

Concurrent and Predictive Validity of the Instrument for Forensic Treatment Evaluation: From Risk Assessment to Routine, Multidisciplinary Treatment Evaluation

Erwin Schuringa^{1,2,3,4,5}, V.E. Heininga^{1,2,3,4,5},
Marinus Spreen^{1,2}, and Stefan Bogaerts^w

Abstract

Besides assessment of forensic patients' risk of future violence and criminogenic needs, knowledge on their responsivity to treatment is equally important. However, instruments currently used for risk assessment are not sensitive enough for treatment evaluation. Therefore, the Instrument for Forensic Treatment Evaluation (IFTE) was developed. The IFTE is a treatment evaluation tool, which uses the dynamic risk items of the Dutch risk assessment tool, the HKT-R (Historical, Clinical, Future-Revised). The IFTE has an extended answering scale, which makes it more sensitive for measuring change and enables clinicians to monitor patients' responsivity to treatment closely. This study examines the concurrent and predictive validity of the IFTE. We found moderate to strong correlations between IFTE items and HKT-30 items (the HKT-30 is the predecessor of the HKT-R), with work and therapy attendance, and positive drug tests. In addition, we found moderate to modest correlations between some IFTE items and work and therapy attendance in a 6-month follow-up period and modest to high discriminative power for some IFTE items for violence and drug use

¹Forensic Psychiatric Centre Dr. S. van Mesdag, Groningen, The Netherlands

²Stenden University of Applied Sciences, Leeuwarden, The Netherlands

³Tilburg School of Social and Behavioral Sciences, The Netherlands

⁴Leuven Institute of Criminology and Forensic Psychiatric Centre Leuven, Belgium

⁵Forensic Psychiatric Centre Kijvelanden, Poortugaal, The Netherlands

Corresponding Author:

Erwin Schuringa, Forensic Psychiatric Centre Dr. S. van Mesdag, Postbox 30.002, 9700 RC Groningen, The Netherlands.

Email: E.Schuringa@fpcvanmesdag.nl

6 months after the measurement. Given its good reliability and validity properties, and comprehensive but short-term nature, implementation of the IFTE in forensic practice likely improves individual treatment of forensic psychiatric patients and has high potential for risk management purposes.

Keywords

Instrument for Forensic Treatment Evaluation (IFTE), concurrent validity, predictive validity, routine outcome monitoring, multidisciplinary

Introduction

In general, forensic treatment is based on the principles of the risk-need-responsivity (RNR) model, which is widely accepted as the most effective approach (Andrews, Bonta, & Hoge, 1990; Polaschek, 2012; Ward, Melsner, & Yates, 2007). The general underlying assumption of the RNR approach is that the intensity of treatment should be appropriate to the risk level of the patient. Furthermore, the nature and extent of the treatment must be geared to the specific needs of the patient and the treatment offered should be tailored to the developmental level of the offender. Finally, the practitioner must be sufficiently responsive to the offender and take into account the learning capacity of the offender; this is called specific responsivity. To assess the level of risk of recidivism, Andrews and Bonta (2010) recommended the use of validated risk assessment schemes, which provide information about specific risks and criminogenic needs of individual patients. However, they do not mention a way to establish specific responsivity. Wooditch, Tang, and Taxman (2014) stated that criminogenic needs should be assessed and re-assessed in short-term intervals to establish any (abrupt) changes that might occur. A way of doing this is by routine outcome monitoring (ROM) using standardized instruments, which not only provides information about criminogenic needs and thus support decision making about treatment goals but also helps to evaluate the condition of a patient and his treatment response (Andrews et al., 1990; Bogaerts, 2010; Knaup, Koesters, Schoefer, Becker, & Puschner, 2009). Using ROM, specific responsivity can be established more objectively. In non-forensic mental health care, ROM systems are commonly used (e.g., Coombs, Stapley, & Pirkis, 2011; Gilbody, House, & Sheldon, 2002) and offer benefits for both patients and practitioners. For example, Knaup et al. (2009) found a largely positive effect on patients' treatment progress, when feedback was given to both patient and clinician at least twice during treatment. In addition, clinicians using ROM information tended to be more effective and more adequate in diagnosing, decision making, and adapting treatment perspective. They were also found to improve their communication with patients (Boswell, Kraus, Miller, & Lambert, 2015; Carlier et al., 2012; Priebe et al., 2007). ROM information also improved the therapeutic alliance, because detection and discussion of slight improvements in treatment may motivate skeptical clients to treatment adherence (Youn, Kraus, & Castonguay, 2012). Furthermore, ROM enhanced effect sizes of treatment and decreased the risk of deterioration of the patient (Anker, Duncan, & Sparks, 2009; Kraus, Castonguay, Boswell, Nordberg, & Hayes, 2011). It is conceivable that applying ROM in forensic psychiatry will be equally beneficial as in regular mental health care. However, the

use of ROM data in forensic mental health care has rarely been implemented; maybe, a lack of a suitable instrument is one of the reasons.

In the following part, criteria will be discussed which an ROM instrument for forensic mental health care should meet, to be useful. In contrast to regular mental health care, the primary goal in forensic mental health care is to reduce the likelihood of future recidivism (Shinkfield & Ogloff, 2014). Periodically measuring the risk of recidivism and making a patient's treatment progress transparent are two core points of focus in forensic psychiatry, which are linked directly to each other. Insights into the level of criminogenic needs, such as the severity of impulsiveness and the severity of hostility are very valuable to direct and, possibly, adjust the treatment. It is, therefore, obvious that a forensic ROM instrument should contain similar dynamic items as risk assessment schemes (Douglas & Kropp, 2002; Lewis, Olver, & Wong, 2013). In a structured review by Chambers and colleagues (2009) about outcome measures used in forensic mental health research, only one study was described in which a risk assessment scheme was used as a routine outcome measure, namely that of Belfrage and Douglas (2002). However, Belfrage and Douglas considered the possibility that the clinical items of the risk assessment scheme, Historical, Clinical, and Risk-20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997) are too broadly conceived to detect precise changes in levels of risk of recidivism. Also, Drieschner and Hesper (2008) concluded that for many forensic psychiatric patients, changes of behavior from "not present" to "present" are unlikely. For example, the items of the HCR-20 are scored as "present," "possibly present," "not present" (Webster et al., 1997). Commonly speaking, risk assessment instruments are not designed to measure small behavioral changes.

