
VISITING EXPERTS' PAPERS

ISSUES IN THE ASSESSMENT OF JUVENILE OFFENDERS

*Robert D. Hoge**



I. INTRODUCTION

The assessment process involves the collection, processing, and synthesis of information about the individual. The outcome of the assessment is generally expressed as a judgment or opinion which may, in turn, be expressed as a diagnosis (e.g., bi-polar depression, psychotic, autistic) or categorization (e.g., borderline developmentally delayed, high risk for violent offending). As we will see below, formal and informal assessments are conducted in juvenile justice systems by police, prosecuting attorneys, probation officers, and judges, and these assessments are used as the basis for important decisions about the offender. Although not a primary concern in our discussions, assessments of risk for antisocial behaviours are also important in prevention programmes.

While the assessment process is critical to the quality of decisions made about the youth, many juvenile justice systems depend on badly flawed assessment processes. In some cases there are simply no systematic assessments conducted. In other cases the assessments are conducted by unqualified individuals or reflect an absence of adequate standards and procedures. Ample research now exists to show that justice systems that depend on structured and validated assessment procedures are more effective in producing reduced reoffending rates than those that do not use these procedures. The use of standardized assessments constitutes one of the most important of the principles of best practice (Andrews & Bonta, 2006; Grisso, 2005; Hoge, 1999, 2008; Hoge & Andrews, 1996).

II. CONTEXTS AND PURPOSES OF ASSESSMENTS

Table 1 provides an outline of the major decision areas encountered in juvenile justice systems. These range from the initial police contact with the youth to final discharge from the system.

Table 1
Major Decision Areas in Juvenile Justice Systems

Pre-arrest diversion
Arrest
Criminal charge
Pre-trial detention
Waiver to adult court
Competency to stand trial
Adjudication
Sentencing/disposition
Rehabilitative intervention planning

Some of these decisions relate to legal issues involving, for example, judgments about guilt or innocence. Those issues are outside the scope of our interest. However, many other decisions involve assessments relating to the psychological functioning of the youth or his or her circumstances. For example, a decision of the police or prosecutor to formally charge the youth may be affected by judgments about the youth's cognitive abilities or emotional maturity. Other decisions such as waiver to the mental health system may involve more complex diagnoses relating to the youth's psychiatric status. Disposition or sentencing

* Department of Psychology, Carleton University, Ottawa, Ontario, Canada.

decisions made by a judge or magistrate may be affected by a probation officer's assessment of the family circumstances of the youth. These are all important decisions and highlight the importance of conducting careful and valid assessments of the youth.

As suggested by the above examples, the focus of assessments will vary. In some cases the concern is with describing the criminal history of the individual. In other cases the concern is with describing or diagnosing internal conditions of the youth, relating, for example, to his or her emotional state or propensity for violence. In still other cases the goal is identification of problems existing with family circumstances, educational achievement, peer group associations, or substance abuse. Still another focus may be on the attitudes and values of the young person.

Many decisions require that the assessment be expressed as an evaluation of risk for engaging in future criminal behaviour (criminogenic risk). This may be a factor in decisions about pre-trial detention or diversion. For example, programmes designed to divert youth out of the criminal justice system without further processing are generally reserved for low risk individuals. Risk level may also be a consideration in deciding on an appropriate disposition following a finding of guilt. This could, for example, be the basis for deciding whether custody or community supervision is the appropriate courses of action.

Assessments may also be expressed as criminogenic needs. In this case we are attempting to identify the risk factors that can be changed through interventions to reduce the probability of future offending. For example, associations with antisocial peers is a risk factor, but it is something we can influence and, to the extent that we succeed, we can reduce the risk level. Needs assessments are very important wherever interventions or treatments are to be provided within the judicial action. We will talk more about risk and need assessments later in the paper.

Under some circumstances legal criteria may be available for guiding the assessment process. For example, specific psychiatric assessments may be indicated where decisions are to be made regarding competence to stand trial (Grisso, 1998, 2004). In still other cases regulations may guide the decision process. We will see in a later paper that the Canadian Province of Ontario has specified procedures to be followed in preparing pre-disposition reports for judges. However, in many cases rules or regulations for the conduct of assessments are not available, leaving open the possibility for decisions to be based on invalid or biased assessments.

