# Criminal Justice and Behavior

http://cjb.sagepub.com

---

**Assessing Adolescent Defendants' Adjudicative Competence:
Interrater Reliability and Factor Structure of the Fitness Interview
Test–Revised**

The online version of this article can be found at:
http://cjb.sagepub.com/cgi/content/abstract/33/4/467

---

Additional services and information for *Criminal Justice and Behavior* can be found at:

**Email Alerts:** http://cjb.sagepub.com/cgi/alerts

**Subscriptions:** http://cjb.sagepub.com/subscriptions

**Reprints:** http://www.sagepub.com/journalsReprints.nav

**Permissions:** http://www.sagepub.com/journalsPermissions.nav

**Citations** http://cjb.sagepub.com/cgi/content/refs/33/4/467

# ASSESSING ADOLESCENT DEFENDANTS' ADJUDICATIVE COMPETENCE

## Interrater Reliability and Factor Structure of the Fitness Interview Test–Revised

JODI L. VILJOEN
*University of Nebraska–Lincoln*

GINA M. VINCENT
*University of Massachusetts Medical School*

RONALD ROESCH
*Simon Fraser University*

As a result of changing legal standards, forensic clinicians have a greater likelihood of being faced with the task of assessing adolescents' adjudicative competence. This study examines the reliability and factor structure of the Fitness Interview Test, Revised Edition (FIT-R), in 152 male and female defendants ages 11 to 17. The interrater reliability of items and sections on the FIT-R is good. Most intraclass correlation coefficients (ICCs) for items fall between .60 and .91, and ICCs for section summary scores range from .82 to .91. Consistent with the

design of the FIT-R, confirmatory factor analysis supports a three-factor model, which includes understanding and reasoning about legal proceedings, appreciation of the charges and possible consequences of proceedings, and the ability to communicate with counsel. These factors are united by a dominant superordinate factor. Recommendations are made regarding the clinical use of the FIT-R in the assessment of adolescent competency.

Assessments of adjudicative competence, or competence to stand trial, are the most common forms of forensic pretrial evaluations, occurring at an estimated rate of 60,000 evaluations per year (Bonnie & Grisso, 2000). Historically, such evaluations were restricted to adult defendants because competency was deemed unnecessary and irrelevant to juvenile court proceedings, given the rehabilitative ideals of the early juvenile justice system (Bonnie & Grisso, 2000; Grisso, Miller, & Sales, 1987).

During the past few decades, the consequences of juvenile adjudicative proceedings have become more severe and transfers to adult court more common (McGuire, 1997; Snyder & Sickmund, 1999). As such, courts have increasingly required that juvenile defendants be competent to understand and participate in legal proceedings against them (Grisso, 1997; Redding & Frost, 2001). Therefore, forensic clinicians have a greater likelihood of being faced with the task of assessing juvenile defendants' adjudicative competency (see Redding & Frost, 2001).

There is considerable need for reliable and valid adjudicative competence assessment tools appropriate for juvenile defendants. Youth aged 15 and younger appear to be at heightened risk for impairments in the relevant legal abilities, relative to adults (Grisso et al., 2003). However, competence may be difficult to evaluate in children and adolescents because the relevant legal standards are vague, the psychological literature is limited, and developmental factors are likely to complicate assessments (Bonnie & Grisso, 2000; Redding & Frost, 2001).

Because of an absence of assessment instruments designed specifically for youth, many juvenile court clinicians report routine use of adult instruments, or modified versions of adult instruments, in juvenile

evaluations (Quinlan & Grisso, 2004). Similarly, researchers have also measured adolescents' adjudicative capacities using instruments that were designed for adults (e.g., Grisso et al., 2003). However, it is unclear whether such instruments are appropriate for adolescents. Grisso et al. (2003), for instance, expressed concern that the Appreciation subscale of the MacArthur Competency Assessment Tool for Criminal Adjudication (MacCAT-CA; Hoge, Bonnie, Poythress, & Monahan, 1999; Poythress et al., 1999), which is a competency assessment instrument designed for mentally ill adults, may not have adequately assessed appreciation in their adolescent sample.

To address this gap in knowledge about the developmental appropriateness of competence assessment tools, this study examined the applicability of a validated adult instrument to juvenile defendants. Specifically, interrater reliability, factor structure, and internal consistency of the Fitness Interview Test, Revised Edition (FIT-R; Roesch, Zapf, Eaves, & Webster, 1998) was evaluated among a sample of defendants aged 11 to 17. Prior to discussing the FIT-R, legal standards and models of adjudicative competence are reviewed.

## LEGAL STANDARD OF COMPETENCE

The *Dusky* standard of competence requires that defendants must have "factual understanding" of legal proceedings, "rational understanding" (or appreciation), and the "ability to communicate with counsel" (*Dusky v. the United States*, 1960). This legal standard is applied throughout the United States for adult and juvenile defendants adjudicated within the *adult criminal justice system* (Melton, Petrila, Poythress, & Slobogin, 1997; Redding & Frost, 2001) and is very similar to standards applied in Canada (Roesch et al., 1998).[1] Many courts remain undecided as to which standards should be adopted for youth adjudicated within the *juvenile justice system*. However, existing rulings have typically relied on the *Dusky* standard (Redding & Frost, 2001).

