

Challenges for the Theory and Application of Dynamic Risk Factors

Jacinta R. Cording, Sarah M. Beggs Christofferson, and Randolph C. Grace

University of Canterbury, Christchurch, New Zealand

Address correspondence to:

Jacinta R. Cording
c/o Randolph C. Grace
University of Canterbury
Department of Psychology
Private Bag 4800
Christchurch, New Zealand
Email: jacinta.cording@gmail.com

Abstract

The inclusion of dynamic factors in risk assessment measures used with forensic populations has largely been considered an improvement in both the accuracy and utility of these assessments in informing treatment and sentencing. However, there are important challenges associated with the current approach to the conceptualisation, identification, and use of dynamic factors in risk assessment. Whereas some of these challenges relate to applied settings (such as the use of measures with different offender populations), there are also deeper questions regarding the construct validity of dynamic risk measures and the methodological strategies used to identify them. More emphasis on theoretically-driven research is needed, to identify causal and explanatory relationships between dynamic risk factors and recidivism. We hope that highlighting these challenges can help to build a consensus on a future research agenda for dynamic risk factors.

Keywords: dynamic risk, risk assessment, sexual offenders, violent offenders, treatment change

Challenges for the Theory and Application of Dynamic Risk Factors

The accurate measurement of offender risk is becoming an increasingly important focus of forensic research, due in part to the major implications that offender risk level can have for sentencing, custodial and community living arrangements, and the decision to indefinitely detain certain high-risk offenders. In addition to informing these decisions, risk assessment plays an important role in determining appropriate treatment for offenders in that there is now a large empirical base to suggest that the level of intervention should match the level of risk posed by an offender Andrews and Bonta (2010). Risk assessment has undergone a stark transformation from the first generation practice of relying largely upon unstructured clinical judgement (Bonta, 1996). First was the development of the empirically-based second generation tools based on static risk factors, followed by third generation measures that placed a greater emphasis on dynamic risk factors (also referred to as criminogenic needs) and were more theoretically informed. These third generation measures are the primary focus of the current article. Table 1 lists selected dynamic risk measures that have been developed with adult offender populations, including information about predictive validity and the samples used. As Table 1 shows, these dynamic risk measures have generally high levels of predictive accuracy that are at least on par with static measures of risk. [INSERT TABLE 1 NEAR HERE].

Although their causative and explanatory nature is still not clearly understood (Heffernan & Ward, 2015; Ward, this issue), dynamic risk factors are generally defined as situational and personal characteristics that are both empirically linked to an increased chance of future offending and are, theoretically at least, able to change (Andrews & Bonta, 2010). Although dynamic factors have received overall less attention than static factors (risk factors that are not amenable to change, e.g., age at first criminal offense) in research on risk

assessment, they remain highly promising not only with regards to the accurate prediction of recidivism, but also in identifying useful and effective treatment targets for offenders.

The growing information on dynamic risk factors has enabled the development of a number of measures containing either exclusively dynamic risk factors, or a combination of both static and dynamic factors. This change in the focus of risk assessment measures has been justified by findings that dynamic measures are able to provide incremental predictive validity beyond static factors in the measurement of future reoffending rates (Craissati & Beech, 2003; Gendreau, Little, & Goggin, 1996; Hanson & Harris, 2000). The proliferation of risk assessment measures for different offender populations has greatly improved the ability of professionals to provide informed judgements about the risk posed by individual offenders.

However, there remain a number of concerns about the measurement of dynamic risk that have important implications for assessment in applied settings; our goal in this paper is to identify these problems, both theoretical and practical, so that we might help to generate discussion and build a consensus on a future research agenda. We begin with a consideration of current evidence for the reliability and validity of various risk assessment tools, extending this assessment to the application of risk assessment with different populations of offenders, and the use of tools in applied contexts. Next we discuss the difficulties that arise as a result of the multi-faceted and indistinct nature of many dynamic risk factors, and the issues posed by a data-driven, rather than theory-driven, approach to the identification and conceptualisation of dynamic risk. We discuss possible solutions, such as a greater focus on confirmatory factor analysis and causal modelling in theory construction and testing. Third, we discuss the challenges faced in applying risk assessment in practice contexts, including the assessment of offender change. Finally, we consider the implications of socially-desirable responding for the accuracy of dynamic risk assessment based upon self-report measures.

The Validity of Dynamic Risk Assessment

Construct validity. The most common interpretation of construct validity, as applied to dynamic risk factors, relates to whether the particular measure represents or measures what it is supposed to; this has been referred to as “fundamental” construct validity (Colliver, Conlee, & Verhulst, 2012). As such, much research has focused on determining the concurrent validity – the extent to which a particular measure correlates with existing measures of the same constructs – and the predictive validity – the extent to which a measure accurately predicts a specific relevant outcome measure – of dynamic risk assessments. Overall, research has supported the concurrent validity of common dynamic risk assessment tools, in that offenders who are categorised as high risk on one particular measure are likely to also be categorised as high risk on other measures (Beggs & Grace, 2010; WAGDY Loza, Dhaliwal, Kroner, & Loza-Fanous, 2000; Nunes & Babchishin, 2012). Thus on the surface it appears as if there is a strong empirical basis for the construct validity of dynamic risk assessment. However, scholars have recently questioned the idea that concurrent and predictive validity are the two most important measures of construct validity (Borsboom, Mellenbergh, & van Heerden, 2004; Colliver et al., 2012; Haig, 2012). Instead, they argue that a given measure should be considered to have good construct validity if it is able to demonstrate a causal or explanatory link between the attributes it measures and the outcome of interest (Borsboom et al., 2004).

The implication is that dynamic risk assessment could be considered to have good construct validity only if researchers are able to demonstrate a causative or explanatory link between dynamic risk measures and recidivism. In order to assess whether this is possible given the current evidence base, we first must take a step back and ask an important question: what *are* dynamic risk factors? As defined by Andrews and Bonta (2010, p.7), dynamic risk factors are theoretically changeable factors that are “predictors of the criminal futures of

individuals”, and that with changes, “we see changes in the chances of criminal activity” (p.48). While this conceptualisation of dynamic risk factors, and therefore the assessment of dynamic risk, appears to at least theoretically meet the explanatory requirement of construct validity, there has been some recent doubt cast on whether the current reliance on correlational analyses and significance testing is a valid method of identifying truly causal factors relating to recidivism risk (Haig, 2012; Heffernan & Ward, 2015; Ward & Beech, 2014). Ward (this issue) discusses this issue in more depth; suffice to say, it is unclear whether the current conceptualisation of what constitutes as a “risk factor” is demonstrably valid in a more meaningful sense of the term.

A further problem with validity in the area of risk assessment relates to the multi-dimensional and indistinct nature of many dynamic risk factors (Heffernan & Ward, 2015). “Cognitive distortions” provides a good example. Typically, cognitive distortions are conceived of as non-normative belief structures that include justifications and rationalisations for sexual offending, and are regarded as a dynamic risk factor (Gannon, Ward, & Collie, 2007). However, Ó Ciardha and Gannon (2011) noted that “cognitive distortions” has been applied to a multitude of different constructs including “maladaptive beliefs” (Ward, Hudson, Johnston, & Marshall, 1997), “defensiveness” (Rogers & Dickey, 1991), “rationalisations” (Neidigh & Krop, 1992), “incorrect or deviant cognitive practices” (Ward & Casey, 2010), and “etiological cognitions” (Ó Ciardha & Gannon, 2011). Although definitions may play a larger role in the conceptualisation of scientific phenomena than perhaps they should (Haig, 2012), the degree of variation in how ‘cognitive distortions’ are defined poses a significant problem for developing valid measures of dynamic risk. Without clear definitions of risk factors, it is uncertain which features of each factor are linked to recidivism, and how to create clear scoring guidelines. Such uncertainty will not only potentially degrade the accuracy and discriminative validity of a measure, but also inter-rater and test-retest

reliability. It also has implications for construct validity – how can we be sure that we are measuring what we want to measure, when we are unable to define clearly what that is?