The primary concern in our sessions will be with diversion and disposition decisions based on risk and need assessments. These assessments will be used to make decisions about the proper placement of the youth in the system and the kinds of interventions or treatments they should receive while in the system. We turn next to some issues relevant to this matter.

III. ISSUES IN ASSESSMENT

The following is an introduction to some key concepts relevant to the assessment process. The reader is referred to Grisso (1998), Grisso, Vincent, and Seagrave (2005), Hoge (1999, 2008), and Hoge and Andrews (1996) for further discussions of these issues.

A. Screening versus Assessment

A distinction is sometimes made between screening and assessment procedures, although the line between these is not always entirely clear. Screening instruments are generally relatively simple measures designed for use with all individuals within a group. The *Massachusetts Youth Screening Instrument—Version 2 (MAYSI-2)*; Grisso & Barnum, 2000), for example, is a self-report form used as a preliminary screening device for detecting emotional, behavioural, and psychological disturbances. It does not yield psychological diagnoses, but it does provide initial information about symptoms that may require more intensive assessments. Screening measures of this sort generally do not require a high level of training or expertise for administration or interpretation.

Psychological assessments, on the other hand, involve more thorough analyses of psychological or behavioural functioning. This might, for example, involve a comprehensive evaluation of cognitive and

personality functioning through the use of standardized tests and clinical interviews. This would be appropriate where signs of serious disorder are present. Many psychological assessments will require the services of a mental health practitioner such as a psychiatrist or psychologist.

However, as we will see below there are other assessments that can be conducted by non-mental health professionals such as probation officers, youth workers, or teachers. These generally involve a period of training and supervision in the use of some standardized measure.

B. Clinical versus Standardized Assessments

We can distinguish two general categories of assessment. Clinical assessments involve the unstructured collection of information and the interpretation of that information on the basis of past clinical experience. A probation officer, for example, might conduct an open-ended interview with a youth, interview parents, examine file information and conclude that the youth is high risk for continued criminal activity. Some clinical assessments do involve a certain amount of structure, but clinical interviews are relatively unstructured and allow the assessor considerable latitude in collecting and interpreting information. Decisions within juvenile justice systems are often based on clinical assessments.

Standardized assessments, on the other hand, represent more structured procedures for forming assessments. These are instruments or procedures with (a) fixed stimulus, response, and scoring formats; (b) yielding quantitative scores; and (c) for which normative and psychometric data are available. The *Wechsler Intelligence Scale for Children-IV (WISC-IV)* (Wechsler, 2004) and the *Minnesota Multiphasic Personality Inventory – Adolescent (MMPI-A)* (Butcher et al., 1992) are examples of standardized psychological measures.

Mechanical or actuarial prediction represents a special form of standardized assessment procedures whereby the procedures yield a specific prediction regarding an outcome. We will examine some comprehensive risk/need assessment instruments later and see that they yield specific estimates of risk for reoffending. These mechanical or actuarial predictions are always based on empirical analyses and provide information on reliability and validity.

We will examine another form of standardized assessment later. This is referred to as guided professional assessment or structured professional judgment (Webster, Hucker, & Bloom, 2002). This involves the use of clinical judgments within a structured framework. These instruments or procedures are based more or less directly on empirical data and may be accompanied by psychometric information. We will present examples in a later section.

Considerable research is now available demonstrating that the use of standardized assessments, particularly actuarial or guided professional assessments, is preferable to a dependence on clinical assessments (Borum & Verhaagen, 2006; Hoge, 1999, 2008).

C. Normative Scores

Norms or normative scores are also important assessment terms, particularly with reference to standardized measures. Normative scores reflect performance on a measure relative to the performance of a group of respondents. The *Wechsler Intelligence Scale for Children-IV (WISC-IV)* (Wechsler, 2004), for example, is scored in terms of a large and representative sample of children. An individual's score is expressed relative to that sample. A full-scale *WISC-IV* score of 100 for a nine-year old would indicate that the child's raw score was at the mean level for the normative sample of nine-year olds, while a score of 115 would indicate a performance one standard deviation above the mean. The quality of the normative data forms one basis for evaluating the utility of a standardized measure.

D. Evaluating Assessment Measures

It is very important to pay attention to the quality of a screening or assessment measure when deciding whether to incorporate it into the system. Reliability and validity are the primary qualities we look for when evaluating quality, and the basic forms of these constructs are defined in Table 2. Only a brief introduction will be provided to these terms in this paper, and you are referred to Grisso (2005), Hoge (2008), and Hoge and Andrews (1996) for more thorough discussions.