Although there is general agreement that *Dusky* is the primary legal standard, there is variation in its interpretation, which has led to several competing conceptual models of adjudicative competence (Rogers, Jackson, Sewell, Tillbrook, & Martin, 2003; Zapf, Skeem, & Golding, 2005). First, the *Dusky* standard can be interpreted quite

literally to mean that competency comprises three discrete abilities, namely, factual understanding, rational understanding, and communication with counsel. This model, which is referred to by Rogers et al. (2003, p. 345) as the "discrete-abilities model," is consistent with the design of the FIT-R, the instrument of focus in the present study.

As an alternative conceptualization, referred to as the "domains model" (Rogers et al., 2003, p. 345), Melton et al. (1997) proposed that competency can be thought of as two distinct abilities, specifically rational and factual understanding and the capacity to communicate with counsel. Rogers et al. (2003) described a second two-factor model, the "cognitive-complexity model" (p. 345), in which the relevant legal abilities are divided into basic "factual understanding" and the more cognitively complex "rational abilities" (rational understanding and the ability to communicate with counsel). Finally, Bonnie (1992, p. 297) in his influential theoretical model of competency, proposed that competency consists of "competency to assist counsel" (foundational competency) and "decisional competence" (the ability to reason and make specific legal decisions).

Several factor analytic studies have examined the fit of these models of competence in adult samples. Using confirmatory factor analysis (CFA), Rogers et al. (2003) compared the discrete-abilities, domains, and cognitive-complexity models of competence and determined that only the three-factor discrete-abilities model adequately explained the factor structure of the Evaluation of Competency to Stand Trial–Revised (Rogers, Tillbrook, & Sewell, 2004). Zapf et al. (2005) compared Bonnie's model to a three-factor model that roughly corresponded to the *Dusky* standard and found that the three-factor solution best explained the factor structure of the MacCAT-CA (Hoge et al., 1999; Poythress et al., 1999; see also Rogers, Grandjean, Tillbrook, Vitacco, & Sewell, 2001).

However, other studies, using various instruments and exploratory factor analytic techniques, have not found support for a three-factor structure (e.g., Bagby, Nicholson, Rogers, & Nussbaum, 1992; Rogers, Ustad, Sewell, & Reinhart, 1996; Ustad, Rogers, Sewell, & Guarnaccia, 1996). To date, no published studies have investigated the factor structure of competency assessment instruments with adolescent samples. Therefore, it is unclear whether these models are generalizable to young defendants.

## THE FIT-R

The FIT-R is a semistructured clinical interview that takes approximately 30 to 45 minutes to administer and comprises three scales that are consistent with the three-factor discrete-abilities model of competence just described. The first section, Understanding (Factual Knowledge), examines a defendant's understanding of the arrest process, current charges, role of key participants, legal process, pleas, and court procedures. The second section, Appreciation (Understanding of the Possible Consequences of Proceedings or Rational Understanding), examines a defendant's appreciation of the possible penalties, available legal defenses, and likely outcome. The third section, Communication with Counsel (Participation), examines a defendant's ability to communicate facts, relate to lawyers, plan legal strategy, engage in the defense, challenge witnesses, testify relevantly, and manage courtroom behavior.

Research with adult samples has provided evidence as to the FIT-R's reliability and validity. Viljoen, Roesch, and Zapf (2002) reported that the interrater reliability for overall clinical judgments of competency on the FIT-R was excellent (Average intraclass correlation coefficient [ICC] single raters [ICC1] = .88). ICCs for items ranged considerably but typically sufficiently met the criteria for fair or good reliability. In contrast, ICCs for clinical ratings of impairment on the three sections fell in the poor range.

Zapf and Roesch (2001) found support for the FIT-R's concurrent validity, reporting that it was moderately correlated with the MacCAT-CA, which is another assessment tool for adjudicative competence. The FIT-R also appears to have predictive validity, as it is able to distinguish between defendants who are and are not judged to be incompetent by clinicians (Zapf & Roesch, 1997; Zapf, Roesch, & Viljoen, 2001). If the FIT-R is to be used with child and adolescent defendants, however, further research is required on its psychometric properties with this population.

### RESEARCH NEEDS

*Interrater reliability*. No research has examined the interrater reliability of the FIT-R specifically in adolescent samples. Also, the low interrater reliability for section ratings found by Viljoen et al. (2002)

is somewhat disconcerting. The reliability of sections might potentially be enhanced through summing item scores rather than relying on separate structured clinical judgments of impairment, and/or through the provision of example ratings to anchor examiners' ratings.

*Structural validity*. The FIT-R is based on the three-factor model of competence, which includes factual understanding, rational understanding or appreciation, and the ability to communicate with counsel. To date, however, this proposed structure has not been empirically investigated. Defining a test's structure is important for establishing its construct validity. According to the latent-trait theory of measurement, the first step in evaluating test performance is to determine whether a test conforms to its theoretical factor structure (Maraun & Jackson, 2001; McDonald, 1999; Thissen, Steinberg, Pyszczynski, & Greenberg, 1983).

Psychological or psycholegal constructs, such as adjudicative competency, are typically conceptualized as latent traits, or unobservable variables, that are inferred from observable, behavioral processes (Embretson & Reise, 2000). Latent-trait methods for investigating structural validity, such as CFA, purport to link item responses directly to unobservable, latent traits by removing random error and differences in examinee ability. As such, these procedures can inform us as to the number of traits underlying a given competency test, the relative fit of various models or conceptualizations of competence, and they may even provide guidance in understanding the "structure" of competence.

## PURPOSE OF THIS STUDY

In the interest of identifying assessment tools that will be useful for juvenile populations, the psychometric properties of the FIT-R were evaluated among a sample of juvenile defendants. Specifically, this study examined interrater reliability for items, sections, and overall determinations of competence and compared the interrater reliability of structured clinical ratings to numerical summary scores. It was hypothesized that numerical summary scores would be more reliable than structured clinical ratings.

