.

The limited number of inter-rater agreement studies for the Stable-2007 and VRS-SO demonstrate that these measures appear to achieve a minimum level of reliability for forensic use (Heilbrun, 1992). It is conceivable that the Stable-2007 and VRS-SO total scores could be used to establish a Static-99R reference-group-selection system. On the other hand, data from Sachsenmaier et al. (2011) show the reliability for the SRA-FV total score falls far below minimum standards for forensic practice. There is an urgent need to revise the SRA-FV scoring system to increase its reliability, if this is even possible. Until then, clinicians who choose to conduct SRA-FV assessments would be hard pressed to justify its use statistically (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2003; Anastasi & Urbina, 1997) and ethically (American Psychological Association, 2011 & 2010) when selecting Static-99R reference groups and, for that matter, in other forensic applications. Even if the reliability of the Stable-2007, SRA-FV, and VRS-SO achieve acceptable levels for forensic practice, this does not ensure that the resulting models to select Static-99R reference groups are valid.

Validity of Selection Models

Hanson and Thornton (2012), Phenix et al. (2012), and Thornton et al. (2010) contend that that hypothesized preselection effect can be quantified based on total scores from measures of risk factors external to the Static-99R. Using data obtained from Hanson and Thornton (2012), the results from the analysis were consistent with the developers' hypotheses in part. On the one hand, the Stable-2007, SRA-FV, and VRS-SO mean total scores were in the expected direction (e.g., higher mean scores for higher base rate groups), as predicted by Hanson and Thornton (2012). On the other hand, the Stable-2007 and SRA-FV selection models overlapped by such wide margins that it would be virtually impossible to select a single Static-99R reference group.

The Stable-2007 and SRA-FV selection models assume that mutually exclusive cut-scores define the levels of preselection (i.e., low, moderate, or high) associated with the Static-99R reference group (RC, PTN, or PHRN, respectively). The total instrument score should result in the selection of a single Static-99R reference group. The results of this analysis refute this hypothesis, as the total instrument score assigned to an evaluee would likely be consistent with choosing at least two if not all three of the Static-99R reference group options. Only the extreme upper ends of the score distribution for the Stable-2007 and SRA-FV were unaffected by score duplication. This finding appears of limited utility forensically as it affects a tiny proportion of the score distribution. It is clear that these data directly contradict the contention of the developers (Phenix et al., 2012; Thornton, 2011; Thornton et al., 2010) that the density of risk factors external to the Static-99R can be used to select a single reference group.

The results from the VRS-SO studies were similar to the findings from the Stable-2007 and SRA-FV. As seen in Table 3, the range of pretreatment dynamic risk total scores overlapped substantially for the PTN and PHRN reference groups. The participants from the PHRN reference group were distinguished from the PTN group by the fact that they did not score below nine, but this garners little forensic utility because scores of nine or greater are also representative of the PTN reference group score distribution. The members of the PHRN reference group also scored slightly higher overall on the pretreatment dynamic risk total score than their PTN counterparts, but such discrimination has little utility since it affects a very small proportion of the score distribution. Like the other two instruments, it would be expected that the range (proxy of score distribution) of pretreatment dynamic risk total scores for the RC reference group would overlap substantially with the PTN and PHRN reference groups. Taken together, these results suggest that the prospect is poor for constructing a valid selection model based on mutually exclusive VRS-SO pretreatment dynamic-risk cut-scores.

Possible Reasons for Invalidity of the Selection Models

While the Stable-2007, SRA-FV, and VRS-SO data from Hanson and Thornton (2012) did not test inter-rater agreement, this potential source of error would affect the validity of the reference-group-selection methods. Low levels of inter-rater agreement for the total SRA-FV, Stable-2007, and VRS-SO scores may explain the variability in score ranges that caused the wide overlap in cut-scores associated with each Static-99R reference group. The necessity for precise measurement cannot be over-emphasized. However, adequate reliability does not ensure that the Static-99R selection system is valid. It appears that the three selection models tested in this study failed to perform as intended because, likely, the magnitude of risk factors external to the Static-99R was an incorrect assumption to explain the statistically significant differences in sexual recidivism base rates between the three reference groups.

The failure to establish the validity of the selection models is not surprising when considering that the developers (Helmus, 2009) employed post-hoc logical inference to hypothesize the preselection effect. Devising a selection system based on non-overlapping score ranges runs counter to the expectation that the range of observed instrument scores is likely to be homogeneous among different groups. For instance, the selection system that Thornton (2011; 2010) devised is akin to the expectation that Static-99R scores within each reference group would be restricted to a certain range without overlapping other reference groups such that the Static-99R scores for members of the RC reference group would be ≤ 2, 3 to 4 for the PTN reference group, and ≥ 5 for the PHRN reference group. To the contrary, in all Static-99R reference groups, the participants demonstrate a full range of scores. This observation mirrors the results from this study demonstrating similar dispersion of Stable-2007, SRA-FV, and VRS-SO total scores across the three reference groups.