Predictive validity. Predictive validity in the area of offender risk assessment is the extent to which risk scores derived from risk assessment methods or tools predict offender outcomes, typically in the form of reoffending or reconviction rates post release from custody. Because the prevention of reoffending is a major goal for offender treatment and the justice system as a whole, developing tools or assessment methods that can accurately predict recidivism has understandably been a key research focus.

A strong empirical base has developed over recent years which supports the view that measures of dynamic risk can significantly improve the accuracy of risk prediction over and above the ability of static risk alone (Beggs & Grace, 2010; Hanson & Harris, 2000). In general, studies indicate that empirically-derived actuarial measures (i.e., measures with pre-determined items identified as risk factors in previous literature, and structured methods of calculating total risk) are significantly more predictive of reoffending compared to both structured clinical judgement (i.e., measures with pre-determined items but with no set method for calculating level of risk) and unstructured clinical judgement (Gendreau et al., 1996; Hanson & Morton-Bourgon, 2009). Somewhat unsurprisingly, studies have also found that the predictive accuracy of a particular measure changes depending on the match between the target behaviour being predicted and the type of behaviour the measure was developed to predict; for instance, a measure developed to measure risk of future sexual reoffending is typically more predictive of sexual recidivism than violent or general recidivism (Hanson & Morton-Bourgon, 2009). With this condition in mind, most common actuarial measures containing a large focus on dynamic items appear to perform relatively similarly in terms of predictive validity, with moderate rates of predictive validity for these measures in predicting

sexual, violent, and general recidivism (Campbell, French, & Gendreau, 2009; Hanson & Morton-Bourgon, 2009).

While the predictive validity of dynamic risk measures has been clearly demonstrated in previous studies, questions remain about the generalisability of these measures to different groups, such as female, young, or aboriginal offenders. Research focussing on the identification of important dynamic risk factors, and the development and validation of related risk assessment tools, is overwhelmingly conducted using samples of adult male offenders (Reisig, Holtfreter, & Morash, 2006; Schwalbe, 2008) under the general assumption that dynamic risk factors are broadly consistent across gender and age (for discussion of the theory underlying this, see Andrews, Bonta, & Wormith, 2006). This view has been described by Yesberg and colleagues as the “gender-neutral” assumption (Yesberg, Scanlan, Hanby, Serin, & Polaschek, 2015) and is currently the predominant view within the area of offender risk assessment. However, a number of scholars have rejected the gender-neutral assumption, instead asserting that there are a number of important differences between men and women which challenge the idea that dynamic risk factors – and therefore measures of dynamic risk – informed solely by male populations are equally valid for females. These includes differences in motivations for offending (Veysey & Hamilton, 2007), experiences of victimisation (Reisig et al., 2006), base rates of offending (Odgers, Moretti, & Reppucci, 2005) and pathways to crime (Belknap, 2014; also see Yesberg et al., 2015). A meta-analysis that investigated the predictive validity of the Level of Service Inventory – Revised (LSI-R; a widely-used dynamic risk measure in North America) for both male and female samples provided support for this view (Holtfreter & Cupp, 2007). They found that although the LSI-R performed moderately well for women, the measure was often more accurate for male offenders, and the predicted risk for female offenders was commonly too high; this was particularly true in cases where females had followed a pathway into crime that

was dissimilar from typical male offenders (e.g., women with high levels of abuse and/or neglect, poverty and victimisation), highlighting the importance of considering risk factors that are unique to female offenders.

Although Holtfreter and Cupp's (2007) results challenged the assumption of gender-neutral risk assessment, this conclusion is not straightforward because other studies with Level of Service (LS) risk measures have reported contradictory results. One recent meta-analysis found that LS tools in fact had greater predictive validity for female offenders (mean effect size = 0.53) than for male offenders (mean effect size = 0.39), for both adult and youth populations (Andrews et al., 2012). Support for the gender-neutral assumption has also been found with risk assessment tools developed for offenders in community settings, again with researchers finding that the predictive validity of the Dynamic Risk Assessment for Offender Re-entry (DRAOR; Serin, Mailloux, & Wilson, 2012) was greater for female offenders than for male offenders (Yesberg et al., 2015). The mixed results in research on gender and risk assessment highlight the need for further research in this area. Although overall results support the view that risk assessment tools are equally valid for males and females, it is important to investigate whether the inclusion of gender-specific items would further improve the predictive ability of risk assessment measures for females (Yesberg et al., 2015). It is also important to acknowledge that the generalisation of risk assessment measures to other minority offender populations, such as different ethnic groups, has also been challenged by the results of various studies that suggest a reduction in the accuracy of risk-assessment tools for non-White populations within North America (Chenane, Brennan, Steiner, & Ellison, 2015; see Shepherd, Luebbers, & Dolan, 2013, for review).

We suggest that one of the major reasons for the lack of research on validating dynamic risk measures with different offender groups is because of the data-driven nature of research of dynamic risk factors and assessment. This leads to an over-reliance on empirical

evidence that is available to researchers – which may be affected by particular choices of questionnaires or measures used as dynamic risk indicators - and under-reliance on theory or aetiology to guide our understanding of dynamic risk factors as scientific phenomena, and how these factors might be combined into an overall meaningful measure of risk; as explained by Haig (2013, p. 137), '[d]ata themselves are of scientific interest and importance only because they serve as evidence for the phenomena under investigation.' We consider the implications of the data-driven approach in the next section.

The Data-Driven Approach to Dynamic Risk

Overreliance on the hypothetico-deductive methodology in psychological research, where empirical data are used to identify, describe and/or discover correlates of constructs to inform theory, has long been criticised (e.g., Cohen, 1994; Falk & Greenbaum, 1995; Rozeboom, 1960, 1997). By contrast, the abductive method may represent a more meaningful and valid approach to research (Borsboom et al., 2004; Haig, 2005, 2009, 2014). . Haig (2014) describes the abductive approach as 'reasoning from factual premises to *explanatory* conclusions' (p.60), noting that 'phenomena, not data, serve as evidence for the abducted theories' (p.61). This differs from a purely inductive approach, whereby conclusions or theories are the 'same in kind' as the data used to generate them, meaning that they are more descriptive than explanatory in nature (Haig, 2014). As such, in using an abductive approach to theory generation we are more likely to develop meaningful knowledge about the phenomena of interest, including an understanding of aetiology, causal networks, and the potential for change or adaptation.

The abductive critique of the hypothetico-deductive method is clearly applicable to prior research on dynamic risk assessment, which has been largely data-driven, attempting to identify the best predictors of recidivism among a set of candidate measures using regression-

based statistical methods, rather than theory-driven (Heffernan & Ward, 2015; Ward, this issue). One could argue that variables which were studied as potential dynamic risk factors – such as lack of empathy for victims – were selected based on prior theoretical grounds (e.g., Marshall, Hamilton, & Fernandez, 2001), however acceptance of these dynamic risk factors was reliant largely on evidence of their ability to predict recidivism, ideally beyond the contribution made by static factors. For example, an influential series of articles by Beech and colleagues showed that cluster analysis (a data-driven exploratory technique) could be used to classify sexual offenders as ‘high deviance’ or ‘low deviance’, and the deviance classification was subsequently shown to predict recidivism beyond the Static-99 (Beech, 1998; Beech, Friendship, Erikson, & Hanson, 2002; Fisher, Beech, & Browne, 1999).