A forensic ROM instrument should be able to detect changes in risk behaviors with sufficient sensitivity and specificity, even when measurements are repeated in a relatively short interval of 6 months (Schuringa, Spreen, & Bogaerts, 2014; Wong, Gordon, & Gu, 2007; Wooditch et al., 2014; Youn et al., 2012). A 5-point Likert-type scale is not suitable for this purpose unless multiple items are used to measure one construct (Drieschner & Hesper, 2008). However, multiple items to measure one construct will lead to time-expensive unpractical instruments. Larger scales, which are more subtle (10-points or larger Likert-type scales) offer advantages above crude ones: They are more likely to be normally distributed, more sensitive to detect important minimal change, and more beneficial to researchers by producing more accurate data (Leung, 2011; Pearse, 2011).

Doyle and Logan (2012) specified that by completing a risk assessment, the awareness of risk factors is heightened for practitioners. According to them, monitoring of these factors should be done by the client and others engaged in the treatment (nurses, therapists, psychiatrists, etc.). However, if the client is lacking insight into his own behavior or if his motivation for monitoring is limited, then the information of others becomes more important. The Violence Reduction Program (Wong & Gordon, 2013) also points out that an integrated approach of risk monitoring will result in a more complete picture of a patient's behavior over time by observing the patient in the broader context, such as the therapeutic living environment, the behavior at work, sports, leisure, and cooking. Because risk behavior is dynamic in nature, repeated treatment evaluation by various professionals is designated to optimize treatment (Douglas & Kropp, 2002; Lewis et al., 2013).

To summarize, the use of ROM data in forensic mental health care can be of much benefit. By extending clinical scales of risk assessment schemes to make them more sensitive to changes in criminogenic needs, they then can be applied as a forensic ROM instrument. To tailor treatment and to measure behavioral change more objectively, a

multidisciplinary and repetitive approach in the assessment of a patient's criminogenic needs across time is highly recommended. The Instrument for Forensic Treatment Evaluation (IFTE; Schuringa et al., 2014) is one of the first instruments developed within forensic mental health care based on such an approach. By closely evaluating the criminogenic needs of a patient, specific responsivity to treatment can be monitored.

In this study, after a description of the IFTE, we report on the criterion validity of the IFTE by examining its concurrent and predictive validity. Concurrent validity will be determined by comparing the IFTE with the Dutch risk assessment scheme, Historische Klinische Toekomst-30 (HKT-30; Historical, Clinical, Future-30; Workgroup Risk Assessment Forensic Psychiatry, 2002), and with variables collected by the administration department during the same period as the IFTE assessment: work attendance, therapy attendance, and (illegal) drug test outcomes. These variables were independently collected from the IFTE evaluations. Predictive validity will be examined by comparing the degree of work and therapy attendance in the period of 5 to 7 months after the IFTE assessment, with IFTE scores. Furthermore, discriminative power of the IFTE for future violence and positive illegal drug test outcomes is examined. Finally, the implications of implementing IFTE in forensic mental health care are discussed.

Method

Setting and Sample

Data for this study were drawn from the ROM system of the Forensic Psychiatric Center Dr. S. van Mesdag in Groningen, the Netherlands, a maximum-security hospital for mentally disordered offenders, who are hospitalized under the judicial measure of "TBS-order." This order is a "provision in the Dutch criminal code that allows for a period of treatment following a prison sentence for mentally disordered offenders" (van Marle, 2002, p. 83). The hospital has approximately 250 residential male offenders. Due to outplacement and new entries in the institution, in the period April 2010 until October 2014, 277 patients were included in the current study (see Table 1).

Table 1. Description of the Sample.

Sample		Index offences	
Number of patients	277	Homicide	105 (38%)
Age (years)	36.7	Violence	37 (13%)
SD	9.6	Sexual offence	70 (25%)
Range	20-68	Theft with and without violence	17 (6%)
	43.6	Arson	19 (7%)
Mean time of admission (months)		Other	29 (10%)
SD	36.4		
Range	2-203		
Diagnoses³			
Axis 1		Axis 2	
Schizophrenia or other psychotic disorder	134 (48%)	Cluster A Personality disorder	7 (3%)
Mood and anxiety disorder	33 (12%)	Cluster B Personality disorder	131 (47%)
Development disorder	69 (25%)	Cluster C Personality disorder	17 (6%)
Substance abuse	363 ^b	Personality disorder NOSc	86 (31%)
Pedophilia/paraphilia	67 (24%)	Postponed	1 (0%)
Other	20 (7%)	Mental retardation	33 (12%)
Number of patients with at least one substance (ab) use-related diagnosis	218 (79%)	Other	44 (16%)

^aDiagnoses according to the *Diagnostic and Statistical Manual of Mental Disorders* (4 th ed., text rev.; DSM- IV-TR; American Psychiatric Association, 2000).

^bThe number of substance abuse diagnoses is larger than the total population, because some patients had multiple substance abuse diagnoses. ^cNot otherwise specified

Mean age on intake in the institution was 36.7 years (range = 20-68, $SD = 9.6$) and mean duration of hospitalization until the measurement was 43.6 months (range = 2-203, $SD = 36.4$). Forty-eight percent of this population had a diagnosis of schizophrenia or another psychotic disorder. Forty-seven percent had a cluster B personality disorder and 31% a personality disorder not otherwise specified. Seventy-nine percent of this population had at least one substance use-related diagnosis. A lot of co-morbidity existed in this population, resulting in a mean amount of 3.6 diagnosis per patient (range = 1-6, $SD = 1.3$). For the study of concurrent validity, a smaller subsample of 232 patients was used, because the institution stopped collecting administrative data in January 2013.