The mean instrument scores observed by Hanson and Thornton (2012) produce statistically significant differences in the levels of preselection effect across all three instruments. The significant differences in mean scores for the three measures suggest that,

as the base rate increases across reference groups, there is a concomitant escalation in levels of external risk factors. This observation might reflect a true difference between reference groups based on total average scores from the Stable-2007, SRA-FV, and VRS-SO, or it might represent a high correlation between the Static-99R and the measures of external risk factors. If the Static-99R and external risk measures account for a high proportion of the same variance associated with sexual recidivism, then the significant differences in mean instrument scores across preselection levels may only represent the high degree of co-variation. Alternatively, if a sufficient number of external risk factors redundant with the Static-99R produce error of sufficient magnitude, then it would confound the sensitivity of a few items contributing to the hypothesized preselection effect. The extent of shared variance among measures needs further study but, nevertheless, it is obvious that the homogeneous dispersion of instrument scores across the three levels of preselection presents an insurmountable obstacle in designing a Static-99R reference-group-selection model that relies on non-overlapping critical regions as proposed by Thornton (2011) or when using the existing Stable-2007 cut-off score system (Hanson & Harris, 2008).

Finally, there is a fundamental problem with proposing a selection model that is premised on post-hoc logical inference. It is simply unknown whether risk factors external to the Static-99R even explain the differences in base rates among the RC, PTN, and PHRN reference groups. The developers used post-hoc logical inference to hypothesize the preselection effect as discriminating different Static-99R reference groups without empirical validation. The selection models tested in this analysis may not have performed as hypothesized because of the speculative nature of their foundation.

Before advocating that clinicians employ untested selection models to choose Static-99R reference groups, it seems reasonable to expect the developers to prove the validity and reliability of their methods. As a first step, the developers could generate a list of possible risk factors external to the Static-99R. Next, the intercorrelation between the Static-99R and identified external risk factors could be computed to determine which factors add unique variance above the Static-99R. Any risk factors found to have statistical independence from the Static-99R could then be tested to determine if a reliable and valid selection model could be developed. If such a model were established, it would need to undergo sufficient replication studies (Schmidt, 2009) before it is deployed for forensic use. Until then, it seems scientifically and ethically indefensible for clinicians to apply speculative reference-group-selection methods in forensic practice.

Implications for Clinical and Forensic Practice

The most obvious conclusion for clinicians to draw from this analysis is to abandon the use of the Static-99R because it lacks a reliable or valid method by which to choose one of three or four reference groups to compare to the individual being assessed. Clinicians cannot have any level of confidence in their predictions of sexual recidivism risk when it is uncertain whether the correct reference group was selected. Short of scuttling the Static-99R reference-group-selection system, this article concludes by explor-

ing two alternate selection procedures. The first method is referred to as *base-rate matching* and relies on using the current Static-99R reference groups. The second selection procedure rejects choosing among four options in favor of a single reference group along the lines recommended by Wollert (2010). It is obvious that an alternate method for selecting a reference group is unnecessary when clinicians have a Static-99R experience table developed and validated on a local group of sexual offenders that is representative of the individual being assessed.

Base-rate matching. The base-rate-matching approach is premised on the concept of selecting a Static-99R reference group that has a base rate of sexual recidivism most similar to the group representing the individual being assessed. This method is grounded in mathematics showing that the recidivism estimate at each score on an actuarial instrument is a function of the observed base rate of sexual recidivism for the group (Donaldson & Wollert, 2008). Matching the base rate of sexual recidivism between the group representing the individual being assessed and a Static-99R reference group will produce the most accurate score-wise estimate to compare to the evaluee.

Donaldson and Wollert (2008) show how the risk estimate at a given score is determined largely by the observed base rate of sexual recidivism in the offender group. This is a direct effect whereby the risk estimate at each score on the actuarial measure increases as the base rate rises. The recidivism rate at each score is also further determined by the score-wise likelihood ratio. The risk estimates for the selected Static-99R reference group will be inaccurate as applied to the group representing the evaluee when the base rates of sexual recidivism differ substantially between the local population of sexual offenders and the selected Static-99R reference group. Helmus (2009) and Helmus, Hanson et al. (2012) document significant variability in base rates across the samples comprising the Static-99R experience tables. As a result, users cannot assume the base rate of sexual recidivism expressed in an experience table is consistent with the group representing the evaluee when applying either of the two recommended selection procedures. Rather, it makes statistical sense to select a Static-99R reference group that reflects a base rate of sexual recidivism similar to the local group representing the evaluee.