In addition, the structural validity of the FIT-R was investigated using CFA to test whether the FIT-R adhered to its theoretical

three-factor discrete-abilities structure (Understanding, Appreciation, and the Ability to Communicate With Counsel). The fit of this model was compared to a simple unidimensional (one-factor) model and CFA models based on other conceptual frameworks of adjudicative competence reviewed earlier (e.g., the domains model and cognitive-complexity model).

## METHOD

### PARTICIPANTS

Participants included 152 pretrial defendants (73 females and 79 males), aged 11 to 17 ($M = 14.52$, $SD = 1.68$), held in a detention facility in the state of Washington. The majority of defendants remanded to this facility were 15 and older. The sample was stratified by age (11 to 13, 14 to 15, and 16 to 17) to ensure that younger defendants were adequately represented. An equal number of invitations to participate were extended to adolescents randomly selected from each age group. The rate of agreement for participation was 94.4%. Defendants who did not participate ($n = 9$) appeared representative of the larger sample in terms of age, gender, race, and current charge. All participants indicated that English was their first language, or the language they spoke at home or at school.

The average IQ of participants was 82.57 ($SD = 13.91$). Although low, this is comparable to other samples of delinquent youth (e.g., Grisso et al., 2003). Sixty percent of participants ($n = 92$) in the overall sample were non-Hispanic Caucasian, 26.3% ($n = 40$) were African American, 7.9% ($n = 12$) were Hispanic, 3.9% ($n = 6$) were Native American, and 1.3% ($n = 2$) were Asian. The majority of participants (66.7%, $n = 96$) were classified as being at the two lowest socioeconomic levels (Levels IV and V), according to Hollingshead's (1975) classification system. Most participants had committed a violent offense (37.5%, $n = 57$) or a property offense (36.8%, $n = 56$).[2] The remaining participants (25.7%, $n = 39$) had committed another type of offense, such as a drug offense, obstruction, or failure to appear in court.

**MATERIALS**

The FIT-R is a semistructured clinical interview that assesses legal abilities relevant to adjudication, including understanding and appreciation of adjudicative proceedings, and the ability to communicate with counsel. Items on the FIT-R are clinically rated on a 3-point scale according to one's level of impairment (0 = *definite impairment*, 1 = *possible impairment*, and 2 = *no impairment*). Clinical judgments regarding level of impairment are made for each section and for the overall test using similar 3-point scales. In rating sections and making overall judgments, evaluators are instructed to consider impairment on individual items. However, item scores are not summed to calculate section and global impairment scores (Roesch et al., 1998). Instead, these ratings constitute a separate clinical judgment. In addition to these structured clinical judgments of impairment, item scores in the present study were summed to form numerical summary scores for each section and the overall instrument. Also, a more detailed scoring system for the FIT-R was developed for this study, in which three scoring examples were provided for each item. These examples were derived from actual participant responses.

**PROCEDURES**

All study procedures were approved by the appropriate review boards of Simon Fraser University and the juvenile detention facility that was the study site and were consistent with current ethical procedures. Confidentiality was assured except in cases of risk of harm to self or others. Identifying information was not recorded, and participants were instructed not to provide details on the current charge for which they were undergoing adjudication.

Information about the study was presented orally to individuals who expressed interest in participating, and a form was also provided so that potential participants could read the information presented. Participants were tested, using an adapted version of the MacCAT for Clinical Research (Appelbaum & Grisso, 2001) to assess if they understood and appreciated study procedures and were able to make a stable choice about participation. If a participant showed inadequate comprehension of a concept, this concept was

reexplained. In addition, the institutional administrator provided consent for all participating defendants.

The FIT-R was administered to all study participants by a doctoral student in clinical psychology trained in the use of study instruments. Training on the FIT-R included familiarization with the test manual, viewing a training videotape, and completion of several practice protocols. To assess interrater reliability, 26 randomly selected protocols were recoded by a second doctoral student in clinical psychology, with similar training as the original rater.

On completion of the study, participants were given 10 points as compensation for their time. These points could be used to pay for food and toiletry items at the facility and was enough to buy approximately two chocolate bars or a small bottle of shampoo. This amount was thought to be sufficient to compensate participants for their time but not enough to coerce participation.

### DATA ANALYSIS

Analyses were performed to assess interrater reliability, factor structure, and internal consistency. Interrater reliability was assessed using ICCs for single raters with a two-way mixed effect model (Model 2, McGraw & Wong, 1996). The FIT-R's factor structure was tested using fit indices derived from CFA in Mplus (Muthén & Muthén, 2001). Because the scale of measurement for FIT-R items is best described as ordered categorical, CFA fit indices were based on the weighted root mean square residual (WRMR) estimator intended for categorical data. Although many studies of similar scales have reported fit indices based on maximum-likelihood estimators, intended for continuous data, the use of these factor analytic techniques with categorical data may not be appropriate, especially if the data are skewed (Maraun & Rossi, 2001; van Schuur & Kiers, 1994). In testing models, residual variances of variables were allowed to be parameters in the model, but scale factors were not.