Reliability refers to the stability or consistency of a measure. More formally, it refers to the relative proportion of true and error variance in a measure.

Table 2
Definitions of Psychometric Terms

Reliability	The stability or consistency of a measure; formally defined as the relative proportion of true or error variance within a measure.
Content Validity	The adequacy with which a measure represents the conceptual domain it is expected to encompass.
Construct Validity	The theoretical meaning of scores from a measure; the accuracy with which the measure represents the construct in question.
Criterion-Related Validity	Extent to which scores from a measure relate to a criterion of performance; the two forms of criterion-related validity are concurrent and predictive validity.
Dynamic Predictive Validity	The sensitivity of a measure to changes in the dimension being assessed; also referred to as treatment validity.
Incremental Predictive Validity	The extent to which a measure exhibits improvements in prediction relative to other procedures.

Three standard procedures are available for evaluating reliability: test-retest, inter-rater agreement, and internal consistency. Each provides a somewhat different approach to detecting the extent to which extraneous or error factors are affecting scores on a measure. Reliability coefficients are generally expressed through correlation coefficients.

Reliability constitutes an essential condition in a measure. Lack of stability or consistency in a measure seriously interferes with its utility in applied assessment situations. If, for example, we found that scores on a personality test were affected by factors not related to the personality trait being assessed and that scores fluctuated in a more-or-less random fashion over time, we could have little confidence in that measure.

Validity is a more difficult construct to define since it is used in a number of different ways in different contexts. However, when referring to psychological tests or procedures, the term refers in its broadest sense to the meaningfulness of scores from a measure (Messick, 1995). Table 2 defines a number of different forms of validity, but only two will be noted in our discussion.

Construct validity is sometimes regarded as the key form of validity and may be defined as referring to the theoretical meaning or accuracy of a measure. It is also often defined as referring to the extent to which a measure is measuring what it says it is measuring.

Some illustrations of the definition may be useful. In raising a question of the construct validity of an intelligence test we would be raising a question about the meaningfulness of scores from the test. Just what does a full-scale score of 113 mean so far as the cognitive functioning of the youth is concerned? We could also ask how well that score reflects what we consider the meaning of "intelligence." Consider a second example. If we raised a question about the construct validity of a measure of behavioural pathology we would be asking about the actual meaning of scores from the measure. How do those scores define behavioural pathology? Construct validity may be evaluated through theoretical and empirical procedures.

Criterion-related validity is a second form of validity important for our purposes. It refers to the extent to which scores on a measure relate to some criterion of performance. The two forms of criterion-related validity are concurrent validity (where predictor and criterion scores are collected at the same time) and predictive validity (where predictor scores are collected at one point and criterion scores at a later time).

Criterion-related predictive validity is particularly important in applied settings because we often

need to know how well scores from a measure predict future behaviour or performance. For example, the comprehensive risk/need measures described below are designed to identify the current risk and need factors exhibited by the youth as a means of estimating their likelihood of engaging in continued criminal activity. Data from criterion-related predictive studies would provide us with that kind of information.

It is important to have some familiarity with the meaning of the reliability and validity constructs, and it is extremely important to obtain information about the reliability and validity of instruments being considered. Measures that do not display adequate levels of reliability and validity are of no value to us.

IV. TYPES OF ASSESSMENT INSTRUMENTS

This section will provide a brief overview of the major categories of assessment instruments and procedures relevant to juvenile justice settings. More thorough discussions may be found in Grisso (1998), Grisso et al. (2005), Hoge (2008), and Hoge and Andrews (1996). A further discussion of tools for assessing risk for violence in juveniles will be provided in the second paper in this series.

Two broad categories of assessment instruments can be identified. The first includes measures developed for general application but relevant to assessment in juvenile justice settings and the second includes instruments and procedures specifically developed for forensic application.

A. General Application Measures

A large number of personality tests, structured interview schedules, rating/checklist measures and attitude measures have been developed that have proven useful sometimes in assessing juvenile offenders.

The *Minnesota Multiphasic Personality Inventory-Adolescent (MMPI-A)*; Butcher, Williams, Graham, Archer, Tellegen, & Ben Porath, 1992) and *Reynolds Adolescent Depression Scale (RADS)*; Reynolds, 2002) are just two examples of the many standardized personality tests useful in the psychological assessment of adolescent offenders. These tests generally require special training in scoring and interpretation.