Instruments

The HKT-30 was until January 2015 the most used risk assessment scheme for adults in forensic psychiatry in the Netherlands. The HKT-30 uses structured professional judgment to establish the risk of future violence. The historical scale consists of 11 items, the clinical

scale of 13, and the future scale counts six items. All items are scored on a 5-point scale, ranging from 0 (*low risk*) to 4 (*high risk*). The score is obtained by consensus between the treatment coordinator and an independent rater. The HKT-30 has been rated as being very useful (Singh et al., 2014) and has good inter-rater reliability, sufficient to good internal consistency, and sufficient predictive validity (Blok, de Beurs, de Ranitz, & Rinne, 2010; Hildebrand, Hesper, Spreen, & Nijman, 2005; Workgroup Risk Assessment Forensic Psychiatry, 2002). The HKT-30 has been revised into the HKT-R (Spreen, Brand, ter Horst, & Bogaerts, 2014), which is legally obliged since January 2015. The HKT-R was recently validated in a nationwide saturation sample of 347 forensic patients, who were released into the community from maximum-security forensic psychiatric hospitals in the period 2004-2008. The psychometric results from this study showed sufficient inter-rater reliability, good internal consistency, and good predictive validity of the HKT-R for most forensic target groups (Spreen et al., 2014).

The IFTE is the ROM tool based on the clinical and future subscales of the HKT-R. The IFTE was specifically designed to measure the progress of behaviors and insights over time to support treatment decisions. In a psychometric study among 232 forensic psychiatric male patients, the IFTE showed good internal consistency (factor Protective Behaviors = .90, factor Problematic Behaviors = .86, factor Resocialization Skills = .88), test-retest reliability (a range = .62-.91), and inter-rater reliability (intra-class correlations range = .65-.92; Schuringa et al., 2014).

The IFTE comprises 22 observational behavior items (see Table 2), including all 14 clinical items of the HKT-R. The Atascadero Skills Profile (Vess, 2001) inspired three additional items: “skills to prevent drug use,” “skills to prevent physically aggressive behavior,” and “skills to prevent to sexually deviant behavior.” In collaboration with clinicians, five other items, which they found important for treatment evaluation, were added: “sexually deviant behavior,” “manipulative behavior,” “financial skills,” “balanced daytime activities,” and “medication use.” Other studies have already shown the importance of mainly antipsychotic medication use in preventing violence by forensic patients with a severe mental illness (Hodgins & Riaz, 2011; Swartz et al., 1998a, 1998b). To sensitively detect behavioral changes in rather short measurement periods, the IFTE items can be scored on a 17-point scale with five anchor points (ranging from 0 to 4) describing characteristic behaviors in global terms. The scale allows raters to depict a score between the five anchor points (for instance a 1+, 1.5, or 2-) (see Figure 1).

The 22 items are divided into three factors (see Table 2). The factor Protective Behaviors pertains to behaviors that are related to risk reducing prosocial behaviors and skills, the factor Problematic Behaviors are typical “forensic” risk behaviors, and **Table 2**. Items of the IFTE.

Protective Behaviors

- Does the patient show problem insight?
- Does the patient cooperate with your treatment?
- Does the patient admit and take responsibility for the crime(s)?
- Does the patient show adequate coping skills?
- Does the patient use his medication in a consistent and adequate manner?
- Does the patient show skills to prevent drug and alcohol use?
- Does the patient show skills to prevent physically aggressive behavior?

Does the patient show skills to prevent sexually deviant behavior?

- Does the patient show impulsive behavior?
- Does the patient show antisocial behavior?
- Does the patient show hostile behavior?
- Does the patient show sexually deviant behavior?
- Does the patient show manipulative behavior?
- Does the patient comply with the rules and conditions of the center and/or the treatment? Does the patient have antisocial associates?
- Does the patient have psychotic symptoms?
- Does the patient use any drugs or alcohol?

Resocialization Skills

- Does the patient have balanced daytime activities?
- Does the patient show sufficient labor skills?
- Does the patient show sufficient common social skills?
- Does the patient show sufficient skills to take care of oneself?
- Does the patient show sufficient financial skills?

Note. IFTE = Instrument for Forensic Treatment Evaluation.

the factor Resocialization Skills is characterized by those behaviors that individuals need to “survive” in society. For a member of a treatment team, it takes approximately 10 min to complete an IFTE.

The IFTE is independently completed by all therapists involved in the treatment of one patient, including the psychologist, psychiatrist, nurses in the ward, psychomotor therapist, work therapists, and skills trainers. All data are summarized in a report, which displays graphically the level of functioning of the patient on the three factors and all items separately, the level of agreement between all raters per item, and the level of change per item compared with the last and to the initial measurement. The level of functioning provides information about criminogenic need factors. The level of agreement provides insight into how the behavior is generalized in

Someone with problem insight has insight in his own mental processes and their influence on his behavior. A patient with problem awareness is troubled with the problems his behavior causes (he realizes he has a problem), but he has no insight in what causes his behavior or how he could influence *N.E.I.* his behavior.

0				1				2				3				4
None			Rarely			Sometimes			Often			Always				

Does the patient show problem insight?

- 1 No problem insight and **minor** problem awareness.
- 2 No problem insight. He has problem **awareness**, but does not behave accordingly.
- 3 **Some** problem insight. He does no always behave accordingly.
- 4 He has **sufficient** problem insight and behaves accordingly.

Figure 1. Example of an IFTE item.

0 No problem insight and no problem awareness, does not accept external control.

Note. IFTE = Instrument for Forensic Treatment Evaluation; N.E.I. = not enough information.

different situations and the level of change provides information about the respon- sivity of the patient.

Statistical Procedures

To measure concurrent validity, Kendall’s tau (τ) was used to examine the relationship between the IFTE items and the corresponding 12 dynamic risk items of the HKT-30 (Arndt, Turvey, & Andreasen, 1999). The HKT-30 was used, because at the time of this study, the HKT-R had not been implemented yet. Only patients were included, whose IFTE observation period of 6 months was completely overlapped by the 12-month observation period of the HKT-30. A treatment coordinator and an independent researcher scored the HKT-30. All team members, including the treatment coordinator filled out the IFTE.