A number of fit indices were used to evaluate the CFA models, including the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and WRMR. The use of multiple indices of fit is recommended given discrepancies in opinions as to which index is best (Byrne, 1994;

Hu & Bentler, 1998). Consistent with suggestions of psychometricians (Byrne, 1994; Hu & Bentler, 1998; Schwarzer, 1986), a model was judged to have adequate fit when the CFI and TLI fell between .90 and .94 and the RMSEA fell between .05 and .09 and good fit when the CFI and TLI were .95 or higher, the RMSEA was less than .05, and the WRMR was less than .90 (Yu & Muthén, 2001).

# RESULTS

## ITEM AND SECTION DISTRIBUTIONS

Means, standard deviations, and skew are presented in Tables 1 and 2 for FIT-R items, sections, and global ratings. Item means ranged from 0.84 to 1.64, with most falling between 1 and 1.5. An item or scale can be considered skewed when the skew is greater than twice the standard error. Based on this criterion, 8 of the 16 items were skewed (see Table 1), providing further justification for the use of WRMR estimator techniques.

## INTERRATER RELIABILITY

Interrater reliability was evaluated for item scores, section summary scores and structured clinical ratings, and total test scores and ratings. With the exception of three items with relatively low interrater reliability (i.e., Item 14: Capacity to Communicate Facts; Item 15: Capacity to Testify Relevantly; and Item 16: Capacity to Manage Behaviour), item ICCs for single raters ($ICC_1$s) were generally good (see Cicchetti, Showalter, & Tyrer, 1985), with most falling between .60 and .75 (see Table 1). $ICC_1$s for summary scores were excellent (Cicchetti et al., 1985) for the sections measuring Understanding (.91), Appreciation (.82), and Communication (.83), with a .91 for the overall test score. For structured clinical ratings, agreement was somewhat lower but still acceptable, ranging from .59 to .80 (see Table 2).

## FACTOR STRUCTURE

*FIT-R's three-factor structure*. First, the theoretical three-factor structure of the FIT-R was tested by loading Items 1 through 6, Items

**TABLE 1:  FIT-R Item Distributions and Reliability**

|  | M | SD | *Skew* | *ICC* |
|---|---|---|---|---|
| Section 1: Understanding | | | | |
| 1. Understanding of arrest process | 1.49 | .63 | −0.86[a] | .74 |
| 2. Understanding of current charges | 1.51 | .69 | −1.06[a] | .60 |
| 3. Understanding of key participants | 1.19 | .78 | −0.35 | .83 |
| 4. Understanding of legal process | 0.84 | .74 | 0.26 | .72 |
| 5. Understanding of pleas | 0.90 | .74 | 0.15 | .71 |
| 6. Understanding of court procedures | 1.08 | .73 | −0.12 | .73 |
| Section 2: Appreciation | | | | |
| 7. Appreciation of possible penalties | 1.30 | .78 | −0.57[a] | .53 |
| 8. Appraisal of available defenses | 1.12 | .85 | −0.23 | .64 |
| 9. Appraisal of likely outcome | 1.51 | .69 | −1.09[a] | .69 |
| Section 3: Communication | | | | |
| 10. Capacity to communicate facts | 1.64 | .53 | −1.10[a] | .08 |
| 11. Capacity to relate to lawyer | 1.28 | .68 | −0.43[a] | .65 |
| 12. Capacity to plan strategy | 1.06 | .79 | −0.11 | .73 |
| 13. Capacity to engage in defense | 1.24 | .67 | −0.33 | .63 |
| 14. Capacity to challenge witnesses | 1.11 | .71 | −0.16 | .73 |
| 15. Capacity to testify relevantly | 1.59 | .57 | −1.03[a] | .39 |
| 16. Capacity to manage behavior | 1.64 | .52 | −1.04[a] | .16 |

*Note.* FIT-R = Fitness Interview Test, Revised Edition; ICC = intraclass correlation coefficient.
a. Items that are significantly skewed.

7 through 9, and Items 10 through 16, respectively, onto three distinct factors representing the test's sections. This model had adequate fit with respect to tests of absolute fit (CFI and TLI) but inadequate FIT-R on indices of relative fit (RMSEA and WRMR) (see Table 3). To determine whether this three-factor structure was better represented as a nested model, another CFA was conducted after constraining each first-order factor to load equally onto a dominant superordinate factor with the covariance of the dominant factor set at one. Again, this model had adequate fit with respect to tests of absolute FIT-R but inadequate fit on indices of relative fit.

*A priori tests of other competence conceptual models.* The less than optimal fit of the FIT-R's prescribed factor structure called for further investigations to identify a better factor structure. To rule out the possibility of a common-factor structure, the fit of a simple

**TABLE 2: FIT-R Section Scores and Reliability**

|  | M | SD | *ICC* |
|---|---|---|---|
| Section 1: Understanding |  |  |  |
| Structured clinical rating | 1.34 | 0.69 | .65 |
| Numerical summary score | 7.01 | 3.10 | .91 |
| Section 2: Appreciation |  |  |  |
| Structured clinical rating | 1.34 | 0.78 | .80 |
| Numerical summary score | 3.92 | 1.81 | .82 |
| Section 3: Communication |  |  |  |
| Structured clinical rating | 1.47 | 0.63 | .59 |
| Numerical summary score | 9.54 | 3.26 | .83 |
| Global rating |  |  |  |
| Structured clinical rating | 2.47 | 0.74 | .69 |
| Numerical summary score | 20.35 | 7.21 | .91 |

*Note.* FIT-R = Fitness Interview Test, Revised Edition; ICC = intraclass correlation coefficient. Structured clinical ratings for sections and the global rating can be scored as 0, 1, or 2. The total possible numerical summary score is 12 for Understanding, 6 for Appreciation, 14 for Communication, and 32 for the total score.

unidimensional model was tested by constraining all of the FIT-R items to load onto one common factor. As reported in Table 3, the unidimensional model also did not achieve acceptable fit.