The base-rate-matching approach involves identifying a sexual recidivism rate representative of the local group to which the evaluee belongs. The local sexual recidivism base rate is relied upon to select a Static-99R reference group that has the most similar base rate. An exact match in base rates may not occur; however, a variation of a few percentage points will not make a meaningful difference in the validity of the score-wise risk estimate as applied to the evaluee. An example involving sexually violent predators illustrates this point.

The author has observed the practice of government evaluators who assess sexually violent predators in California by typically choosing the PHRN reference group when using either of the two Static-99R reference-group-selection procedures. The five-year sexual recidivism base rate of the PHRN reference group is 21%. Data from several

studies in California contradict this base rate as representative of the sexual recidivism potential for individuals being considered for involuntary civil confinement. The California Department of Corrections (2010) and the California Sexual Offender Management Board (2008) studied two separate cohorts of sexual offenders released from prison in 2006 and 2003, respectively, with an aggregate sample size of 11,298. The average three-year sexual recidivism rate was 3.4%. This figure suggests a five-year rate near 6%. A third research project by Padilla (2006) and as reported by Zavodny, DeClue, and Cohen (2012), found that that a group of 93 petitioned sexually violent predators released from custody at the commitment center in California reoffended sexually at a rate of 6% over the 4.7 year follow-up. The rate of sexually violent reoffense, as defined by the California Welfare and Institution Code 6600 (b), was 4.3% over the same period. These base-rate data support the selection of the Static-99R RC reference group that has a 6% sexual recidivism rate over a five-year follow-up. The comparability of sexual recidivism base rates between the local offender group and the RC reference group would produce the most accurate score-wise sexual recidivism rates to compare to the SVP candidate being assessed. Choosing the PHRN reference group would instead result in reporting score-wise risk estimates that are approximately three times greater than would be expected in a lower base rate population.

In practical application, a local base rate of sexual recidivism may not be available to aid in the selection of a Static-99R reference group. In this event, evaluators could turn to nationally representative studies of sexual recidivism base rates to assist in selecting a Static-99R reference group. For instance, Wollert and Waggoner (2009) report a five-year sexual recidivism base rate of 6.5% among 17,697 participants from the United States. Of the 9,691 sexual offenders released from prisons in 15 states during 1994, Langan, Schmitt, and Durose (2003) found that 5.3% were rearrested for another sexual offense within three years. This would suggest an 8.8% sexual reoffense rate over five years.

Clinicians relying upon national studies must be sensitive as to how sexual recidivism base rates vary across samples or jurisdictions (Helmus, Hanson et al., 2012). The average base rate of sexual recidivism determined by the aggregated data may not accurately represent the probability of sexual reoffense for some of the independent groups comprising the combined sample. As a result, the base rate may be incompatible with that of the local jurisdiction representing the evaluee. This potential source of inaccuracy could be alleviated by conducting sexual recidivism studies at the local level (i.e., state-wide). Whenever possible, it is preferable to produce Static-99R experience tables for local jurisdictions.

A limitation of the base-rate-matching approach is the assumption that the score-wise likelihood ratios are similar between the selected Static-99R reference group and the local group representing the evaluee. The accuracy of the score-wise risk estimate as applied to the individual would be affected adversely should the score-wise likelihoods vary substantially between the selected Static-99R experience table and the local population representing the evaluee. The same problem exists when using any method to select an actuarial experience table to compare to the evaluee. Clinicians should rec-

ognize and consider this source of potential error when rendering conclusions and make it known in reports and testimony.

Aggregate reference group. Consistent with the recommendation of Wollert (2010), Table 5 reports data for a single experience table comprising all 23 Static-99R samples and explores its application in risk assessments. The developers have not released this data; however, the author compiled it, as described below, and refers to the data-set as the *Aggregate reference group*. The five-year base rate of sexual recidivism for the Aggregate reference group is 11.1% and the AUC is .714 (95% CI = .694, .734), which reflects moderate selection accuracy. Table 5 provides the observed and predicted sexual recidivism rates for the Aggregate reference group. The method by which the data in Table 5 were produced is explained in the next paragraph.