Recent conceptualisations of validity suggest that a construct (or attribute) is valid only insofar as it is shown to relate causally to a criterion (Borsboom, 2005; Borsboom et al., 2004). However data-driven approaches which demonstrate that dynamic risk factors are correlated with recidivism fall short of providing evidence of a causal linkage. This is likely to result in an incomplete picture of the risk posed by an individual offender, and lacking in an explanation of the aetiology or maintenance of behaviour. Consequently, implications for treatment formulation in terms of the most important needs to target to reduce risk are compromised. A greater understanding of the aetiology of serious offending would allow us to develop more effective strategies for early intervention, ideally to reduce first-time sexual and violent offending rather than reoffending.

Reasons for the data-driven approach are understandable when one considers the definition of dynamic risk factors that was outlined above: risk factors are predictors of increased criminal behaviour (Andrews & Bonta, 2010). Thus it is logical to identify risk factors by their ability to predict reoffending beyond the static, actuarial factors that had already been shown to have predictive validity. Leaving aside the issue of reliance on

statistical significance testing (see Cumming, 2012), there are two major potential problems with the data-driven approach for dynamic risk: 1) there is no guarantee that the risk factors are clinically meaningful in the sense that they can be used to explain the aetiology or maintenance of offending (Heffernan & Ward, 2015; Mann, Hanson, & Thornton, 2010); and 2) the identification of important risk factors is reliant upon the data or measures available to a given researcher.

The variability in the identification of risk factors caused by differences between the information contained in different datasets is displayed clearly by the emergence of competing “second-order” risk domains – composite risk factors that are predictive of reoffending, usually obtained by exploratory factor analysis on data from a psychometric battery. One example of this approach from the sex offender literature is from Allan, Grace, Rutherford, and Hudson (2007). These authors used exploratory factor analysis to identify dynamic risk factors from a large psychometric battery (a total of 20 different measures, including multiple sub-scales for some measures) that had been completed by a sample of sexual offenders against children prior to undergoing prison-based treatment. Four risk domains were identified that were each significantly predictive of sexual recidivism: Social Inadequacy (containing measures relating to low social competence and negative mood); Sexual Interests (containing measures relating to levels of sexual fantasies); Anger/Hostility (containing measures relating to anger expression and regulation); and Pro-offending Attitudes (containing measures relating largely to distorted cognitions and attitudes). Allan et al. combined these risk factors into a measure of ‘Overall Deviance’ which was shown to increase the predictive accuracy for recidivism beyond the Static-99. Previous researchers had also taken this approach for the development of general risk domains, however with slightly different results. For example, Olver, Nicholaichuk and Wong (2014) identified three domains of dynamic risk for sexual offenders – Socioemotional Functioning, Anger/Hostility,

and Misogynist Attitudes; Beech (1998) also identified three risk domains in his analysis of a psychometric battery, although these domains assessed conceptually different types of functioning: Social Competency, Pro-offending Attitudes, and Sexual Interests. Although there is evidence of some overlap between these factors, it is clear that the identification of relevant measures of risk for a given offender population varies depending on the measures available to a given researcher.

Moving beyond a data-driven approach will require different ways to identify risk factors, and possibly in how we conceptualise them. One approach that has been suggested elsewhere (Haig, 2005, 2012) is to modify our research methodologies to be more in line with an abductive approach to science, whereby theories are formed to explain patterns identified within the data (also called “phenomena”), rather than data analysis being used to directly generate theories. In terms of research into dynamic risk assessment, this would require increased utilisation of techniques such as confirmatory factor analysis, which can be used to test proposed models for dynamic risk factors. These could then be used to generate theories and hypotheses relating to relationships between certain risk factors or domains and recidivism (Haig, 2005). It would also require a greater emphasis on validating and replicating the findings of other research in order to ensure that we are identifying true phenomena rather than idiosyncrasies of particular datasets (Beech, 1998; Cumming, 2012). It is hoped that through changing how we identify dynamic risk factors, we might be able to develop a deeper and more meaningful understanding of how these factors contribute to the generation and maintenance of offending, as well as how these factors combine to determine the overall level of risk of an offender.

The Use of Dynamic Risk Measures in Practice Contexts

Most studies of the reliability and validity of dynamic risk assessment measures have been conducted in research rather than applied settings, in that the measures are generally used by researchers or developers rather than by professionals in a correctional context (e.g. parole officers, forensic psychologists and custodial officers). This raises questions about whether the reliability and validity of these measures will be maintained when they are no longer being scored by trained researchers or research assistants, but instead by staff who may have many other responsibilities.

Reasons why we might expect differences in the scoring of these measures by correctional compared to research staff are the lack of specialised, standardised training; high levels of work-related stress (National Institute of Justice, 2007); and large workloads and time pressures leading to a greater reliance on clinical or professional judgment rather than a strict adherence to scoring guidelines (Jones, Brown, & Zamble, 2010; Public Safety Canada, 2008). This would be particularly salient for measures that use a largely unstructured scoring format, allowing for a greater influence of personal heuristics and cognitive biases (Payne, Bettman, & Johnson, 1993). It has also been suggested by some researchers that this reliance on clinical judgement rather than the structured scoring criteria might result from a reluctance of professionals to accept the idea that their judgements might be less accurate than purely quantitative methods of assessing risk (Schlager, 2009; Schneider, Ervin, & Snyder-Joy, 1996). There is also some evidence to suggest that the fear of political and professional implications of having rated someone as low risk who later goes on to reoffend (even if the rating was correct), leads correctional staff to manipulate dynamic scores in a way that overestimates risk, with large resource and financial implications for the corrections service as a whole (Lanterman, Boyle, & Ragusa-Salerno, 2014; Schlager, 2009; Schneider et al., 1996).

A further threat posed to the validity of risk assessment in applied settings is the quality of training provision for these measures. Previous research has indicated that there is

a positive association between the quality of training provided to scorers and the predictive validity of risk assessment (Lowenkamp, Latessa, & Holsinger, 2004). For example, formal instruction on the use of the measures led by trained practitioners is superior to “bootstrap” training led by inexperienced or untrained colleagues, and the provision for hands-on practice improves quality (Lowenkamp et al., 2004; U.S. Department of Justice, 2007). The importance of training with risk assessment tools was highlighted by Andrews et al. (2011), who reported a series of meta-analyses using a total of 101 validation studies on the Level of Service Inventory (LSI) risk assessment tool. The studies were assessed on a number of factors that might moderate the predictive validity of the LSI, including length of follow-up, sample characteristics (such as gender and nationality), and LSI “allegiance” (defined as the level of involvement of the LSI developers in data collection for the study). Andrews et al. found that LSI allegiance was a significant moderator of the predictive validity of the LSI, with the stronger the allegiance, the higher the predictive validity. They suggested that LSI allegiance might be best understood as a proxy for the quality of the implementation and integrity in research methods (such as the selection of appropriate outcome measures). Andrews et al.’s (2011) results highlight the importance of a close adherence to scoring and implementation guidelines that are developed during quality training, in order to ensure maximal utility of the risk assessment tool in question.

Because performance in applied settings is essential for the utility of a risk assessment measure, it is important that we understand fully how reliability and validity of a given measure are affected by extending use from a research to an applied setting. Jones et al. (2010) investigated the extent to which the predictive ability of risk as assessed by parole officers differed from the predictive ability of risk scored by researchers. They used a prospective design where risk was measured at multiple time points in order to best emulate real-world use of the measures. In order to further emulate real-world circumstances, parole

officers provided crude proxy ratings of each area of dynamic risk based on their perceptions of offender circumstances and were not subject to quality assurance processes. Researchers, on the other hand, provided detailed and structured assessments of risk based on multi-dimensional case review (including semi-structured interviews and file reviews) and were subjected to routine quality assurance (such as inter-rater reliability checks). Contrary to expectation, researchers found that the predictive ability of the ratings of the parole officers and researchers were not significantly different, with AUCs of .76 and .79, respectively, indicating medium-high levels of predictive validity. In addition, ratings of external acute risk factors (such as employment) were highly correlated between the two groups, although those for internal acute risk factors (such as stress) were not significantly correlated. Thus, the ability of parole officers to predict recidivism based on crude proxies of risk was equal to that of highly structured and multi-dimensional assessments of risk. Jones et al. suggested that perhaps the extensive level of interaction between parole officers and offenders enabled them to gain a better picture of important collateral information about the offender, such as family, education and interaction with other health providers.