Next, two additional models of competence were examined. Specifically, the domains model was tested using a two-factor structure that included Understanding (Items 1 through 9) and Communication With Counsel (Items 10 through 16), and the cognitive-complexity model was tested using a different two-factor structure that included Factual Understanding (Items 1 through 6) and Rational Abilities (Items 7 through 16). Both two-factor structures were examined with and without the inclusion of a dominant factor. None of these models achieved an acceptable fit on indices of relative fit (see Table 3).

*Adjusted discrete-abilities model.* Based on the results of Zapf et al. (2005), it was hypothesized that the FIT-R items that appeared to tap into reasoning abilities (i.e., Items 12 and 13) might be more

**TABLE 3:  Confirmatory Factor Analyses**

| Model | CFI | TLI | RMSEA | WRMR |
|---|---|---|---|---|
| Theoretical FIT-R factor structure | | | | |
| Discrete-abilities model | .929 | .959 | .107 | 1.038 |
| Discrete-abilities model, dominant factor added | .939 | .962 | .102 | 1.023 |
| Other conceptual models of competence | | | | |
| One factor model | .904 | .941 | .128 | 1.205 |
| Domains model | .933 | .959 | .107 | 1.038 |
| Domains model, dominant factor added | .933 | .959 | .107 | 1.038 |
| Cognitive-complexity model | .914 | .950 | .118 | 1.101 |
| Cognitive-complexity model, dominant factor added | .914 | .950 | .118 | 1.101 |
| Adjusted models | | | | |
| Adjusted discrete-abilities model | .943 | .968 | .095 | 0.935 |
| Adjusted discrete-abilities model, dominant factor added | .956 | .973 | .088 | 0.940 |

*Note.* CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; WRMR = Weighted Root Mean Square Residual; FIT-R = Fitness Interview Test, Revised Edition.

strongly related to the latent variable for understanding, and that items that appeared to tap into personally relevant and case-specific information (i.e., Items 1 and 2) might be more indicative of the latent variable for appreciation. Therefore, an adjusted discrete-abilities model, which consisted of Understanding and Reasoning (Items 3, 4, 5, 6, 12, and 13), Appreciation of Case-Specific Information (Items 1, 2, 7, 8, and 9), and Communication With Counsel (Items 10, 11, 14, 15, and 16), was examined.

This adjusted discrete-abilities model achieved adequate to good fit when a dominant superordinate factor was added (CFI = .956, TLI = .973, RMSEA = .088, and WRMR = .940). The three factors correlated fairly highly, with .81 for Factors 1 and 2, .75 for Factors 1 and 3, and .73 for Factors 2 and 3. These factors also had adequate internal consistency, according to standards provided by Cronbach (1990) and Nunnally (1978). Alpha was .85 for Factor 1 (Understanding/ Reasoning), .78 for Factor 2 (Appreciation), and .80 for Factor 3 (Communication).

## DISCUSSION

In a recent survey of practices for evaluating juvenile competency, most psychologists reported that they considered the use of competency assessment instruments to be an essential or recommended component of these evaluations (Ryba, Cooper, & Zapf, 2003b). However, all existing competency instruments have been developed for adult defendants, and little research has investigated the psychometric properties of such instruments with adolescents. In response to this limitation, this study examined the reliability and structural validity of the FIT-R for a sample of adolescent defendants.

### INTERRATER RELIABILITY

The FIT-R was designed as a structured clinical assessment instrument. Ratings on the FIT-R are made based on structured clinical judgments rather than actuarial scoring algorithms. However, previous research has found that the reliability of structural clinical judgments of sections on the FIT-R is lower than desired (Viljoen et al., 2002). As such, this study compared the reliability of structured clinical ratings of sections to numerical summary scores.

Results indicated that numerical summary scores for sections and overall judgments of competency were more reliable than structured clinical ratings, although structured clinical ratings still had adequate reliability. Specifically, ICCs for section summary scores ranged from .82 to .93, whereas ICCs for structured clinical ratings ranged from .59 to .80. The ICCs of the structured clinical ratings of sections were higher in this study than in the Viljoen et al. (2002) study, which may partly stem from the provision of scoring examples in the present study.

The FIT-R manual instructs examiners to consider clinical observations in ratings of defendants' performance (Roesch et al., 1998). In this study, however, the second rater scored the FIT-R based on written transcripts of examinee responses and did not have access to test observation information. Interrater reliability was lowest for items placing higher weight on clinical observations; namely, Ability to Communicate Facts, Capacity to Manage Behavior, and Capacity to Testify Relevantly. For these items, examiners are

advised to consider the defendant's attention, impulsivity, and thought processes (Roesch et al., 1998). It is possible that the ICCs of items and sections on the FIT-R would have been higher had the second rater used live observation or videotaped responses in scoring. Therefore, these findings may underestimate interrater reliability, which was good despite the lack of clinical observations.

### FACTOR STRUCTURE

The FIT-R was developed based on a three-factor interpretation of Canadian and American legal standards for adjudicative competence, which includes understanding of the nature and object of the legal proceedings (factual understanding), appreciation (rational understanding), and the ability to communicate with counsel. The results indicate that the factor structure of the fit is generally consistent with this framework. Specifically, the best fit was obtained for a three-factor adjusted discrete-abilities model, which included understanding and reasoning about legal proceedings, appreciation of case-specific information, and the ability to communicate with counsel. These factors were united by a dominant unidimensional factor. Viewed in combination with the poor fit of one-factor and two-factor models, competence, at least as measured by the FIT-R, appears to consist of three related but distinct abilities.