The observed recidivism rates for the Aggregate reference group, at each score and for the entire sample, were compiled by combining the frequency data from the NRC group (15 studies) and the RC group (8 studies). These data are contained in the detail recidivism tables provided by the developers.[2] Each score-wise observed recidivism rate was calculated by taking the number of recidivists at a given score divided by the corresponding total number of participants. Predicted estimates of sexual recidivism for the Aggregate reference group were computed using Logistic Regression. The slope parameter (B_1) to compute the predicted risk estimates is based on the frequency data for the aggregate sample. This computation differs from the one used by the developers. The Static-99R developers calculated the average slope parameter from twenty studies and used it in the regression analysis to arrive at the predicted score-wise risk estimates for each reference group (Helmus, 2009). Using the average slope parameter for twenty studies for the Logistic Regression has been found to produce spurious predicted risk estimates within reference groups, which is especially pronounced in the PHRN reference group (Donaldson, Abbott, & Michie, 2012). Precision of the observed score-wise risk estimate was computed using the 95% binomial exact CI (Clopper & Pearson, 1934), which is a preferred method when dealing with dichotomous outcome data. The accuracy of the risk estimate as applied to a presumed individual from the actuarial sample is represented by the 95% confidence interval for the individual ("CII"). Further details regarding the methods for calculating the predicted risk estimates and accuracy data about them have been previously explained by Donaldson and Abbott (2011).

---

[2] Available at http://www.static99.org/pdfdocs/detailed_recid_tables_static99r_2009-11-15.pdf

Table 5
Static-99R Aggregate Reference Group Experience
Table: Five-Year Observed and Predicted Risk Estimates

Static-99R Score	Total n of Recidivists	Total n	Observed Recidivism Rate (%)	95% CI	Predicted Recidivism Rate (%)[+]	95% CI[+]	95% CII[+]
-3	0	78	0.0	0.0, 4.0	2.0	1.0, 2.0	0.0, 93.0
-2	4	113	3.5	0.97, 8.8	2.0	2.0, 3.0	0.0, 95.0
-1	14	497	2.8	1.6, 4.7	3.0	3.0, 4.0	0.0, 96.0
0	27	619	4.4	2.9, 6.3	4.0	4.0, 5.0	0.0, 97.0
1	42	735	5.7	4.2, 7.7	6.0	5.0, 6.0	0.0, 98.0
2	51	743	6.9	5.2, 8.9	7.0	7.0, 8.0	0.0, 98.0
3	75	770	9.7	7.7, 12.1	10.0	9.0, 11.0	0.0, 99.0
4	88	729	12.1	9.8, 14.7	13.0	12.0, 14.0	0.0, 99.0
5	107	593	18.0	15.0, 21.4	16.0	15.0, 18.0	0.0, 99.0
6	93	404	23.0	19.0, 27.4	21.0	19.0, 24.0	0.0, 100
7	66	259	25.5	20.2, 31.3	27.0	24.0, 30.0	0.0, 100
8	44	137	32.1	24.4, 40.6	33.0	29.0, 37.0	0.0, 100
9	21	57	36.8	24.5, 50.7	40.0	35.0, 45.0	0.0, 100
10	10	22	45.5	24.4, 67.8	47.0	41.0, 54.0	0.0, 100
11	0	3	0	0.0, 70.8	55.0	50.0, 60.0	0.0, 100
Total	642	5759	11.1	10.4, 12.0	-	-	-

+ Data from Donaldson and Abbott (2011).

There appear to be some advantages when using the Aggregate reference group only to compare to individuals being assessed, such as reducing potential classification error associated with selecting among a variety of reference groups (Wollert, 2010) and the greater variability in the data results in increased reliability (Anastasi & Urbina, 1997). The benefits of using a single actuarial table are outweighed by two limitations. One, the sexual recidivism base rate of 11% may not be accurate as applied to the group representing the evaluee and this would produce an inaccurate estimation of risk. Second, the Aggregate reference group combines data from 23 convenience samples. The membership of this group was not selected using representative sampling methods so the use of a single experience table raises concerns as to the generalizability of the risk data to other groups of sexual offenders. As a stand-alone experience table, it cannot be assumed the risk data from the Aggregate reference group accurately represents the recidivism potential of sexual offenders who are scored on the Static-99R. For these reasons, it seems most prudent for clinicians to consider the experience table from the Aggregate reference group as another option to select when applying the base-rate-matching approach.

Recommendations for Selecting Static-99R Reference Groups. For clinical and forensic practitioners who conduct risk assessments with the Static-99R, it is always preferable to consider an experience table containing score-wise risk estimates

that were developed and validated on a local group of sexual offenders representing the evaluee. When this choice is unavailable, it is recommended that users follow the base rate-matching approach, as described previously. The following paragraph summarizes how to employ this procedure.