Although Jones et al. (2010) results are encouraging in that they suggest that risk assessment can be valid in an applied setting, it is important to note that the parole officers in this study were not expected to strictly adhere to scoring guidelines for each measure, but instead rated their perceptions of how a risk factor related to a given individual, in a similar way to the procedure used for structured clinical judgement tools. In other words, while it appears as though risk assessment can be accurate and valid in applied contexts, it is not so clear that the validity of actuarial tools can be transferred as successfully between research and practical contexts. It is important to note that as actuarial measures of dynamic risk continue to improve in their level of predictive validity and in the provision of estimated base-rates of offending for different risk bands (e.g. for the VRS:SO; Olver, Beggs

Christofferson, Grace, & Wong, 2014), it will become increasingly important that professionals are able to utilise structured risk assessment tools accurately. The ability of professionals to provide estimates of recidivism rates by risk level becomes even more significant as the possible sentencing options for high-risk offenders become increasingly restrictive and intrusive in the lives of offenders (e.g., preventive detention and extended supervision orders in New Zealand; Ryan, Wilson, Kilgour, & Reynolds, 2013). Thus, given that the extant literature largely supports the notion that the validity of risk assessment can change substantially depending on implementation and adherence to guidelines, more effort is warranted to ensure protocols are in place for effective training and ongoing quality assessment for those responsible for risk assessment in a professional context.

Measuring Change in Dynamic Risk

In addition to identifying those most at risk of reoffending, another vital task for professionals in applied settings is determining whether and to what extent individuals' levels of risk have changed. The link between this function and the concept of dynamic risk is clear – after all, a crucial component of the concept of dynamic risk is that it is just that – dynamic, or in other words, changeable. Although we acknowledge the criticism made by some authors that the measurement of change in dynamic risk is not meaningful in terms of addressing the causes of offending (e.g., see Ward, this issue), nonetheless it may have practical importance in terms of predicting future risk. However, this changeability has seemed to be more readily accepted in theory than actually tested in research. For some oft-cited dynamic risk factors, although an empirical link with recidivism has been established, and there may be face validity in terms of being changeable, studies have often only examined them at one point in time (e.g., Beech et al., 2002; Dempster & Hart, 2002).

Where more recent studies have begun to explore change, findings have been mixed. For example, a negligible amount of change was found in one study on psychometric test scores intending to tap into dynamic risk factors for violence, across a 20-month treatment programme for inpatient forensic mental health patients (Hildebrand & de Ruiter, 2012). This raises a problem evident in the psychometric assessment of dynamic risk: the difficulty in determining whether null results regarding change across treatment are the result of true lack of change (i.e., a poor treatment effect due to programme ineffectiveness or participant factors such as poor motivation), insensitive measurement relating to the tests chosen, or, that the 'dynamic' factors under investigation are not really dynamic. In contrast, other studies have reported significant improvements between pre-treatment and post-treatment psychometric test scores relating to dynamic risk areas such as pro-criminal attitudes, family/marital relationships, and education/employment (Brooks Holliday, Heilbrun, & Fretz, 2012). Likewise, substantial apparent changes from pre- to post-treatment have tended to be found when dynamic factors have been measured psychometrically among sex offenders (e.g., Hudson, Wales, Bakker, & Ward, 2002; Marques, Wiederanders, Day, Nelson, & Van Ommeren, 2005). However, it is possible that such findings reflect a social desirability response bias, which we consider below.

In many studies in which dynamic risk changes have been explored, recidivism outcomes were not included in the investigations. Arguably, as well as needing to be empirically linked to recidivism, and changeable, it is also inherent in the definition of dynamic risk factors that any changes should be *meaningful* (i.e., linked to changes in actual reoffending risk). In fact, this is a central tenet of the needs principle of offender rehabilitation as described by Andrews & Bonta (2010). However, as noted by Serin, Lloyd, Helmus, Derkzen, & Luong (2013), the question of whether changes (i.e., in dynamic risk

factors) are reliably associated with recidivism likelihood is relatively unexplored. Research has only more recently begun to test the assumption empirically.

In one test of the link between treatment change and outcome, Beggs and Grace (2011) demonstrated that specific within-treatment changes in dynamic risk factors in the desired direction, measured psychometrically, can be linked with reduced recidivism at follow-up. In doing so, they reported on the problematic nature of analysing raw change scores: on any given test, individuals with pre-treatment scores towards the more problematic end of the scale (indicating higher levels of dynamic risk) have the opportunity to show greater levels of change across treatment, as they have more 'room to move'. This is also relevant to the issue of socially desirable responding in self-report testing, which we discuss in greater detail, with reference to the Beggs and Grace study, below. However a further problem is also clear: given that both riskier scores and lower change should theoretically be linked with higher recidivism, yet those with riskier scores have the opportunity to attain higher change scores, use of raw change scores is inherently flawed. To manage this issue, Beggs and Grace employed a method of regression in which pre-treatment scores were partialled out of the prediction equation. This allowed a more meaningful pattern of results to emerge, in which positive treatment change was associated with reduced sexual recidivism overall, and for three out of four dynamic risk domains (employing the Allan et al. (2007) framework: social inadequacy, sexual interests, and anger/hostility; the fourth domain, pro-offending attitudes, approached significance). This technique has subsequently been applied by Olver, Nicholaichuk, Kingston, and Wong (2014) in their exploration of therapeutic change and recidivism using a psychometric risk prediction instrument (the Violence Risk Scale-Sexual Offender Version; VRS-SO).

Other studies have employed a different method known as clinically significant change methodology to explore the link between within-treatment changes in dynamic risk

and recidivism (e.g., Barnett, Wakeling, Mandeville-Norden, & Rakestrow, 2013; Olver, Beggs Christofferson, & Wong, 2015). This method avoids the problems associated with raw change scores, as in addition to considering change magnitude, post-treatment scores are evaluated against non-deviant norms to determine whether the individual has qualitatively “improved”, “recovered”, is “already ok” (i.e., never scored outside the normative range), or remained “unchanged”. While this method offers a user-friendly and readily interpretable classification system for individuals based on what their dynamic risk test scores say about their treatment outcome, Olver and colleagues (2015) and others (e.g., Barnett et al., 2013) have overviewed the limitations of the method and noted mixed findings, in particular that the usefulness of the output is dependent on the quality of the measures used. In general, Olver et al. (2015) suggested that the use of a single, purpose-designed risk tool containing multiple dynamic factors, such as the VRS-SO or the STABLE 2007 (see Hanson, Harris, Scott, & Helmus, 2007), may offer advantages over the psychometric battery approach for the consistent and meaningful applied measurement of dynamic risk factors, and change in these across treatment.