Structured interview formats designed for assessing behavioural and emotional pathologies may also play a role in these forensic assessments. Examples include the *Diagnostic Interview Schedule for Children (DISC)*; Shaffer et al., 2000) and *Child and Adolescent Functional Assessment Scale (CAFAS)*; Hodges, 2000).

Standardized rating and checklist measures have also proven very useful in these assessments. These may serve as screening tools for the preliminary identification of problems or as part of more intensive psychological assessments. The parent, teacher, clinician, and youth forms of the *Child Behavior Checklist (CBCL)*; Achenbach & Rescorla, 2001) has proven invaluable in identifying behavioural pathologies in youth. More focused rating instruments such as the *Massachusetts Youth Screening Instrument-Version 2 (MAYSI-2)*; Grisso & Barnum, 2000) and *Aggression Questionnaire (AQ)*; Buss & Warren, 2000) may also be useful. Some of the rating/checklist measures are only suitable for use by mental health professionals, but others can be used by probation officers, teachers, or youth workers with some training in scoring and interpretation.

B. Forensic Assessment Instruments

This category includes instruments and procedures specifically developed for assessments in juvenile justice systems. Some of these are specialized measures designed for evaluating legally relevant considerations. An example is the *MacArthur Competence Assessment Tool-Criminal Adjudication (MacCAT-CA)*; Poythress et al., 1999) instrument.

Several standardized self-report measures of antisocial attitudes, values, and beliefs are also available, including the *Criminal Sentiments Scale - Modified (CSS-M)*; Simourd, 1997) and *How I Think Questionnaire (HIT)*; Gibbs, Barriga, & Potter, 2001). The latter is especially important because it helps identify specific aspects of defective reasoning that can lead to antisocial actions.

Comprehensive risk/need assessment instruments constitute another important category of measures. These are generally in the form of structured checklists and employ either an actuarial or structured professional judgment approach. The instruments are designed to guide disposition and treatment planning

decisions by identifying risk and need factors in the youth. Two examples include the *Youth Level of Service/Case Management Inventory (YLS/CMI)*; Hoge, 2005; Hoge & Andrews, 1996) and *Estimate of Risk of Adolescent Sexual Offence Recidivism (ERASOR)*; Worling & Curwen, 2001). This type of measure will be discussed further in the second paper in the series.

V. PRACTICAL AND ETHICAL ISSUES

We will conclude the paper with brief introductions to some practical and ethical matters you should consider in designing an assessment system.

A. Selecting Relevant Measures

A choice of assessment measure or procedure should be guided, first, by the purposes of the assessment. There would be little value, for example, in using a personality test to aid in a decision about pretrial detention or an intelligence test to guide a decision about length of probation. It is also important to ensure that the assessment instrument is appropriate for the individual being assessed. This depends on the relevance of available normative, reliability, and validity data for the youth. For example, a personality test developed and evaluated with samples of boys between 8 and 12 years may not be relevant for a 17-year-old girl. Age, gender, ethnic group membership, and the presence of physical or mental handicaps are among the factors that should be considered in selecting assessment tools. Many of the standardized aptitude, personality, and behavioral measures have been evaluated for a wide range of respondent types, but this is not true of all instruments, and it is important to keep this issue in mind in selecting assessment tools.

B. Evaluating the Measures

We have already commented on the importance of researching the reliability and validity of the measures being considered. Information about these properties will be available from manuals or guides accompanying the instrument and from reference materials such as the *Mental Measurements Yearbook* and any number of web-based sources.

C. Cost

The cost of test materials and their administration is also a factor to be considered in evaluating the suitability of measures. Psychological services are sometimes expensive, and it is important to weigh those costs against the potential benefits of using the services. However, research shows that following principles of best practice, including the use of standardized assessment procedures, can lead to significantly reduced levels of reoffending. The savings there will often offset the costs of the assessment.

D. Professional Expertise

Standardized assessment instruments and procedures require varying levels of training and experience, and this must be considered in planning the assessment. As we have seen, some of the tools can be used by professionals such as probation officers, youth workers, or teachers with some special training. In other cases, however, the assessments must be conducted by qualified mental health professionals such as psychiatrists or psychologists. All professionals involved in the assessment should have a thorough understanding of child and adolescent development. A related point is that the use of the assessment instruments and procedures should be continually monitored, with periodic retraining provided for those using the measures.