The base rate matching approach requires users to determine a sexual recidivism base rate at either the five-year or ten-year follow ups that are most consistent with the group representing the evaluee. The user should be confident that the evaluee is a member of the selected base rate group, which may be found among the local population of sexual offenders or in another representative sample. As stated earlier, the base rate match does not need to be exact, as a variation of a few percentage points would not make a significant difference in the accuracy of the score-wise risk estimates as applied to the group representing the evaluee. Table 6 provides a reference by which to select a Static-99R reference group when applying the base rate matching procedure. Users can obtain the experience table for the Aggregate reference group from Table 5 and the experience tables for the remaining reference groups are located at the following link.[3]

Table 6
Guidelines for Selecting a Static-99R Reference
Group Using the Base Rate Matching Approach

Reference group to select	When 5-year sexual recidivism base rate is approximately:	When 10-year sexual recidivism base rate is approximately:
RC	6.0%	N.A.[+]
PTN	9.0%	13.0%
Aggregate	11.0%	N.A.[+]
NRC	15.0%	20.0%
PHRN	21.0%	29.0%

[+] Not available

The following example illustrates how to employ Table 6 to select a Static-99R reference group. A clinician is tasked with determining the sexual recidivism potential of an individual before he is released from prison on parole as the outcome will determine the level of community supervision methods to which the offender will be subjected. The clinician completes the Static-99R and assigns the evaluee a score of four. The clinician is aware that the local corrections agency recently conducted a sexual recidivism study for sex offenders released from prison and found a five-year rate of 5.2%. The clinician determines the evaluee is represented by the membership of the study sample. Comparing the local base rate of 5.2% to the values contained in Table 6 informs the clinician to select the RC reference group. The clinician then accesses the current Static-99R RC experience table from the Static99.org website to obtain the five-year predicted estimate at the score of four.

---

[3]Available at http://www.static99.org/pdfdocs/detailed_recid_tables_static99r_2009-11-15.pdf

When reporting the score-wise risk estimate, forensic and clinical practitioners are reminded that the base-rate-matching approach assumes similarity in the score-wise likelihood ratios between the selected Static-99R reference group and the group representing the individual. To the extent the score-wise likelihood ratios diverge substantially, the Static-99R sexual recidivism rates may over or under estimate the risk potential of the group representing the evaluee. This limitation should be described in reports or testimony. Readers who are unfamiliar with the interaction between the sexual recidivism base rate and score-wise likelihood ratios are referred to Donaldson and Wollert (2008) for a full explanation.

Acknowledgements

References

Abbott, B. R. (2009). Applicability of the new Static-99 experience tables in sexually violent predator risk assessments. *Sexual Offender Treatment,* 1, 1-24.

American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2003). *Standards for Educational and Psychological Testing,* Washington, DC: American Educational Research Association.

American Psychological Association. (2011). *Specialty guidelines for forensic psychology.* Washington, D.C.: American Psychological Association.

American Psychological Association. (2010). *Ethical principles of psychologists and code of conduct.* Washington, D.C.: American Psychological Association.

Anastasi, A. & Urbina, S. (1997). *Psychological Testing: Seventh Edition.* New Jersey: Prentice-Hall.

Bartosh, D. L., Garby, T., & Lewis, D., & Gray, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology, 47,* 422-438. doi: 10.1177/0306624X03253850.

Beggs, S. M. & Grace, R. C. (2010). Assessment of dynamic risk factors: An independent validation study of the Violence Risk Scale: Sexual Offender Version. *Sexual Abuse: A Journal of Research and Treatment,* 22(2), 234-251. doi: 10.1177/ 1079063210369014.

Bigras, J. (2007). La prédiction de la récidive chez les délinquants sexuels [Prediction of recidivism among sex offenders]. *Dissertations Abstracts International, 68* (09). (UMI No. NR30941).

Boccaccini, M. T., Murrie, D. C., Caperton, J. D., & Hawes, S. W. (2009). Field validity of the Static-99 and MnSOST-R among sex offenders evaluated for civil commitment as sexually violent predators. *Psychology, Public Policy, and Law,* 15(4), 278-314. doi: 10.1037/a0017232.

Boer, A. (2003). *Evaluating the Static-99 and Static-2002 risk scales using Canadian sexual offenders.* Unpublished master's thesis, University of Leicester, Leicester, United Kingdom.

California Department of Corrections (2010). *2010 Adult Institutions Outcome Evaluation Report.* Retrieved on November 10, 2010, from: http://www.cdcr.ca.gov/adult_research_branch/research_documents/arb_fy0506_outcome_evaluation_report.pdf

California Sexual Offender Management Board. (2008, January). *An Assessment of Current Management Practices of Adult Sex Offenders in California: Initial Report.* Retrieved March 3, 2008, from www.cdcr.ca.gov/News/docs/SOMB%20Report022108.pdf.

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. doi: 10.1037/1040-3590.6.4.284.

Clopper, C. & Pearson, S. (1934). The use of confidence or fiducial limits illustrated in the case of the Binomial. *Biometrika ,* 26, 404-413. doi: 10.1093/biomet/26.4.

Cohen, R. J. & Swerdlik, M. (2001). *Psychological testing and assessment: An introduction to tests and measurement-* Fifth Edition. Columbus, OH: McGraw Hill.