One common factor in the change studies discussed above is that the assessment of dynamic risk occurred at only two points in time – prior to, and then following, treatment. Whilst studies employing this design have been very useful in terms of establishing empirical relationships between specific changes in dynamic risk and decreased recidivism, and exploring the clinical measurement of within-treatment changes, they have focused exclusively on the period of treatment engagement as the change mechanism for risk. However theoretically speaking, other factors could influence the presence or expression of dynamic risk (resulting in change), such as maturation (Hirschi & Gottfredson, 1983), social context and influences (Sampson & Laub, 1995), or in the case of sex offenders, age-related decline in sexual response (Blanchard & Barbaree, 2005). It has also been suggested that

including at least three waves of assessment increases the probability of detecting change (Brown, Amand, & Zamble, 2009). A recent multi-wave study by Greiner, Law, and Brown (2014) illustrated the tracking of seven major theorised dynamic risk factors (employment, personal/emotional factors, substance use, criminal attitudes, criminal associates, family functioning, and community functioning) among female offenders following their release from prison, across four assessment waves at six-monthly intervals. They found that all seven factors were significantly related to survival time without reoffending in the community, and that prediction was improved by their use of multiple assessments of dynamic risk across time. On the other hand, change across multiple waves using a well-validated dynamic risk tool for sex offenders, the STABLE 2007, has been found to not be associated with recidivism (Hanson et al., 2007). As such, although assessing dynamic risk factors at multiple time points both during and after treatment appears to be a promising technique in terms of improving the assessment of change, it is apparent that there are other factors that contribute to the mixed results of studies on change. It is possible that part of the problem lies with our current conceptualisation of what constitutes a dynamic risk factor, as discussed in previous sections.

Clearly, more research is needed on the assessment of changes in dynamic risk, with numerous challenges having been identified for applied settings. As discussed above, dynamic risk evaluations in applied settings typically have a great impact on individuals' progress through the criminal justice system, and assessments of change (across treatment or with continued repeat assessments) are certainly no different. For clinicians this carries a great responsibility, and the need to ensure that the methods we select to assess the changes made by our clients are both capable of detecting change that has occurred, and meaningful in terms of being predictive of actual reductions in the likelihood of recidivism. Although a number of the factors that affect the accuracy of risk assessment (including the issue of

generalisability discussed above) can be mitigated by the selection of appropriate methods on the part of the clinician, there are a number threats to predictive validity that are not so easily avoided when using current assessment tools. We now turn to a discussion of one of the more prominent of these threats: Socially desirable responding.

Threats to Validity of Dynamic Risk Measures: Socially Desirable Responding

One important threat to the validity of dynamic risk assessments, particularly those in which offender self-reports play a role, is impression management or socially desirable responding (SDR; see Tan & Grace, 2008, for review). SDR refers to the tendency of some individuals to respond in ways that are likely to elicit approval from others, and to refrain from responding in ways that would be met with disapproval (Crowne & Marlowe, 1964). In terms of self-report measures, this tendency means that individuals may be influenced to respond to individual items not only based on their beliefs relating to the item content, but also what they believe to be a socially appropriate response. Such patterns of responding pose a unique challenge within an offender population, where an idiosyncratic desire to appear “overly positive” (Paulhus, 2002) is further augmented by a penal system that creates clear incentives for individuals to present in a positive way for parole boards, judges, probation officers, and other individuals making decisions affecting the length and type of custodial sentences (Davis, Thake, & Weekes, 2012).

To the extent that offender self-reports are considered in classification and parole decisions, assessing the credibility of those reports is obviously important. The inaccuracies in measurement that could potentially result from SDR threaten not only the classification and parole decisions, but also affect the ability to assess accurately the level of dynamic risk or need of an individual offender. Some psychometric measures used to assess dynamic risk factors have highly transparent items, so that it is fairly obvious to the responder as to what

the test is measuring and therefore what the socially acceptable responses might be. For example, the Abel-Becker Cognitions Scale (ABCS; Abel et al., 1989), which is commonly used to assess offence-supportive beliefs and attitudes with sexual offenders against children, includes items such as “I show my love and affection to a child by having sex with her (him)” and “A child who doesn’t physically resist an adult’s sexual advances really wants to have sex with the adult”. Using transparent measures makes it relatively easy for offenders to minimise or deny problematic attitudes, and to exaggerate any positive or pro-social traits.

Because of these potential problems with offender self-reports, researchers have often used measures of SDR as part of psychometric batteries to assess dynamic risk (e.g., Beech, 1998). Variance associated with SDR is then partialled out prior to making a dynamic risk classification (cf. Saunders, 1991). Although this is a common practice, there is little evidence that correcting for SDR in this manner improves the accuracy of decision making in applied settings in general (McGrath, Mitchell, Kim, & Hough, 2010). Results with forensic samples are similar. Mills and Kroner (2006) found that using the impression management subscale of the Balanced Inventory of Desirable Responding (BIDR; Paulhus & John, 1998) to correct the self-reports of incarcerated offenders on a measure of dynamic risk decreased, not increased, the predictive validity for recidivism (although note that the decrease was not statistically significant). Recently, Stevens, Tan and Grace (2015) showed that correcting sexual offenders’ self-reported dynamic risk scores for SDR (measured by the Marlowe-Crowne Social Desirability Scale; Crowne & Marlowe, 1964) had virtually no effect on predictive validity for sexual recidivism.

Because removal of SDR variance does not improve the correlation of self-report dynamic risk measures with recidivism, researchers have suggested that SDR scales like the BIDR or MCSDS may actually be measuring a personality trait or enduring disposition related to need for social approval. According to this view, SDR may be correlated with

dynamic risk factors (indeed, SDR was negatively correlated with dynamic risk in Mills and Kroner (2006) and Stevens et al. (2015); see also Mathie and Wakeling (2010), but is not related to recidivism risk directly). This view is consistent with a recent reinterpretation of SDR by Uziel (2010), who suggested that instead of response bias, measures of SDR should be regarded as ‘interpersonally oriented self-control’, that is, SDR reflects the individual’s ability to adjust to social situations and seek approval from others.

Thus, more research is needed on the nature of SDR in forensic settings, but there appears to be little justification for using SDR measures to adjust for response bias. However, this does not imply that response bias is not a problem, or that the transparent nature of many self-report measures is not a cause for concern when assessing dynamic risk. Indeed, there is strong evidence that impression management has a major impact on offender’s self reports. As discussed above, Beggs and Grace (2011) compared responses of sexual offenders to a psychometric battery both before and after treatment and found that medium-to-large gains (in terms of effect size) were reported by men across all variables in the battery. Pre-treatment scores were strongly correlated with change scores (average $R^2 = .33$). Interestingly, Beggs and Grace found that the predictive validity of change scores for recidivism increased when variance associated with pre-treatment scores was partialled out, exemplifying suppressor effects (with some correlations actually reversing direction). That is, measures of treatment gain based on self-reports were more valid predictors of recidivism when the initial level of risk was taken into account. They interpreted this as evidence that offenders with higher pre-treatment risk scores had more potential to show greater change, given the fixed range of possible scores in the battery. Thus their results showed that self-reported change scores should not be taken at face value and were likely biased by impression management, but when the pre-treatment levels of risk were controlled for,

provided a potentially valid measure of treatment change. Similar results were reported subsequently by Olver, Beggs Christofferson, et al. (2014).

Although further research is needed in order to more fully understand what SDR actually is, how it is structured, and how it affects risk assessment in different offender populations, overall research suggests that SDR may not be the large challenge that many assume it to be, and that its impact on the ability to measure dynamic risk with self-reports may be less severe than originally thought.

Conclusion

In terms of the overall accuracy and utility of risk assessment tools, the move towards a greater consideration of dynamic risk factors when assessing risk has been a promising step forward. Not only has the inclusion of dynamic factors shown to improve the predictive validity of actuarial tools (Craissati & Beech, 2003; Gendreau et al., 1996; Hanson & Harris, 2000), but these measures are also theoretically able to inform treatment targets, allow for the assessment of change in risk over time, and incorporate more meaningful risk factors that can be connected to the aetiology and maintenance of antisocial behaviour, all of which are important considerations for a rehabilitative approach to criminal justice. However, it is important to note that the way in which these tools are developed and utilised will moderate the relationship between these theoretical benefits of incorporating dynamic risk factors and how these tools function in reality.