E. Ethical and Legal Considerations

Some juvenile justice systems will provide explicit guidelines regarding the conduct of assessments, while others may provide more guidelines, and still others will give no guidance on the matter. One goal should be to encourage all systems to provide explicit guidelines. This will help to ensure that all youth are treated in a fair and consistent manner.

Professional associations in some areas have developed guidelines regarding the conduct of assessments. Examples include *Standards for Educational and Psychological Testing* developed by the American Educational Research Association and American Psychological Association and the *Specialty Guidelines for Forensic Psychologists* from Division 41 of the American Psychological Association. The British and Canadian associations of professional psychologists have also developed assessment guides.

REFERENCES

- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms and Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.
- Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct—Fourth edition*. Cincinnati, OH: Anderson.
- Borum, R., & Verhaagen, D. (2006). *Assessing and managing violence risk in youth*. New York: Guilford.
- Buss, A. H., & Warren, W. L. (2000). *Aggression questionnaire manual*. Los Angeles: Western Psychological Services.
- Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., & Kaemmer, B. (1992). *Minnesota Multiphasic Personality Inventory—Adolescent*. Minneapolis, MN: University of Minnesota Press.
- Gibbs, J. C., Barriga, A. Q., & Potter, G. B. (2001). *How I Think (HIT) questionnaire*. Champaign, IL: Research Press.
- Grisso, T. (1998). *Forensic evaluation of juveniles*. Sarasota, FL: Professional Resource Press.
- Grisso, T. (2005). Why we need mental health screening and assessment in juvenile justice systems. In T. Grisso, T. Vincent, & D. Seagrave (Eds.), *Mental health screening and assessment in juvenile justice* (pp. 3-21). New York: Guilford.
- Grisso, T., & Barnum, R. (2003). *Massachusetts Youth Screening Instrument—Version 2: User's manual and technical report*. Sarasota, FL: Personal Resource Press.
- Grisso, T., Vincent, G., & Seagrave, D. (Eds.). (2005). *Mental health screening and assessment in juvenile justice*. New York: Guilford.
- Hodges, K. (2000). *Child and Adolescent Functional Assessment Scale*. Ypsilanti, MI: Eastern Michigan University.
- Hoge, R. D. (1999). An expanded role for psychological assessments in juvenile justice systems. *Criminal Justice and Behavior*, 26, 251-266.
- Hoge, R. D. (2005). Youth Level of Service/Case Management Inventory. In T. Grisso, T. Vincent, & D. Seagrave (Eds.), *Mental health screening and assessment in juvenile justice* (pp. 283-294). New York: Guilford.
- Hoge, R. D. (2008). Assessment in juvenile justice systems. In R. D. Hoge, N. G. Guerra, & P. Boxer (Eds.), *Treating the juvenile offender* (pp. 54-75).
- Hoge, R. D., & Andrews, D. A. (1996). *Assessing the youthful offender: Issues and techniques*. New York: Plenum.
- Hoge, R. D., & Andrews, D. A. (2002). *Youth Level of Service/Case Management Inventory user's manual*. North Tonawanda, NY: Multi-Health Systems.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Poythress, N. G., Bonnie, R. J., Monohan, J., Otto, R., & Hoge, S. K. (2002). *The MacArthur Competence Assessment Tool—Criminal Adjudication: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Reynolds, W. M. (2002). *Reynolds Adolescent Depression Scale—Second Edition*. Lutz, FL: Psychological Assessment Resources.

Shaffer, D., et al. (2000). *Diagnostic Interview Schedule for Children (DISC-4)*. New York: New York State Psychiatric Institute.

Simourd, D. (1997). The Criminal Sentiments Scale-Modified and Pride in Delinquency Scale: Psychometric properties and construct validity of two measures of criminal attitudes. *Criminal Justice and Behavior*, 24, 52-70.

Webster, C., Hucker, S., & Bloom, H. (2002). Transcending the actuarial versus clinical polemic in assessment risk for violence. *Criminal Justice and Behavior*, 29, 659-665.

Wechsler, D. (2004). *Wechsler Intelligence Scale for Children—Fourth Edition*. San Antonio, TX: The Psychological Corporation.

Worling, J. R., & Curwen, M. A. (2001). *Estimate of Risk of Adolescent Sexual Offense Recidivism (ERASOR)*. Toronto, ON: Thistletown Regional Centre.