Kendall’s tau was also used to examine the relationship between the IFTE and work attendance, therapy attendance, and illegal drug use for the same period. Work attendance and therapy attendance were defined as the percentage of actual attended hours compared with scheduled hours. Drug use was determined by counting the number of positive outcomes of urine tests on THC (tetrahydrocannabinol, the psychoactive

ingredient of cannabis). The variable drug use was divided into no or single drug use, and multiple usage. Following the classifications of Cohen (1988), $\tau < .10$ was used to indicate a weak correlation, $.10 \geq \tau < .29$ a moderate correlation, $.30 \geq \tau < .49$ a modest correlation, and $\tau \geq .50$ indicated a strong correlation.

The predictive validity was studied by examining the relationship between the IFTE and work and therapy attendance for the 6-month period after the IFTE measurement, using Kendall's tau. In addition, by means of calculating the discriminative power of the IFTE, the predictive validity of the IFTE was tested with regard to future violence and illegal drug use. Violence was defined as intentional behavior, which could or did physically harm a person or animal, and/or a form of aggression, which is extremely intimidating or threatening (Troquete et al., 2013). Illegal drug use was established through urine testing or by patients admitting to illegal drug use, and the variable was dichotomized in patients who did not use illegal drugs, or just once and those who used repeatedly. The time-at-risk was 4 to 8 months. Mann-Whitney tests were performed to analyze whether IFTE items discriminated between violators and non-violators. For the IFTE items which were discriminative ($p < .05$), the effect size was calculated using Cohen's d with a pooled standard deviation, because of the different sample sizes. A d value equal or larger than .80 was considered large (Cohen, 1988).

Results

Concurrent Validity

Table 3 displays the Kendall's tau correlation between IFTE items and corresponding HKT-30 items.

The 22 IFTE items all had modest to strong correlations with their corresponding items of the HKT-30. The three skill items for preventing drug use, physically aggressive behavior, and sexually deviant behavior had modest to moderate correlations with the HKT-30 item "coping skills," which was defined as how adequate a patient confronted with interpersonal or practical problems or situations that require adjustments can integrate and/or solve these problems and situations in a satisfactory way. The IFTE items "balanced daytime activities" and "financial skills" had a modest correlation with the HKT-30 item "skills to take care of oneself." The IFTE item "medication use" had a modest correlation with the HKT-30 item "problem insight." The IFTE item "manipulative behavior" had a modest correlation with the HKT-30 item "hostility," and the IFTE item "antisocial associates" had a moderate correlation with the HKT-30 item "drug use."

Table 4 displays the correlations between IFTE items and other variables collected during the same period (the first two columns with the name, "Same").

For the outcome variable work attendance, most correlations were modest with the exception of "cooperation with treatment," which had a moderate correlation. In addition, "skills to prevent drug use" and the factor Problematic Behaviors had modest correlations with registered illegal drug use. The item "drug use" had showed a strong correlation with registered drug use.

Table 3. Concurrent Validity Between Items of the IFTE and the HKT-30.

IFTE	HKT-30	Kendall's τ (n)
Protective Behaviors		
Problem insight	Problem insight	.60** (162)
Cooperation with treatment	Attitude toward treatment	.49** (162)
Take responsibility for the crime	Admit and take responsibility for the crime	.49** (160)
Coping skills	Coping skills	.52** (161)
Medication use	Problem insight	.46** (135)
Skills to prevent drug and alcohol use	Drug use	.30** (114)
	Coping skills	.32** (113)
Skills to prevent physically aggressive behavior	Coping skills	.48** (115)
Skills to prevent sexually deviant behavior	Coping skills	.25* (57)
Problematic Behaviors		
Impulsive behavior	Impulsivity	.52** (162)
Antisocial behavior	Empathy	-.28** (162)
	Social and relational skills	-.42** (162)
Hostile behavior	Hostility	.42** (162)
Sexually deviant behavior	Sexual preoccupation	.31** (160)
Manipulative behavior	Hostility	.35** (161)
Compliance to rules	Attitude toward treatment	.42** (162)
Antisocial associates	Drug use	.29** (158)
Psychotic symptoms	Psychotic symptoms	.65** (137)
Drug use	Drug use	.62** (133)
Resocialization Skills		
Balanced daytime activities	Skills to take care of oneself	.46** (162)
Labor skills	Skills to take care of oneself	.44** (162)
Social skills	Social and relational skills	.51** (162)
Skills to take care of oneself	Skills to take care of oneself	.46** (162)
Financial skills	Skills to take care of oneself	.42** (161)

Note. For interpretation reasons, the direction of the correlations were made positive, while with HKT-30, items of a high score implies high risk and a high score on IFTE means more observed behavior. For example, "Problem Insight," a high score on the IFTE means a lot of problem insight, whereas with the HKT-30, a high score means no insight at all. IFTE = Instrument for Forensic Treatment Evaluation; HKT = Historische Klinische Toekomst.

* $p < .05$. ** $p < .01$.

Predictive Validity

The last two columns named "Future" in Table 4 display correlations between IFTE items and other variables measured during the follow-up period after the IFTE measurement. "Cooperation with treatment," "balanced daytime activities," and the factor

Table 4. Concurrent and Predictive Validity of IFTE and Other Variables.

	Work attendance	Therapy attendance	Work attendance	Therapy attendance
IFTE	Same (M = 92%, Same (M = 98%, Future (M = 95%, range = 18-100, SD range = 72-100, SD range = 49-100, SD = 14.9; N) = 4.2; N)		Future (M = 97%, range = 75-100, SD = 3.0; N)	
Cooperation with treatment	.211 (142)	-.09 (158)	.16 (60)	.2223 (61)
Balanced daytime activities	.35** (142)	-.08 (158)	.38** (60)	.25** (61)
Labor skills	.33** (141)	.34** (142)		
Factor Resocialization Skills			.29** (60)	.31** (60)
				.21* (61)
	Drug use			
	Same (M = 0.55, range = 0-11, SD = 1.59; N)			
Skills to prevent drug and alcohol use	-.38** (173)			
Drug use Factor	.59** (202)			
Problematic Behaviors	.24** (225)			

Resocialization Skills showed moderate correlations with future therapy attendance. “Balanced daytime activities,” “labor skills,” and the factor Resocialization Skills showed moderate to modest correlations with future work attendance.