As discussed above, the current research methods used to identify dynamic risk factors, and the ways in which we combine and apply these factors to the measurement of risk, is possibly creating a disconnect between the theoretical conceptualisation of dynamic risk and what these tools are measuring in practice. While dynamic risk assessment tools may be reasonably accurate in their predictions of reoffending, it is important that we recognise

that this does not necessarily mean that the risk factors used in these measures are psychologically meaningful, or that they contribute to our understanding of the aetiology of, or indeed to the desistance from, antisocial behaviour (Heffernan & Ward, 2015). As such, it is vital that we consider the methodology utilised when identifying important predictors of risk and remember that the development of meaningful knowledge about risk necessarily includes knowledge of aetiology and causal networks (Borsboom et al., 2004; Haig, 2005, 2012).

It is also important to recognise that the predictive validity of quantitative risk measures is in a large part dependent on the sample used in the validation; for this reason, we highlight the necessity of validating measures in different countries or jurisdictions, and with different types of offenders. In addition to the type and demographics of offenders, the predictive validity of risk assessment tools is also highly impacted on by the way in which the measure is applied, both in terms of the quality of training provided on a given measure, and in the adherence to scoring guidelines demonstrated by the clinician or researcher (Lowenkamp et al., 2004; Schlager, 2009). Assessments of risk that incorporate self-report measures also have the added threat of validity of responses, which might be affected by offender biases, both conscious and sub-conscious; although the literature largely suggests that the concern about the erosion of prediction accuracy as a result of offenders “faking good” is not empirically founded (Beggs & Grace, 2011; Olver, Nicholaichuk, et al., 2014; Stevens et al., under revision) one still must consider the impact of biased responding on the ability of clinicians to gauge the specific level and types of needs that should form the focus of treatment.

Addressing these applied and theoretical challenges related to the assessment of dynamic risk highlights requires a change in the methodologies and analytical techniques used in dynamic risk research (e.g. increased use of causal modelling and confirmatory factor

analysis), as well as a move towards a more theoretically-driven identification of relevant dynamic risk factors. However a significant first step in this process will be widely acknowledging these areas of difficulty, and generating discussion both about the extent of these issues and other ways in which these challenges can be resolved. It is hoped that a careful consideration of these topics can lead to the development of other ways in which these challenges can be faced, and the true potential of dynamic risk factors in their application to risk assessment can be realised.

References

- Abel, G. G., Gore, D. K., Holland, C. L., Camp, N., Becker, J. V., & Rather, J. (1989). The measurement of the cognitive distortions in child molesters. *Annals of Sex Research*, 2, 135-153.
- Allan, M., Grace, R. C., Rutherford, B., & Hudson, S. M. (2007). Psychometric assessment of dynamic risk factors for child molesters. *Sexual Abuse: A Journal of Research and Treatment*, 19(4), 347-367. doi:10.1177/107906320701900402
- Andrews, D. A., & Bonta, J. (1995). *LSI-R: The level of service inventory-revised user's manual*. Multi-Health Systems, Toronto.
- Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct (Fifth Edition)*. Boston: Anderson Publishing Ltd.

- Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, 52(1), 7-27.
doi:10.1177/0011128705281756
- Andrews, D. A., Bonta, J., Wormith, J. S., Guzzo, L., Brews, A., Rettinger, J., & Rowe, R. (2011). Sources of variability in estimates of predictive validity: A specification with Level of Service general risk and need. *Criminal Justice and Behavior*, 38(5), 413-432. doi:10.1177/0093854811401990
- Andrews, D. A., Guzzo, L., Raynor, P., Rowe, R. C., Rettinger, L. J., Brews, A., & Wormith, J. S. (2012). Are the major risk/need factors predictive of both female and male reoffending?: A test with the eight domains of the Level of Service/Case Management Inventory. *International Journal of Offender Therapy and Comparative Criminology*, 56(1), 113-133. doi:10.1177/0306624x10395716
- Barnett, G. D., Wakeling, H., Mandeville-Norden, R., & Rakestrow, J. (2013). Does change in psychometric test scores tell us anything about risk of reconviction in sexual offenders? *Psychology, Crime & Law*, 19(1), 85-110.
- Beech, A. R. (1998). A psychometric typology of child abusers. *International Journal of Offender Therapy and Comparative Criminology*, 42(4), 319-339.
- Beech, A. R., Friendship, C., Erikson, M., & Hanson, R. K. (2002). The relationship between static and dynamic risk factors and reconviction in a sample of U.K. child abusers. *Sexual Abuse: A Journal of Research and Treatment*, 14(2), 155-167.
doi:10.1177/107906320201400206
- 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

- Beggs, S. M., & Grace, R. C. (2011). Treatment gain for sexual offenders against children predicts reduced recidivism: A comparative validity study. *Journal of Consulting and Clinical Psychology, 79*(2), 182-192. doi:10.1037/a0022900
- Belknap, J. (2014). *The invisible woman: Gender, crime, and justice* (4th ed.). Belmont, CA: Thomson Wadsworth Publishing Company.
- Blanchard, R., & Barbaree, H. E. (2005). The strength of sexual arousal as a function of the age of the sex offender: Comparisons among pedophiles, hebephiles, and teleiophiles. *Sexual abuse: a journal of research and treatment, 17*(4), 441-456.
- Bonta, J. (1996). Risk-needs assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18-32). Thousand Oaks, CA: Sage.
- Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge, UK: Cambridge University Press.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*(4), 1061-1071. doi:10.1037/0033-295X.111.4.1061
- Brooks Holliday, S., Heilbrun, K., & Fretz, R. (2012). Examining improvements in criminogenic needs: The risk reduction potential of a structured re-entry program. *Behavioral Sciences & the Law, 30*(4), 431-447.
- Brown, S. L., Amand, M. D. S., & Zamble, E. (2009). The dynamic prediction of criminal recidivism: a three-wave prospective study. *Law and Human Behavior, 33*(1), 25-45.
- Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior, 36*(6), 567-590.
- Chenane, J. L., Brennan, P. K., Steiner, B., & Ellison, J. M. (2015). Racial and ethnic differences in the predictive validity of the Level of Service Inventory–Revised

among prison inmates. *Criminal Justice and Behavior*, 42(3), 286-303.

doi:10.1177/0093854814548195

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49(12), 997-1003.

doi:10.1037/0003-066X.49.12.997

Colliver, J. A., Conlee, M. J., & Verhulst, S. J. (2012). From test validity to construct validity ... and back? *Medical Education*, 46(4), 366-371. doi:10.1111/j.1365-

2923.2011.04194.x

Craissati, J., & Beech, A. (2003). A review of dynamic variables and their relationship to risk prediction in sex offenders. *Journal of sexual aggression*, 9(1), 41-55.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive: studies in evaluative dependence*. New York, NY: Wiley.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Davis, C. G., Thake, J., & Weekes, J. R. (2012). Impression managers: Nice guys or serious criminals? *Journal of Research in Personality*, 46(1), 26-31.

doi:http://dx.doi.org/10.1016/j.jrp.2011.11.001

Dempster, R. J., & Hart, S. D. (2002). The relative utility of fixed and variable risk factors in discriminating sexual recidivists and nonrecidivists. *Sexual Abuse: A Journal of Research and Treatment*, 14(2), 121-138.

Falk, R., & Greenbaum, C. W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory & Psychology*, 5(1), 75-98.

Fisher, D., Beech, A. R., & Browne, K. (1999). Comparison of sex offenders to nonoffenders on selected psychological measures. *International Journal of Offender Therapy and Comparative Criminology*, 43(4), 473-491.