Craissati, J., Bierer, K., & South, R. (2008). *What do sex offenders really get up to? Risk prediction, community failure and "sexually risky behaviours" in a nine-year follow-up study.* Unpublished manuscript.

Cumming, G. & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist,* 60(2), 170-180. doi: 1037/0003-066X.60.2.170.

Donaldson, T. S. & Abbott, B. R. (2011). Prediction in the individual case: An explanation and application of its use with the Static-99R in sexually violent predator risk assessments. *American Journal of Forensic Psychology,* 29(1), 5-35.

Donaldson, T. S., Abbott, B. R., & Michie, C. (2012). Problems with the Static-99R prediction estimates and confidence intervals. *Open Access Journal of Forensic Psychology,* 4, 1-23.

Donaldson, T. & Wollert, R. (2008). A mathematical proof and example that Bayes's theorem is fundamental to actuarial estimates of sexual recidivism risk. *Sexual Abuse: A Journal of Research and Treatment,* 20(2), 206-217. doi: 10.1177/1079063208317734.

Eher, R., Matthes, A., Schilling, F., Hauber-MacLean, T. & Rettenberger, M., (2012). Dynamic Risk Assessment in Sexual Offenders Using Stable-2000 and the Stable-2007: An Investigation of Predictive and Incremental Validity. *Sexual Abuse: A Journal of Research and Treatment,* 24, 5-28. doi: 10.1177/1079063211403164.

Eher, R., Rettenberger, M., Schilling, F., & Pfafflin, F. (2009). [*Data from sex offenders released from prison in Austria*]. Unpublished raw data.

Epperson, D. L. (2003). *Validation of the MnSOST-R, Static-99, and RRASOR with North Dakota prison and probation samples.* Unpublished Technical Assistance Report, North Dakota Division of Parole and Probation.

Fernandez, Y. (2008, October). *An examination of the inter-rater reliability of the Static-99 and Stable-2007.* Poster presentation at the 27th Annual Research and Treatment conference of the Association for the Treatment of Sexual Abusers, Atlanta, GA. Available at:
https://www.dropbox.com/sh/oss0b9ut7hx69me/RH1Uzz2qb7.

Hanson, R. K. & Harris, A. J. R. (2008, February). *Stable-2007 master coding guide.* Unpublished document.

Hanson, R. K., Harris, A. J. R., Scott, T. L., & Helmus, L. (2007). *Assessing risk of sexual offenders on community supervision: The dynamic supervision project.* Retrieved March 23, 2008, from
http://www.static99.org/pdfdocs/hansonharrisscottandhelmus2007.pdf.

Hanson, R. K., Lloyd, C. D., Helmus, L., & Thornton, D. (2012). Developing non-arbitrary metrics for risk communication: Percentile ranks for the Static-99/R and Static-2002/R sexual offender risk tools. *International Journal of Forensic Mental Health, 11*(1), 9-23.

Hanson, R. K., Phenix, A., Helmus, L. (2009, September). *Static-99R and Static-2002R: How to interpret and report in light of recent research.* Pre-Conference workshop at the 28th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, TX. Retrieved on January 2, 2011, from:
http://www.static99.org/pdfdocs/atsapre-conparti_ii_static-99r02r20090930.pdf.

Hanson, R. K. & Thornton, D. (2012, October). *Preselection effects can explain group differences in sexual recidivism base rates in Static-99R validation studies.* Workshop presented at ATSA 31st Annual Research and Treatment Conference, Denver, CO. Available at:
https://www.dropbox.com/sh/oss0b9ut7hx69me/RH1Uzz2qb7

Hanson, R. K. & Thornton, D. (2000). Improving risk assessment for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*(1), 119-136.

Harris, A. J. R., Helmus, L, Hanson, R. K., & Thornton, D. (2008). *Are new norms needed for Static-99?* Workshop presented at the ATSA 27th Annual Research and Treatment Conference on October 23, 2008, Atlanta: GA. Retrieved January 10, 2009: http://www.static99.org/pdfdocs/arenewnormsneededforstatic-99october2008.pdf

Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior, 16,* 257-272. doi: 10.1007/BF01044769.

Helmus, L. (2009). *Re-norming Static-99 recidivism estimates: Exploring base rate variability across sex offender samples* (Master's thesis). Available from ProQuest Dissertations and Theses database. (UMI No. MR58443).

Helmus, L., Hanson, R. K., Thornton, D. Babchishin, K. M., & Harris, A. J. R. (2012). Absolute recidivism rates predicted by Static-99R and Static-2002R sex offender risk assessment tools vary across samples: A meta-analysis. *Criminal Justice and Behavior, 39, 1148-1171. doi:* 10.1177/0093854812443648

Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Assessing risk of older sex offenders: Developing the Static-99R and Static-2002R. *Sexual Abuse: A Journal of Research and Treatment,* 24(1), 64-101. doi: 10.1177/1079063211409951

Knight, R. A. & Thornton, D. (2007). *Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders.* Washington, D.C.: U.S. Department of Justice. Retrieved July 1, 2008, from: https://www.ncjrs.gov/pdffiles1/nij/grants/217618.pdf.