The failure to find a strong first-order unidimensional factor on the FIT-R emphasizes the need to consider specific legal abilities on this instrument rather than solely a global summary score. This notion is consistent with legal standards, which assert that an individual need not show global impairments to be judged incompetent but instead can be incompetent when impaired in only a single domain (Roesch et al., 1998). On the other hand, the evidence for a second-order unidimensional factor suggests that the specific factors on the FIT-R tap into a common construct.

In general, items in the supported model were kept on the scales developed by the test authors with a few exceptions. First, Items 12 (Ability to Plan Legal Strategy) and 13 (Ability to Engage in Defence) were moved from Factor 3 (Communication) to Factor 1 (Understanding/Reasoning). These items appear to measure reasoning and are most closely akin to Bonnie's (1992) concept of decisional

competence.[3] Zapf et al. (2005) similarly found that model fit was improved after moving several reasoning items to the understanding factor of the MacCAT-CA.

In addition, Items 1 (Understanding of the Arrest Process) and 2 (Understanding of the Current Charges) were moved from Factor 1 (Understanding/Reasoning) to Factor 2 (Appreciation). Conceptually, it makes sense that these items measure appreciation. Specifically, it may be necessary for defendants to appreciate their arrest and charges in order to appreciate the possible consequences of legal proceedings. Also, it is notable that the content of Items 1 and 2 pertain to case-specific information, consistent with all the items on Factor 2. Similarly, factor analysis of the MacCAT-CA revealed that items measuring case-specific information tended to load together in adults (Zapf et al., 2005).

### LIMITATIONS AND RESEARCH RECOMMENDATIONS

As mentioned previously, it is likely that this study would have produced more precise estimates of the FIT-R's interrater reliability if the second rater had access to clinical observations. Furthermore, it is important to note that this study did not compare the validity of structured clinical ratings and numerical summary scores. Although structured clinical ratings are less reliable than numerical summary scores, they may be more valid when it comes to legal decision making. For example, courts recognize that it is possible for an individual to be found incompetent on the basis of severe impairment on a single item alone. Therefore, the practice of summarizing item scores may compromise the validity of any instrument designed to assess adjudicative competence (Roesch et al., 1998). Future research is needed to clarify the relative merits of these differing approaches. It is interesting that in the field of violence risk assessments, recent research has suggested that structured clinical judgments may be more valid and equally as reliable as numerical summary scores (Dempster, 1998; Douglas, Ogloff, & Hart, 2003; Kropp & Hart, 2000).

With respect to factor analyses, it is important to note that the model of competence derived in this study does not necessary reflect "the 'truth' about the construct that underlies competency standards" (Zapf et al., 2005, p. 434). One unavoidable problem is that factor

analytic models are largely dependent on the nature of the items included in instruments, which vary considerably across instruments. Another difficulty is that accuracy is compromised when applying factor analytic techniques to measures comprising dichotomous or ordered-categorical items (Maraun & Rossi, 2001; van Schuur & Kiers, 1994). In part, this is because standard maximum-likelihood procedures are linear and dependent on normality within item score distributions. To address this issue, Mplus was used because it was designed specifically to handle ordered-categorical data, and WRMR techniques were applied. However, future research should apply item response theory and other nonparametric techniques to assess test structure and dimensionality.

Finally, it should be emphasized that the results will not necessarily generalize to adult samples. To date, no studies have investigated the factor structure of the FIT-R in adult defendants. Future research might examine potential differences in the factor structure of the FIT-R and other competency assessment instruments in samples of adults and youth of various ages.

**IMPLICATIONS**

Study results indicated that the interrater reliability of the FIT-R is adequate and its factor structure is relatively consistent with its rationale and organization when used with juvenile defendants. These findings provide preliminary empirical support for the psychometric properties of the FIT-R with youth. However, several issues are important to consider in the clinical use of the FIT-R with juvenile defendants. First, evaluators should determine whether the FIT-R is consistent with the legal standards for adolescent competency in their jurisdiction. The FIT-R was designed to measure legal abilities relevant to adjudicative competency in adults (Roesch et al., 1998). Although many courts have applied adult standards of adjudicative competency to adolescents, particularly when adolescents are tried in adult criminal courts, some courts have suggested that lower standards of competency be used for youth adjudicated in juvenile courts (Redding & Frost, 2001). The FIT-R may not be appropriate for use in jurisdictions that apply different competency standards for youth.

Second, it is important to note that competency assessment instruments such as the FIT-R compose only one part of adolescent competency evaluations. Given that courts typically require that legal impairments arise as a result of psychopathology or poor cognitive abilities, evaluators should routinely assess these constructs (Grisso, 2003). It is also recommended that evaluators assess developmental maturity (see Ryba, Cooper, & Zapf, 2003a, for a description of various approaches to assessing maturity), as research has indicated that adolescents may have limitations in legal capacities as a result of developmental immaturity rather than solely because of psychopathology or cognitive deficits (Grisso et al., 2003; see also Scott, Reppucci, & Woolard, 1995; Steinberg & Cauffman, 1996, for a discussion of these issues). Although courts in most jurisdictions have not yet determined whether developmental immaturity constitutes an adequate basis for a possible finding of incompetence (Grisso, 1997), at least one court has ruled that it does (*In re Causey*, 1978).