Table 5 shows the results of the Mann-Whitney tests and the effect size for patients who committed violence and those who did not, and between patients using drugs and the patients with no or single drug use. Of the 277 patients, 53 (19%) engaged in violence and 11 of 56 (19%) used illegal drugs more than once.

For violence, “skills to prevent physically aggressive behavior” had large discriminative power. In addition, the factor Problematic Behaviors and its items “impulsive behavior,” “antisocial behavior,” and “hostile behavior” had large discriminative power. For illegal drug use, the item “skills to prevent drug use” and the factor Problematic Behaviors and its items “impulsive behavior,” “compliance to rules,” and “drug use” had large discriminative power.

Note. IFTE = Instrument for Forensic Treatment Evaluation.
 $3p < .05$. ** $p < .01$.

Table 5. Mann-Whitney Test and Effect Size of Violators Versus Non-Violators.

	Violence				Illegal drug use			
	No		Yes		No		Yes	
	(n = 224)		(n = 53)		M		("= M)	
IFTE	AT (SD; N ^b)	M (SD; N)	P	d	M (SD; N)	M (SD; N)	P	d
Factor Protective Behaviors	1 1.56 (2.61)	10.02 (2.76)	.000		.5713.34 (1.98)	12.22 (1.88)	.049	-.57
Problem insight	10.31 (3.51)	8.99 (2.95)	.004		.4112.33 (2.54)	1 1.83 (2.76)	.665	
Cooperation with treatment	1 1.43 (3.27)	9.69 (3.17; 52)	.000		.5413.12 (2.70)	11.10 (2.39)	.015	-.76
Take responsibility for the crime	10.56 (3.85)	10.00 (3.86)	.344		12.36 (3.26)	12.41 (3.13)	.992	
Coping skills	9.98 (2.63)	8.15 (2.91)	.000		.661 1.65 (2.12)	10.43 (2.32)	.103	
Medication use	13.33 (3.70; 180)	12.34 (3.89; 43)	.064		15.13 (1.76; 37)	14.17(2.06; 10)	.134	
Skills to prevent drug use	12.72 (3.49; 187)	10.38 (4.63; 46)	.003		.5714.46 (2.31; 41)	10.57 (2.62; 1 1)	.001	-1.64
Skills to prevent physically aggressive behavior	13.84 (2.98; 202)	1 1.23 (3.32; 50)	.000		.8315.25 (1.39; 41)	14.43 (1.99; 1 1)	.260	
Skills to prevent sexually deviant behavior	11.01 (4.16; 105)	10.26 (4.10; 19)	.466		1 3.61 (2.97; 20)	17.00 (0.00; 2)	.039	
Factor Problematic Behaviors	4.48 (1.80)	6.66 (2.24)	.000	-1.07	3.35 (1.33)	5.26 (1.15)	.000	1.47
Impulsive behavior	7.40 (3.21)	9.99 (3.07)	.000	-.82	6.03 (2.88)	8.22 (1.97)	.021	.80
Antisocial behavior	5.39 (2.64)	8.53 (3.16)	.000	-1.08	4.00 (2.09)	5.71 (4.44)	.013	.64

(continued)

Table 5. (continued)

(n = 224)_	(n = 53)_	Violence				Illegal drug use			
		No	Yes	(« = Γ')		No	Yes		
		(SD; N ^b)	M (SD; N)	P	d	M (SD; N)	M (SD; N)	P	d
IFTE									
Hostile behavior		4.74 (2.48)	7.58(3.18)	.000	-1.00	3.24 (1.80)	4.66 (1.88)	.027	.78
Sexually deviant behavior		2.18 (1.59)	2.87 (2.32)	.162		2.08 (1.94)	1.98 (1.67)	.559	
Manipulative behavior		5.28 (2.78)	7.38 (3.69)	.000	-.64	3.70 (2.37)	3.94 (2.25)	.650	
Compliance to rules		13.08 (2.23)	10.98 (3.09)	.000	-.78	14.39 (1.98)	11.00 (2.34)	.000	-1.65
Antisocial associates		3.16 (2.36)	5.39 (3.83; 52)	.000	-.70	2.49 (1.61)	3.63 (1.71)	.023	.70
Psychotic symptoms		3.36 (2.87; 175)	4.62 (4.61; 39)	.381		2.37 (1.73; 37)	3.18 (2.14; 8)	.449	
Drug use		3.55 (4.11; 204)	5.93 (5.21; 48)	.020	-.51	2.10 (2.58; 42)	8.10 (4.39; 11)	.001	2.00
Factor Resocialization Skills		11.84 (2.78)	10.70 (2.96)	.012	.40	12.71 (2.00)	12.26 (2.23)	.789	
Balanced daytime activities		11.36 (3.43)	9.81 (4.04)	.009	.41	12.60 (2.28)	11.28 (3.66)	.439	
Labor skills		12.08 (4.02)	10.35 (4.52; 52)	.006	.40	13.80 (2.66; 43)	11.90 (3.25; 11)	.049	-.68
Social skills		10.65 (2.74)	9.00 (2.58)	.000	.62	11.59 (2.43)	12.01 (1.24)	.364	
Skills to take care of oneself		13.13 (3.21)	12.77 (3.64)	.629		13.37 (3.06)	13.95 (2.55)	.665	
Financial skills		12.02 (3.49; 223)	11.63 (4.06)	.740		12.32 (2.70)	12.15 (3.22)	.788	

Note. IFTE = Instrument for Forensic Treatment Evaluation; *p* = significance level; *d* = Cohen's *d* with pooled SD. ^aM on 17-point scale. ^bN = smaller N.