Langan, P. A., Schmitt, E. L., & Durose, M. R. (2003). Recidivism of sex offenders released from prison in 1994. Washington, D.C.: U.S. Dept of Justice. Available at: http://bjs.gov/content/pub/pdf/rsorp94.pdf.

Langström, N. (2004). Accuracy of actuarial procedures for assessment of sexual offender recidivism risk may vary across ethnicity. *Sexual Abuse: A Journal of Research and Treatment, 16,* 107-120.

Marshall, W. L. (2006). Diagnostic problems with sexual offenders. In, W.L. Marshall, Y.M. Fernandez, L.E. Marshall, G.A. Serran (Eds.), *Sexual Offender Treatment: Controversial Issues* (pp. 33-43), West Sussex, UK: John Wiley and Sons.

Murrie, D. C., Boccaccini, M. T., Caperton, J., & Rufino, K. (2012). Field Validity of the Psychopathy Checklist–Revised in Sex Offender Risk Assessment. *Psychological Assessment.* 24(2), 524-549. doi: 10.1037/a0026015

Murrie, D. C., Boccaccini, M. T., Turner, D., Meeks, M., Woods, C., & Tussey, C. (2009) Rater (dis)agreement on risk assessment measures in sexually violent predator

proceedings: evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, and the Law, 15,* 19–53. doi: 10.1037/a0014897.

Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale–Sexual Offender version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19,* 318-329. doi: 10.1037/1040-3590.19.3.318

Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2013). *VRS-SO norms.* Unpublished document.

Padilla, J. (2006, October). *Communication to Jim McEntee, Public Defender.* Available at: https://www.dropbox.com/sh/oss0b9ut7hx69me/RH1Uzz2qb7.

Phenix, A., Helmus, L., & Hanson, R. K. (2012, July). *Static-99R and Static-2002R evaluators' workbook.* Retrieved July 30, 2012, from: http://www.static99.org/pdfdocs/Static-99RandStatic-2002R_EvaluatorsWorkbook2012-07-26.pdf.

Phenix, A., Helmus, L., & Hanson, R. K. (2009, September). *Static-99R Evaluator's Workbook.* Available at: https://www.dropbox.com/sh/oss0b9ut7hx69me/RH1Uzz2qb7.

Sachsenmaier, S., Thornton, D., & Olson, G. (2011, November). *Structured risk assessment forensic version (SRA-FV): Score distribution, inter-rater reliability, and margin of error in an SVP population.* Presentation at the 30th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Toronto, Canada. Available at: https://www.dropbox.com/sh/oss0b9ut7hx69me/RH1Uzz2qb7

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology,* 13(2), 90-100.

Thornton, D. (2012, May). *Scoring guidelines for need assessment: Structured risk assessment- forensic version – 1.52.* Unpublished paper, available at: https://www.dropbox.com/sh/oss0b9ut7hx69me/RH1Uzz2qb7

Thornton, D. (2011, December). *Notes for using SRA-FV level of need index.* Unpublished paper, available at: https://www.dropbox.com/sh/oss0b9ut7hx69me/RH1Uzz2qb7

Thornton, D. (2010). *Using the forensic version of the SRA in sex offender risk assessment.* Presented at workshop sponsored by Central Coast Clinical and Forensic Psychology Services, Atascadero, CA (December 2 & 3, 2010).

Thornton, D., Hanson, R. K., & Helmus, L. (2010). *Moving beyond the standard model for actuarial risk assessment for sexual offenders. The Forum. 22,* Summer.

Thornton, D. & Knight, R. (2009, October). *Using SRA Need domains based on structured judgment to revise relative risk assessments based on Static-2002 and Risk Matrix 2000.* Presentation at the 28th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, Texas.

Wollert, R. (2010, March). *The use of probability mathematics in sexual violent predator (SVP) evaluations.* Workshop presented at California Department of Mental Health, Seaside, CA, March 9, 2010. Available at: https://www.dropbox.com/sh/oss0b9ut7hx69me/RH1Uzz2qb7

Wollert, R. & Waggoner, J. (2009). Bayesian computations protect sexually violent predator evaluations from the degrading effects of confirmatory bias and illusions of certainty: A reply to Doren and Levenson. *Sexual Offender Treatment, 1,* 1-23.