There are a number of reasons why it might be valuable to use competency assessment instruments such as the FIT-R in assessing adolescent competency. In general, it has been argued that competency assessment instruments may be more reliable than traditional clinical judgments and may help ensure that clinicians adequately address the areas relevant to legal competencies (Grisso, 2003; Nicholson, Briggs, & Robertson, 1988). Nevertheless, it is important to recognize that research has not directly compared various approaches to assessing competency in adolescents, and few studies have even compared these approaches in adults (see Schreiber, Roesch, & Golding, 1987, for an exception). Such comparisons could provide clinicians with empirical guidance and possibly bring about a resolution to ongoing debates over which approach should be favored.

## NOTES

1. The Criminal Code of Canada (1985) states that adult defendants must have an "understanding of the nature and object of legal proceedings" (or factual understanding), "understanding of the possible consequences of legal proceedings" (rational understanding or appreciation), and the "ability to communicate with counsel." This standard has been applied to adolescents in Canada as well (e.g., *R. v. D. (W.A.L.-1)*, 2002; *R. v. W.(C.)*, 2001).

2. Charges were classified based on the most serious offense.

3. Item 12 examines a defendant's reasoning about decisions to accept plea bargains, plead guilty, and testify in court. Also, it investigates whether defendants would consult with attorneys or defer to attorneys in these decisions and how they would manage disagreements with their lawyer. Item 13 provides defendants with several scenarios (i.e., their attorney finds a way to get their charges dropped, their attorney recommends to appeal the case, and their attorney is able to get a plea bargain) and questions defendants on whether they would accept this plan and their reasons for this decision.

# REFERENCES

Appelbaum, P. S., & Grisso, T. (2001). *MacArthur Competency Assessment Tool for Clinical Research (MacCAT-CR)*. Sarasota, FL: Professional Resource Press.

Bagby, R. M., Nicholson, R. A., Rogers, R., & Nussbaum, D. (1992). Domains of competency to stand trial: A factor analytic study. *Law and Human Behavior*, *16*, 491-507.

Bonnie, R. J. (1992). The competence of criminal defendants: A theoretical reformulation. *Behavioral Sciences and the Law*, *10*, 291-316.

Bonnie, R. J., & Grisso, T. (2000). Adjudicative competence and youthful offenders. In T. Grisso & R. G. Schwartz (Eds.), *Youth on trial: A developmental perspective on juvenile justice* (pp. 73-103). Chicago: University of Chicago Press.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows*. London: Sage.

Cicchetti, D. V., Showalter, D., & Tyrer, P. J. (1985). The effect of number of rating scale categories on levels of interrater reliability: A Monte Carlo investigation. *Applied Psychological Measurement*, *9*, 31-36.

Criminal Code, R.S.C. 1985, c. C-46, as amended.

Cronbach, L. J. (1990). *Essentials of psychological testing*. New York: HarperCollins.

Dempster, R. J. (1998). *Prediction of sexually violent recidivism: A comparison of risk assessment instruments*. Unpublished master's thesis, Simon Fraser University, Burnaby, British Columbia, Canada.

Douglas, K. S., Ogloff, J. R. P., & Hart, S. D. (2003). *Reliability and validity of clinical decisions regarding violence risk made using the HCR-20: A test of the structured professional judgment approach to risk assessment*. Unpublished manuscript.

Dusky v. the United States, 362 U.S. 402 (1960).

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Kansas City, MO: Multivariate Book Series.

Grisso, T. (1997). The competence of adolescents as trial defendants. *Psychology, Public Policy, and Law*, *3*, 3-32.

Grisso, T. (2003). *Evaluating competencies: Forensic assessments and instruments* (2nd ed.). New York: KluwerAcademic/Plenum.

Grisso, T., Miller, M. O., & Sales, B. (1987). Competency to stand trial in juvenile court. *International Journal of Law and Psychiatry*, *10*, 1-20.

Grisso, T., Steinberg, L., Woolard, J., Cauffman, E., Scott, E., Graham, S., et al. (2003). Juveniles' competence to stand trial: A comparison of adolescents' and adults' capacities as trial defendants. *Law and Human Behavior*, *27*, 333-363.

Hoge, S. K., Bonnie, R. J., Poythress, N., & Monahan, J. (1999). *The MacArthur Competence Assessment Tool—Criminal Adjudication*. Odessa, FL: Psychological Assessment Resources.

Hollingshead, A. (1975). *Four factor index of social status*. New Haven, CT: Department of Sociology, Yale University.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparametrized model misspecification. *Psychological Methods*, *3*, 424-453.

In re Causey, 363 So. 2d 472 (La. 1978).

Kropp, P. R., & Hart, S. D. (2000). The Spousal Assault Risk Assessment (SARA) guide: Reliability and validity in adult male offenders. *Law and Human Behavior*, *24*, 101-118.

Maraun, M. D., & Jackson, J. S. H. (2001). *On the justification of test analytic claims*. Unpublished manuscript.

Maraun, M. D., & Rossi, N. T. (2001). The extra-factor phenomenon revisited: Unidimensional unfolding as quadratic factor analysis. *Applied Psychological Measurement*, *25*, 77-87.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*, 30-46.

McGuire, M. (1997). C. 19, An act to amend the Young Offenders Act and the Criminal Code: "Getting tougher?" *Canadian Journal of Criminology*, *39*, 185-214.

Melton, G. B., Petrila, J., Poythress, N. G., & Slobogin, C. (1997). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers* (2nd ed.). New York: Guilford.