Discussion

In the first IFTE study, this ROM instrument showed good inter-rater and test-retest reliability (Schuringa et al., 2014). This current study was conducted to test concurrent validity and predictive validity of the IFTE. Moderate to strong correlations of the IFTE were found with the clinical items of the HKT-30, and other variables measured in the same period: work attendance, therapy attendance, and illegal drug use. Moderate to modest correlations were found between some IFTE items and other variables measured in the follow-up period: work attendance, therapy attendance, and illegal drug use. Some IFTE items and the factor Problematic Behaviors were found to have a large discriminative power for subsequent violence and drug use.

With regard to concurrent validity on item level, the modest correlations with items of the HKT-30, although the items are basically similar, could be due to small differences between both tools. The score on the HKT-30 was based on behaviors observed in the previous 12 months, whereas the IFTE covered a 6-month period. The HKT-30 item score was based on consensus between two raters, whereas the IFTE item score was the mean score of all raters. In addition, the HKT-30 items had a 5-point scale and the IFTE items had a 17-point scale.

The HKT-30 item “skills to take care of oneself correlated modestly with the following items of the IFTE: “balanced daytime activities,” “labor skills,” “skills to take care of oneself,” and “financial skills.” The HKT-30 item “coping skills” correlated with the following IFTE items: “coping skills,” “skills to prevent drug or alcohol use,” “skills to prevent physically aggressive behavior,” and “skills to prevent sexually deviant behavior.” These two dynamic risk items of the HKT-30 are thus represented in more detail in the IFTE. This is beneficial for treatment purposes, because treatment can then be aimed more precisely. Instead of just enhancing coping skills, it is now more clear which coping skills to enhance. The results show that the three added IFTE skills items also have clear correlations with the dynamic risk items. Other items of the IFTE, which were not present in the HKT-30, still showed modest correlations with dynamic risk items and, therefore, are forensically relevant for treatment purposes. Overall, we can conclude that the items of IFTE are sufficiently associated with the dynamic items of the risk assessment instrument HKT-30.

Concurrent validity was also examined using other variables: work attendance, therapy attendance, and drug use. The items “cooperation with treatment” and “balanced daytime activities” were associated with work attendance but not with therapy attendance. A reason for this, may be that therapy attendance was just a small part of daytime activities and, thus, also a small part for the measurement of treatment cooperation. This is also reflected in the fact that patients can work for a maximum of 20 hr per week, and therapy is offered up to maximum of 4 to 6 hr per week. Furthermore, the item “labor skills,” and the factor Resocialization Skills had significant correlations with work and therapy attendance. The items “skills to prevent drug use,” “drug use,” and the factor Problematic Behaviors had significant correlations with registered drug use. These results emphasize the relevance of the IFTE for treatment evaluation processes, because observed behavior scored on the IFTE is also reflected in these administratively collected data.

The degree of attendance to work and therapy 6 months after the IFTE measurement was used to examine predictive validity. Only the item “balanced daytime activities” and the

factor Resocialization Skills had modest correlations with the degree of work attendance.

Predictive validity of the IFTE items was also studied by monitoring violence and illegal drug use during a 4- to 8-month follow-up period as an outcome variable. The results showed that “skills to prevent physically aggressive behavior” are protective for violence. In addition, patients who committed violence showed more problematic behavior in the prior period than patients who did not commit violence: The items “antisocial behavior” and “hostile behavior” showed large predictive power for violence. The importance of predictive validity for short-term violence was already established in studies of the Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Desmarais, 2009; Desmarais, Nicholls, Wilson, & Brink, 2012; O’Shea, Picchioni, & Dickens, 2016; Troquete et al., 2015) and is a crucial quality of any forensic ROM instrument. This study showed that the IFTE has good predictive validity for violence. Van der Veeken, Lucieer, and Bogaerts (2016) also found that the IFTE showed good predictive validity for inpatient aggression, and marginal to reasonable predictive validity for leave approvals and drug use in the short term, in another forensic psychiatric center in the Netherlands.

This study also, not surprisingly, showed that “skills to prevent drug use” had discriminative power for drug use in the short term. In addition, the factor Problematic Behaviors had discriminative power for drug use. For treatment purposes, it is very informative that “skills to prevent drug use” seems protective for future drug use. The best predictor in this study for future drug use was past drug use as measured with the IFTE. For validity purposes, this is a very satisfactory outcome; the item is relevant for treatment evaluation purposes. From a risk management and treatment perspective, the result is remarkable, because after illegal drug use, treatment and risk management interventions are imposed to prevent future drug use. These results suggest a lack of efficacy of these measures.

Limitations and Strengths

This study has not only many strengths but also some limitations. Some of the strengths are the large and heterogeneous diagnosis in the group of patients included in this study, the long period during which prospective measurements were conducted, and the naturalistic design of the study. The IFTE is an integral part of the treatment evaluation and, therefore, is filled out by multiple and experienced therapists instead of researchers. A limitation of the study is the single site design. Generalizations to other institutions or other countries should be done with care. However, the IFTE is derived from the HKT-R, which was recently validated in a multi-center study, and therefore, we expect no significant difficulties in implementing the IFTE in other forensic psychiatric institutions, which is supported by the van der Veeken et al. (2016) study. This patient group was heterogeneous with respect to diagnosis and it is possible that results found for this patient group will differ, if the group is divided by diagnosis. Nevertheless, we expect that some correlations might even be stronger for more homogeneous groups, than what we encountered with this heterogeneous group. For example, medication use could be much more important for patients with schizophrenia than for patients with a personality disorder.

In the future, studies should focus on sensitivity and specificity of the IFTE for different diagnoses and for different outcomes, such as short-term violence and reoffending after treatment, so the IFTE can aid treatment and risk management even more. In addition, future research should also focus on the link between patients' changing profile scores over time and the potential change in likelihood for re-offending, to further establish validity of the IFTE as a forensic treatment evaluation instrument.