- Gannon, T. A., Ward, T., & Collie, R. (2007). Cognitive distortions in child molesters: Theoretical and research developments over the past two decades. *Aggression and Violent Behavior, 12*(4), 402-416. doi:<http://dx.doi.org/10.1016/j.avb.2006.09.005>
- Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works. *Criminology, 34*(4), 575-608.
- Greiner, L. E., Law, M. A., & Brown, S. L. (2014). Using dynamic factors to predict recidivism among women: A four-wave prospective study. *Criminal Justice and Behavior, 42*(5), 457-480. doi: 10.1177/0093854814553222.
- Haig, B. D. (2005). An abductive theory of scientific method. *Psychological Methods, 10*(4), 371-388. doi:10.1037/1082-989X.10.4.371
- Haig, B. D. (2009). Inference to the best explanation: A neglected approach to theory appraisal in psychology. *The American journal of psychology, 219*-234.
- Haig, B. D. (2012). From construct validity to theory validation. *Measurement: Interdisciplinary Research and Perspectives, 10*(1-2), 59-62.
doi:10.1080/15366367.2012.681975
- Haig, B. D. (2013). Detecting psychological phenomena: Taking bottom-up research seriously. *The American Journal of Psychology, 126*(2), 135-153.
doi:10.5406/amerjpsyc.126.2.0135
- Haig, B. D. (2014). The philosophy of quantitative methods. In *The Oxford handbook of quantitative methods in psychology: Foundations* (pp. 7-31). Oxford, UK: Oxford University Press
- Hanson, R. K., Harris, A., Scott, T., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project*. (Report no. 2007-05). Ottawa, ON: Public Safety and Emergency Preparedness Canada.

- Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene?: Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behavior*, 27(1), 6-35.
doi:10.1177/0093854800027001002
- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment*, 21(1), 1-21. doi:10.1037/a0014421
- Heffernan, R., & Ward, T. (2015). The conceptualization of dynamic risk factors in child sex offenders: An agency model. *Aggression and Violent Behavior*, 24, 250-260.
doi:http://dx.doi.org/10.1016/j.avb.2015.07.001
- Hildebrand, M., & de Ruiter, C. (2012). Psychopathic traits and change on indicators of dynamic risk factors during inpatient forensic psychiatric treatment. *International Journal of Law and Psychiatry*, 35, 276-288. doi:10.1016/j.ijlp.2012.04.001
- Hirschi, T., & Gottfredson, M. (1993). Age and the explanation of crime. *American Journal of Sociology*, 89(3), 552-584.
- Holtfreter, K., & Cupp, R. (2007). Gender and risk assessment: The empirical status of the LSI-R for women. *Journal of Contemporary Criminal Justice*, 23(4), 363-382.
doi:10.1177/1043986207309436
- Hudson, S. M., Wales, D. S., Bakker, L., & Ward, T. (2002). Dynamic risk factors: the Kia Marama evaluation. *Sexual Abuse: A Journal of Research and Treatment*, 14(2), 103-119.
- Jones, N. J., Brown, S. L., & Zamble, E. (2010). Predicting criminal recidivism in adult male offenders: Researcher versus parole officer assessment of dynamic risk. *Criminal Justice and Behavior*, 37(8), 860-882. doi:10.1177/0093854810368924

- Lanterman, J. L., Boyle, D. J., & Ragusa-Salerno, L. M. (2014). Sex offender risk assessment, sources of variation, and the implications of misuse. *Criminal Justice and Behavior, 41*(7), 822-843. doi:10.1177/0093854813515237
- Lowenkamp, C. T., Latessa, E. J., & Holsinger, A. M. (2004). Empirical evidence on the importance of training and experience in using the Level of Service Inventory-Revised. *Topics in Community Corrections, 49-53*.
- Loza, W. (2005). *Self-Appraisal Questionnaire*. Toronto, Canada: Multi-Health Systems.
- Loza, W., Dhaliwal, G., Kroner, D. G., & Loza-Fanous, A. (2000). Reliability, construct, and concurrent validities of the Self-Appraisal Questionnaire: A tool for assessing violent and nonviolent recidivism. *Criminal Justice and Behavior, 27*(3), 356-374. doi:10.1177/0093854800027003005
- Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment, 22*(2), 191-217. doi:10.1177/1079063210366039
- Marques, J. K., Wiederanders, M., Day, D. M., Nelson, C., & Van Ommeren, A. (2005). Effects of a relapse prevention program on sexual recidivism: Final results from California's Sex Offender Treatment and Evaluation Project (SOTEP). *Sexual Abuse: A Journal of Research and Treatment, 17*(1), 79-107.
- Marshall, W., Hamilton, K., & Fernandez, Y. (2001). Empathy deficits and cognitive distortions in child molesters. *Sexual Abuse: A journal of research and treatment, 13*(2), 123-130.
- Mathie, N. L., & Wakeling, H. C. (2010). Assessing socially desirable responding and its impact on self-report measures among sexual offenders. *Psychology, Crime & Law, 17*(3), 215-237. doi:10.1080/10683160903113681

- McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin, 136*(3), 450-470. doi:10.1037/a0019216
- Mills, J. F., & Kroner, D. G. (2006). Impression management and self-report among violent offenders. *Journal of Interpersonal Violence, 21*(2), 178-192.
doi:10.1177/0886260505282288
- National Institute of Justice. (2007). *Helping probation and parole officers cope with stress*. Retrieved from <https://www.ncjrs.gov/pdffiles1/nij/220107.pdf>.
- Neidigh, L., & Krop, H. (1992). Cognitive distortions among child sexual offenders. *Journal of Sex Education and Therapy, 18*(3), 208-215.
- 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*(1), 29-45.
doi:10.1177/1079063211404921
- Ó Ciardha, C., & Gannon, T. A. (2011). The cognitive distortions of child molesters are in need of treatment. *Journal of Sexual Aggression, 17*(2), 130-141.
doi:10.1080/13552600.2011.580573
- Odgers, C. L., Moretti, M. M., & Reppucci, N. D. (2005). Examining the science and practice of violence risk assessment with female adolescents. *Law and Human Behavior, 29*(1), 7-27. doi: 10.1007/s10979-005-1397-z
- Olver, M. E., Beggs Christofferson, S. M., Grace, R. C., & Wong, S. C. P. (2014). Incorporating change information into sexual offender risk assessments using the Violence Risk Scale–Sexual Offender version. *Sexual Abuse: A Journal of Research and Treatment, 26*(5), 472-499. doi:10.1177/1079063213502679
- Olver, M. E., Beggs Christofferson, S. M., & Wong, S. C. P. (2015). Evaluation and applications of the clinically significant change method with the Violence Risk Scale-

- Sexual Offender version: Implications for risk-change communication. *Behavioral Sciences & the Law*, 33(1), 92-110. doi:10.1002/bsl.2159
- Olver, M. E., Nicholaichuk, T. P., Kingston, D. A., & Wong, S. C. P. (2014). A multisite examination of sexual violence risk and therapeutic change. *Journal of Consulting and Clinical Psychology*, 82(2), 312-324. doi:10.1037/a0035340
- Olver, M.E., Nicholaichuk, T.P., & Wong, S. C. P. (2014). The predictive and convergent validity of a psychometric battery used to assess sexual offenders in a treatment programme: an 18-year follow-up. *Journal of Sexual Aggression*, 20(2), 216-239. doi: 10.1080/13552600.2013.816791
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The Role of Constructs in Psychological and Educational Measurement* (pp. 49-69). Mahwah, NJ: Lawrence Erlbaum.
- Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, 66(6), 1025-1060.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*: Cambridge University Press.
- Public Safety Canada. (2008). *Outside the walls*. Ottawa, Canada. Retrieved from <http://www.publicsafety.gc.ca/cnt/cntrng-crm/csc-scc-rvw-pnl/report-rapport/otsd-wll-eng.aspx>.
- Reisig, M. D., Holtfreter, K., & Morash, M. (2006). Assessing recidivism risk across female pathways to crime. *Justice Quarterly*, 23, 384-405.
- Rogers, R., & Dickey, R. (1991). Denial and minimization among sex offenders. *Annals of Sex Research*, 4(1), 49-63.