Muthén, L. K., & Muthén, B. O. (2001). *Mplus user's guide* (2nd ed.) [Computer software and manual]. Los Angeles: Author.

Nicholson, R. A., Briggs, S. R., & Robertson, H. C. (1988). Instruments for assessing competency to stand trial: How do they work? *Professional Psychology: Research and Practice*, *19*, 383-394.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Poythress, N., Nicholson, R., Otto, R. K., Edens, J. F., Bonnie, R. J., Monahan, J., et al. (1999). *The MacArthur Competence Assessment Tool—Criminal Adjudication: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Quinlan, J., & Grisso, T. (2004, March). *Juvenile court clinical services: A nationwide survey*. Paper presented at the annual convention of the American Psychology-Law Society, Scottsdale, AZ.

R. v. D. (W.A.L.-1), 2002 SKPC 38.

R. v. W. (C.), 2001 ABPC 148.

Redding, R., & Frost, L. (2001). Adjudicative competence in the modern juvenile court. *Virginia Journal of Social Policy and the Law*, *9*, 353-410.

Roesch, R., Zapf, P. A., Eaves, D., & Webster, C. D. (1998). Fitness Interview Test (Rev. ed.). Burnaby, British Columbia, Canada: Mental Health, Law and Policy Institute, Simon Fraser University.

Rogers, R., Grandjean, N. R., Tillbrook, C. E., Vitacco, M. J., & Sewell, K. W. (2001). Recent interview-based measures of competency to stand trial: A critical review augmented with research data. *Behavioral Sciences and the Law*, *19*, 503-518.

Rogers, R., Jackson, R. L., Sewell, K. W., Tillbrook, C. E., & Martin, M. A. (2003). Assessing dimensions of competency to stand trial: Construct validation of the ECST-R. *Assessment*, *10*, 344-351.

Rogers, R., Tillbrook, C. E., & Sewell, K. W. (2004). *Evaluation of Competency to Stand Trial™–Revised (ECST–R)*. Lutz, FL: Psychological Assessment Resources.

Rogers, R., Ustad, K. L., Sewell, K. W., & Reinhart, V. (1996). Dimensions of incompetency: A factor analytic study of the Georgia Court Competency Test. *Behavioral Sciences and the Law*, *14*, 323-330.

Ryba, N. L., Cooper, V. G., & Zapf, P. A. (2003a). Assessment of maturity in juvenile competency to stand trial evaluations: A survey of practitioners. *Journal of Forensic Psychology Practice*, *3*, 23-45.

Ryba, N. L., Cooper, V. G., & Zapf, P. A. (2003b). Juvenile competence to stand trial evaluations: A survey of current practices and test usage among psychologists. *Professional Psychology: Research and Practice*, *34*, 499-507.

Schreiber, J., Roesch, R., & Golding, S. (1987). An evaluation of procedures for assessing competency to stand trial. *Bulletin of the American Academy of Psychiatry and Law*, *15*, 187-203.

Schwarzer, R. (1986). Evaluation of convergent and discriminant validity by use of structural equations. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaires: Current issues in theory and measurement* (pp. 191-213). Berlin, Germany: Springer-Verlag.

Scott, E. W., Reppucci, N. D., & Woolard, J. L. (1995). Evaluating adolescent decision making in legal contexts. *Law and Human Behavior*, *19*, 221-244.

Snyder, H. N., & Sickmund, M. (1999). *Juvenile offenders and victims: 1999 national report*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Steinberg, L., & Cauffman, E. (1996). Maturity of judgment in adolescence: Psychosocial factors in adolescent decision making. *Law and Human Behavior*, *20*, 249-272.

Thissen, D., Steinberg, L., Pyszczynski, T., & Greenberg, J. (1983). An item response theory for personality and attitude scales: Item analysis using restricted factor analysis. *Applied Psychological Measurement*, *7*, 211-226.

Ustad, K. L., Rogers, R., Sewell, K. W., & Guarnaccia, C. A. (1996). Restoration of competency to stand trial: Assessment with the GCCT-MSH and the CST. *Law and Human Behavior*, *20*, 131-146.

van Schuur, W. H., & Kiers, H. A. (1994). Why factor analysis often is the incorrect model for analyzing bipolar concepts, and what model to use instead. *Applied Psychological Measurement*, *18*, 97-110.

Viljoen, J. L., Roesch, R., & Zapf, P. A. (2002). Interrater reliability of the Fitness Interview Test across four professional groups. *Canadian Journal of Psychiatry*, *47*, 945-952.

Yu, C. Y., & Muthén, B. (2001). *Evaluation of model FIT-R indicates for latent variable models with categorical and continuous outcomes*. Unpublished manuscript.

Zapf, P. A., & Roesch, R. (1997). Assessing fitness to stand trial: A comparison of institution-based evaluations and a brief screening interview. *Canadian Journal of Community Mental Health*, *16*, 53-66.

Zapf, P. A., & Roesch, R. (2001). A comparison of the MacCAT-CA and the FIT-R for making determinations of competency to stand trial. *International Journal of Law and Psychiatry*, *24*, 81-92.

Zapf, P. A., Roesch, R., & Viljoen, J. L. (2001). The utility of the Fitness Interview Test for assessing fitness to stand trial. *Canadian Journal of Psychiatry*, *46*, 426-432.

Zapf, P. A., Skeem, J. L., & Golding, S. L. (2005). Factor structure and validity of the MacArthur Competence Assessment Tool–Criminal Adjudication. *Psychological Assessment, 17*, 433-445.