Conclusion

Taken together, the results demonstrate that the IFTE is a useful multidisciplinary forensic psychiatric treatment evaluation instrument and, thus, capable of monitoring responsiveness. Because the IFTE showed modest to high concurrent and short-term predictive validity when using the parameters available in this specific study, the instrument likely has high potential for risk management purposes in other institutions also. Replication is warranted, but given its good reliability properties (Schuringa et al., 2014), and given its comprehensive but short-term nature, implementation of IFTE in forensic practice likely improves individual treatment of forensic psychiatric patients. By doing so, we could move beyond the one-size-fits-all approach, and move toward a more tailored and, therefore, a presumably more effective "personalized" approach.

Acknowledgments

The authors thank Miss H. Visser, MSc, for providing language help.

Declaration of Conflicting Interests

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

Funding

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

References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law*, *16*, 39-55. doi:10.1037/a0018362
- Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior*, *17*, 19-52. doi:10.1177/0093854890017001004
- Anker, M. G., Duncan, B. L., & Sparks, J. A. (2009). Using client feedback to improve couple therapy outcomes: A randomized clinical trial in a naturalistic setting. *Journal of Consulting and Clinical Psychology*, *77*, 693-704. doi:10.1037/a0016062
- Arndt, S., Turvey, C., & Andreasen, N. C. (1999). Correlating and predicting psychiatric symptom ratings: Spearman's r versus Kendall's tau correlation. *Journal of Psychiatric Research*, *33*, 97-104. doi:10.1016/S0022-3956(98)90046-2
- Belfrage, H., & Douglas, K. S. (2002). Treatment

- effects on forensic psychiatric patients measured with the HCR-20 violence risk assessment scheme. *International Journal of Forensic Mental Health*, 1, 25-36. doi:10.1080/14999013.2002.10471158
- Blok, G. T., de Beurs, E., de Ranitz, A. G. S., & Rinne, T. (2010). Psychometrische stand van zaken van risicotaxatie-instrumenten voor volwassenen in Nederland [Psychometric state-of-the-art of risk assessment instruments for adults in the Netherlands]. *Tijdschrift voor Psychiatrie*, 52, 331-341.
- Bogaerts, S. (2010). Emerging international perspectives in forensic psychology: Individual level analysis. *Journal of Forensic Psychology Practice*, 10, 263-266. doi:10.1080/15228932.2010.481229
- Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research*, 25, 6-19. doi:10.1080/10503307.2013.817696
- Carlier, I. V., van Meuldijk, D., van Vliet, I. M., van Fenema, E. M., van der Wee, N. J. A., & Zitman, F. G. (2012). Empirische evidentie voor de effectiviteit van routine outcome monitoring; een literatuuronderzoek [Empirical evidence for the effectiveness of routine outcome monitoring: A literature review]. *Tijdschrift voor Psychiatrie*, 54, 121-128.
- Chambers, J. C., Yiend, J., Barrett, B., Burns, T., Doll, H., Fazel, S., . . . Fitzpatrick, R. (2009). Outcome measures used in forensic mental health research: A structured review. *Criminal Behaviour and Mental Health*, 19, 9-27. doi:10.1002/cbm.724
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Coombs, T., Stapley, K., & Pirkis, J. (2011). The multiple uses of routine mental health outcome measures in Australia and New Zealand: Experiences from the field. *Australasian Psychiatry*, 19, 247-253. doi:10.3109/10398562.2011.562507
- Desmarais, S. L., Nicholls, T. L., Wilson, C. M., & Brink, J. (2012). Using dynamic risk and protective factors to predict inpatient aggression: Reliability and validity of START assessments. *Psychological Assessment*, 24, 685-700. doi:10.1037/a0026668
- Douglas, K. S., & Kropp, P. R. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behavior*, 29, 617-658. doi:10.1177/009385402236735
- Doyle, M., & Logan, C. (2012). Operationalizing the assessment and management of violence risk in the short-term. *Behavioral Sciences & the Law*, 30, 406-419. doi:10.1002/bsl.2017
- Drieschner, K. H., & Hesper, B. L. (2008). *Dynamic risk outcome scales*. Boschoord, The Netherlands: Trajectum.
- Gilbody, S. M., House, A. O., & Sheldon, T. A. (2002). Outcomes research in mental health. Systematic review. *The British Journal of Psychiatry*, 181, 8-16. doi: 10.1192/bjp.181.1.8
- Hildebrand, M., Hesper, B. L., Spreen, M., & Nijman, H. L. I. (2005). *De waarde van gestructureerde risicotaxatie en van de diagnose psychopathie: een onderzoek naar de betrouwbaarheid van de HCR-20, HKT-30 en PCL-r* [The value of structured risk assessment and the diagnosis psychopathy: A study of the reliability of the HCR-20, HKT-30 and PCL-r]. Utrecht, The Netherlands: Expertisecentrum Forensische Psychiatrie.
- Hodgins, S., & Riaz, M. (2011). Violence and phases of illness: Differential risk and predictors. *European Psychiatry*, 26, 518-524. doi:10.1016/j.eurpsy.2010.09.006
- Knaup, C., Koesters, M., Schoefer, D., Becker, T., & Puschner, B. (2009). Effect of feedback of treatment outcome in specialist mental healthcare: Meta-analysis. *The British Journal of Psychiatry*, 195, 5-21. doi:10.1192/bjp.bp.108.053967
- Kraus, D., Castonguay, L., Boswell, J., Nordberg, S., & Hayes, J. (2011). Therapist effectiveness: Implications for accountability and patient