- Rozeboom, W. W. (1960). The fallacy of the null-hypothesis significance test. *Psychological bulletin*, 57(5), 416-428.
- Rozeboom, W. W. (1997). Good science is abductive, not hypothetico-deductive. In L. L. Harlow, S. A. Mulaik & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 335-392). Mahwah, NJ: Lawrence Erlbaum Associates Publishers
- Ryan, J., Wilson, N., Kilgour, G., & Reynolds, N. (2013). Balancing efficiency and accuracy: Guidelines for the psychological assessment of extended supervision orders for child sex offenders in New Zealand. *Psychiatry, Psychology and Law*, 21(1), 139-157.
doi:10.1080/13218719.2013.793154
- Sampson, R. J., & Laub, J. H. (1995). *Crime in the making: Pathways and turning points through life*. Cambridge, Mass: Harvard University Press.
- Saunders, D. G. (1991). Procedures for adjusting self-reports of violence for social desirability bias. *Journal of Interpersonal Violence*, 6(3), 336-344.
doi:10.1177/088626091006003006
- Schlager, M. D. (2009). The organizational politics of implementing risk assessment instruments in community corrections. *Journal of Contemporary Criminal Justice*, 25(4), 412-423. doi:10.1177/1043986209344555
- Schneider, A. L., Ervin, L., & Snyder-Joy, Z. (1996). Further exploration of the flight from discretion: The role of risk/need instruments in probation supervision decisions. *Journal of Criminal Justice*, 24(2), 109-121. doi:http://dx.doi.org/10.1016/0047-2352(95)00059-3
- Schwalbe, C. S. (2008). A meta-analysis of juvenile justice risk assessment instruments: Predictive validity by gender. *Criminal Justice and Behavior*, 35(11), 1367-1381.
doi:10.1177/0093854808324377

- Serin, R. C., Lloyd, C. D., Helmus, L., Derkzen, D. M., & Luong, D. (2013). Does intra-individual change predict offender recidivism? Searching for the Holy Grail in assessing offender change. *Aggression and violent behavior, 18*(1), 32-53.
- Serin, R. C., Mailloux, D. L., & Wilson, N. J. (2012). *The dynamic risk assessment of offender re-entry (DRAOR)*. Unpublished user manual.
- Shepherd, S. M., Luebbers, S., & Dolan, M. (2013). Gender and ethnicity in juvenile risk assessment. *Criminal Justice and Behavior, 40*(4), 388-408.
doi:10.1177/0093854812456776
- Smith, P., Cullen, F. T., & Latessa, E. J. (2009). Can 14,737 women be wrong? A meta-analysis of the LSI-R and recidivism for female offenders. *Criminology & Public Policy, 8*(1), 183-208. doi:10.1111/j.1745-9133.2009.00551.x
- Stevens, C. D. J., Tan, L., & Grace, R. C. (2015). Socially-desirable responding and psychometric assessment of dynamic risk in sexual offenders against children. Under revision for *Psychology, Crime & Law*.
- Tan, L., & Grace, R. C. (2008). Social desirability and sexual offenders: A review. *Sexual Abuse: A Journal of Research and Treatment, 20*(1), 61-87. Retrieved from <http://sax.sagepub.com/content/20/1/61.abstract>
- U.S. Department of Justice. (2007). *The importance of assessment in sex offender management: An overview of key principles and practices*. Silver Spring, MD. Retrieved from http://www.csom.org/pubs/assessment_brief.pdf.
- Uziel, L. (2010). Rethinking social desirability scales: From impression management to interpersonally oriented self-control. *Perspectives on Psychological Science, 5*(3), 243-262.

- Veysey, B. M., & Hamilton, Z. (2007). Girls will be girls: Gender differences in predictors of success for diverted youth with mental health and substance abuse disorders. *Journal of Contemporary Criminal Justice*, 23(4), 341-362. doi:10.1177/1043986207309435
- Ward, T. (in press). Dynamic risk factors: Scientific kinds or predictive constructs. *Psychology, Crime & Law*.
- Ward, T., & Beech, A. R. (2014). Dynamic risk factors: a theoretical dead-end? *Psychology, Crime & Law*, 21(2), 100-113. doi:10.1080/1068316X.2014.917854
- Ward, T., & Casey, A. (2010). Extending the mind into the world: A new theory of cognitive distortions in sex offenders. *Aggression and Violent Behavior*, 15(1), 49-58.
Retrieved from <http://www.sciencedirect.com/science/article/pii/S1359178909001001>
- Ward, T., Hudson, S. M., Johnston, L., & Marshall, W. L. (1997). Cognitive distortions in sex offenders: An integrative review. *Clinical Psychology Review*, 17(5), 479-507.
- Wong, S., & Gordon, A. E. (1999-2003). *The Violence Risk Scale*. Unpublished manuscript. University of Saskatchewan, Saskatoon, Saskatchewan, Canada.
- Wong, S., Olver, M. E., Nicholaichuk, T. P., & Gordon, A. (2003). *The Violence Risk Scale - Sexual Offender version (VRS-SO)*. Saskatoon, Saskatchewan, Canada: Regional Psychiatric Centre and University of Saskatchewan.
- Yesberg, J. A., Scanlan, J. M., Hanby, L. J., Serin, R. C., & Polaschek, D. L. L. (2015). Predicting women's recidivism: Validating a dynamic community-based 'gender-neutral' tool. *Probation Journal*, 62(1), 33-48. doi:10.1177/0264550514562851

Table 1

Measure	Prediction of sexual recidivism	Prediction of violent recidivism	Prediction of general recidivism	Countries validated in	Offender types validated on
----------------	--	---	---	-----------------------------------	--

Violence Risk Scale - Sex Offender Version (VRS-SO; Wong, Olver, Nicholaichuk & Gordon, 2003)	AUC = .72, $r = .34$	AUC = .57 $r = .11$	N/A	Canada, New Zealand	Incarcerated male rapists and child molsters
STABLE 2007 (Hanson, Harris, Scott & Helmus, 2007)	AUC = .67	AUC = .66	AUC = .66	Austria, Canada, New Zealand, United States	Male adult parolees/probationers; rapists, child molesters and non-contact offenders
Violence Risk Scale (VRS; Wong & Gordon, 1998-2003)	N/A	AUC = .75 $r = .40$	AUC = .75 $r = .43$	Canada, New Zealand, The Netherlands	Incarcerated male offenders, forensic psychiatry inpatients, and personality-disordered offenders
Level of Service Inventory - Revised (LSI-R; Andrews & Bonta, 1995)	N/A	$r = .26$	$r = .37$	9 countries, including Singapore, United States, China, and Australia	Aboriginal, Asian, Black, and Hispanic offenders ; Female and male offenders
Self-Appraisal Questionnaire (SAQ; Loza, 2005)	N/A	$r = .32$	$r = .49$	Australia, Canada, England, Singapore, United States	Female and male offenders; Asian, African American, Hispanic and Aboriginal Australian offenders
Dynamic Risk Assessment for Offender Re-entry (Serin et al., 2012)	N/A	<i>n.s</i>	AUC = .62	New Zealand	Male and female offenders

Note. Predictive validity scores are sourced from the first validation studies for each measure, apart from the LSI-R for which a meta-analysis was the source (Gendreau, Goggin & Smith, 2002).