Wong, S. C. P., Olver, M., Nicholaichuk, T., & Gordon, A. (2003). *Violence risk scale: Sexual Offender Version.* Unpublished scoring manual. Available at: http://shop.psynergy.ca/3-VRS-SO-Manual-VRSSO01.htm.

Zavodny, D. L., DeClue, G., & Cohen, J. (2012, August). *Sexual recidivism base rates relevant to SVP assessments.* Paper presented at the 120th Annual Convention of the American Psychological Association, Orlando, FL.

Appendix A: References for Studies Listed in Table 2

Stable 2007

Routine Corrections

Gillespie, L., & Anderson, D. (2008). Using validated assessments to guide supervision: Partnerships for Sustainability. Department of Public Safety, Hawaii.

Hanson, R. K., Harris, A. J. R., Scott, T. L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (Corrections Research User Report No. 2007-05). Ottawa, ON: Public Safety Canada.

Preselected Treatment Need

Bélair, S. (2008). *La prédiction de la récidive sexuelle, violente et criminelle chez les agresseurs de femmes adultes et d'enfants.* Ph.D. thesis, University of Montréal.

Coles, R. L. (2011). *Dynamic risk assessment, personality disorder, and key developmental variables in sexual offenders.* Ph.D. thesis, University of Birmingham.

Cookman, C. L. (2010). The utility of the Assessment of Risk Management of Intellectually Disabled Individuals who Offend – Sexually (ARMIDILO-S) for a community-based service. Ph.D. dissertation. Massachusetts School of Professional Psychology.

Eher, R., Matthes, A., Schilling, F., Haubner-MacLean, T., & Rettenberger, M. (2012). Dynamic risk assessment in sexual offenders using STABLE-2000 and the STABLE-2007: An investigation of predictive and incremental validity. *Sexual Abuse: A Journal of Research and Treatment, 24,* 5-28. doi:10.1177/1079063211403164

Nunes, K. L., & Babchishin, K. M. (2012). Construct validity of Stable-2000 and Stable-2007 scores. Sexual Abuse: A Journal of Research and Treatment, 24, 29-45. doi:10.1177/1079063211404921

Willis, G. M., & Johnston, L. (2012). Planning helps: The impact of release planning on subsequent re-entry experiences of child sex offenders. *Journal of Sexual Aggression,* 18, 194-208. doi:10.1080/13552600.2010.506576

Preselected High Risk Need

Sowden, J., & Olver, M. (2012, June). *Examining the relationship between treatment readiness, therapeutic change, and recidivism among sex offenders.* CPA Conference, Halifax, NS, Canada.

SRA-FV

Preselected Treatment Need

Knight, R. A., & Thornton, D. (2007). *Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders* (Document No. 217618). Submitted to the U.S. Department of Justice.

Trapold, J. (2012). *Identifying sexually violent interests in various types of sex offenders using items on the Structured Risk Assessment: Forensic Version Light.* Ph.D. thesis. Alliant International University.

Preselected High Risk Need

D'Orazio, D., & Thornton, D. (2011). *Levels of criminogenic need in outpatient and SVP populations.* ATSA presentation, Toronto.

Knight, R. A., & Thornton, D. (2007). *Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders* (Document No. 217618). Submitted to the U.S. Department of Justice.

VRS-SO

Preselected Treatment Need

Beggs, S. M., & Grace, R. C. (2010). Assessment of dynamic risk factors: An independent validation study of the Violence Risk Scale: Sexual Offender Version. *Sexual Abuse: A Journal of Research and Treatment, 22,* 234-251. doi:10.1177/1079063210369014

Kingston, D. A. (2010). The offence progression in sexual offenders: An examination of the self-regulation model of the offence process. Ph.D. thesis. University of Ottawa.

Nunes, K. L., & Cortoni, F. (2008). Assessing treatment change in sexual offenders. Correctional Services Canada.

Preselected High Risk Need

Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale – Sexual Offender Version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19,* 318-329.

Sowden, J., & Olver, M. (2012, June). *Examining the relationship between treatment readiness, therapeutic change, and recidivism among sex offenders.* CPA Conference, Halifax, NS, Canada.

State v. Ronald Love

No. 32555-5-III

Certificate of Service by email

I Patrick Mayovsky, declare under penalty of perjury under the laws of the state of Washington that the following is true and correct:

That on the 26th day of January, 2015, I caused a true and correct copy of the **Brief of Appellant** to be served on the party / parties designated below by email per agreement of the parties pursuant to GR30(b)(4) and/or by depositing said document in the United States mail.

Thomas Howe
Attorney General's Office
thomash1@atg.wa.gov
crjsvpef@atg.wa.gov

Ronald Love
Special Commitment Center
P.O. Box 88600
Steilacoom, WA 98388

Signed in Seattle, Washington this 26th day of January, 2015.

x _Patrick Mayovsky_