- care. *Psychotherapy Research*, 21, 267276. doi:10.1080/10503307.2011.563249
- Leung, S.-O. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *Journal of Social Service Research*, 37, 412-421. doi:10.1080/014 88376.2011.580697
- Lewis, K., Olver, M. E., & Wong, S. C. (2013). The Violence Risk Scale: Predictive validity and linking changes in risk with violent recidivism in a sample of high-risk offenders with psychopathic traits. *Assessment*, 20, 150-164. doi: 10.1177/1073191112441242
- O'Shea, L. E., Picchioni, M. M., & Dickens, G. L. (2016). The predictive validity of the Short-Term Assessment of Risk and Treatability (START) for multiple adverse outcomes in a secure psychiatric inpatient setting. *Assessment*, 23, 150-162. doi:10.1177/1073191115573301
- Pearse, N. (2011). Deciding on the scale granularity of response categories of Likert type scales: The case of a 21-point scale. *Electronic Journal of Business Research Methods*, 9, 159-171.
- Polaschek, D. L. L. (2012). An appraisal of the Risk-Need-Responsivity (RNR) model of offender rehabilitation and its application in correctional treatment. *Legal and Criminological Psychology*, 17, 1-17. doi:10.1111/j.2044-8333.2011.02038.x
- Priebe, S., McCabe, R., Bullenkamp, J., Hansson, L., Lauber, C., Martinez-Leal, R., . . . Wright, D. J. (2007). Structured patient-clinician communication and 1-year outcome in community mental healthcare: Cluster randomised controlled trial. *The British Journal of Psychiatry*, 191, 420-426. doi: 10.1192/bjp.bp.107.036939
- Schuringa, E., Spreen, M., & Bogaerts, S. (2014). Inter-rater and test-retest reliability, internal consistency, and factorial structure of the Instrument for Forensic Treatment Evaluation. *Journal of Forensic Psychology Practice*, 14, 127-144. doi:10.1080/15228932.2014.897536
- Shinkfield, G., & Ogloff, J. (2014). A review and analysis of routine outcome measures for forensic mental health services. *International Journal of Forensic Mental Health*, 13, 252-271. doi:10.1080/14999013.2014.939788
- Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., . . . Otto, R. K. (2014). International perspectives on the practical application of violence risk assessment: A global survey of 44 countries. *International Journal of Forensic Mental Health*, 13, 193-206. doi:10.1080/14999013.2014.922141
- Spreen, M., Brand, E., ter Horst, P., & Bogaerts, S. (2014). *Handleiding HKT-R* [Manual of the HKT-R]. Groningen, The Netherlands: Stichting FPC Dr. S. van Mesdag.
- Swartz, M. S., Swanson, J. W., Hiday, V. A., Borum, R., Wagner, H. R., & Burns, B. J. (1998a). Taking the wrong drugs: The role of substance abuse and medication noncompliance in violence among severely mentally ill individuals. *Social Psychiatry & Psychiatric Epidemiology*, 33, 75-80.
- Swartz, M. S., Swanson, J. W., Hiday, V. A., Borum, R., Wagner, H. R., & Burns, B. J. (1998b). Violence and severe mental illness: The effects of substance abuse and nonadherence to medication. *The American Journal of Psychiatry*, 155, 226-231.
- Troquete, N. A. C., Van den Brink, R. H. S., Beintema, H., Mulder, T., van Os, T. W. D. P., Schoevers, R. A., & Wiersma, D. (2013). Risk assessment and shared care planning in outpatient forensic psychiatry: Cluster randomised controlled trial. *The British Journal of Psychiatry*, 202, 365-371. doi:10.1192/bjp.bp.112.113043
- Troquete, N. A. C., Van den Brink, R. H. S., Beintema, H., Mulder, T., van Os, T. W. D. P., Schoevers, R. A., & Wiersma, D. (2015). Predictive validity of the Short-Term Assessment of Risk and Treatability for violent behavior in outpatient forensic psychiatric patients. *Psychological Assessment*, 27, 377-391. doi:10.1037/a0038270
- Van der Veen, F. C. A., Lucieer, J., &

- Bogaerts, S. (2016). Routine outcome monitoring and clinical decision-making in forensic psychiatry based on the Instrument for Forensic Treatment Evaluation. *PLoS ONE*, *11*, e0160787. doi:10.1371/journal.pone.0160787
- van Marle, H. J. C. (2002). The Dutch Entrustment Act (TBS): Its principles and innovations. *International Journal of Forensic Mental Health*, *1*, 83-92. doi:10.1080/14999013.2002.10471163
- Vess, J. (2001). Development and implementation of a functional skills measure for forensic psychiatric inpatients. *The Journal of Forensic Psychiatry*, *12*, 592-609. doi:10.1080/09585180110092001
- Ward, T., Melsler, J., & Yates, P. M. (2007). Reconstructing the Risk-Need-Responsivity model: A theoretical elaboration and evaluation. *Aggression and Violent Behavior*, *12*, 208-228. doi:10.1016/j.avb.2006.07.001
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence (Version 2)*. Burnaby, British Columbia, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Desmarais, S. (2009). *Manual for the Short-Term Assessment of Risk and Treatability (START) (Version 1.1)*. Port Coquitlam, British Columbia, Canada: Forensic Psychiatric Services Commission and St. Joseph's Healthcare.
- Wong, S. C. P., & Gordon, A. (2013). The Violence Reduction Programme: A treatment programme for violence-prone forensic clients. *Psychology, Crime & Law*, *19*, 461-475. doi:10.1080/1068316X.2013.758981
- Wong, S. C. P., Gordon, A., & Gu, D. (2007). Assessment and treatment of violence-prone forensic clients: An integrated approach. *The British Journal of Psychiatry*, *190*, 66-74. doi:10.1192/bjp.190.5.s66
- Wooditch, A., Tang, L. L., & Taxman, F. S. (2014). Which criminogenic need changes are most important in promoting desistance from crime and substance use? *Criminal Justice and Behavior*, *41*, 276-299. doi:10.1177/0093854813503543
- Workgroup Risk Assessment Forensic Psychiatry. (2002). *Handleiding HKT-30, versie 2002* [Manual HKT-30, version 2002]. The Hague, The Netherlands: Dutch Justice Department.
- Youn, S. J., Kraus, D. R., & Castonguay, L. G. (2012). The treatment outcome package: Facilitating practice and clinically relevant research. *Psychotherapy*, *49*, 115-122. doi:10.1037/a0027932