Factor Structure and Validity of the MacArthur Competence Assessment Tool—Criminal Adjudication

Patricia A. Zapf
The City University of New York
Jennifer L. Skeem
University of California, Irvine
Stephen L. Golding
University of Utah

Examination of the available literature regarding the development of the MacArthur Competence Assessment Tool—Criminal Adjudication (MacCAT–CA) reveals 2 theoretical factor structures on which the MacCAT–CA was based: one in which 3 lower-order constructs are proposed (understanding, reasoning, appreciation) and one in which 2 higher-order constructs are proposed (competence to assist counsel and decisional competence). Confirmatory factor analyses were conducted with the MacCAT–CA’s original normative sample (N = 729) to test both the relative fit of these 2 theoretical factor structures and models that combine the 2 factor structures. Analyses were also completed to examine the convergent and discriminant validity of the MacCAT–CA. Results are discussed in terms of the strengths and weaknesses of the nomothetic nature of the MacCAT–CA.

Keywords: MacCAT–CA, competency to stand trial, forensic assessment, factor structure, validity

Aspects of the Competency Construct That Inform Instrument Evaluation

Three key aspects of the adjudicative competency construct are relevant to the development and evaluation of competency assessment instruments. First, competency is a different construct than either intelligence or mental disorder. The presence of cognitive disability or mental disorder is merely a threshold issue that must be established to “get one’s foot in the incompetency door” (Bonnie, 1992; Golding & Roesch, 1988; Grisso, Appelbaum, Mulvey, & Fletcher, 1995; Skeem, Golding, Cohn, & Berge, 1998). Symptoms are relevant only when linked with impairment in the abilities necessary to understand legal proceedings and participate in one’s defense. Based on this principle, scores on competency assessment measures should be associated with measures of intelligence and psychopathology only to the extent that the latter themselves are nomologically related to the competency construct. Because subgroups of individuals with different levels of intelligence or psychopathology will produce different average levels of psycholegal abilities, there will be some valid covariation across these subgroups (Meehl, 1995), as depicted in Areas 4 and 5 of Figure 1. An ideal measure of competency, however, would perfectly overlap the variability in the psycholegal abilities associated with adjudicated competent or incompetent individuals (i.e., Areas 2 and 3 of Figure 1 would be zero, Area 6 would be isomorphic with Area 7, and Area 4 would be isomorphic with Area 5). In the language of test construction, “good” items on competency assessment instruments (Campbell, 1960; Krause, 1972) covary with intelligence or mental disorder only to the extent that such constructs covary with competency. In the language of Baron and Kenny (1986), there should exist an indirect path between measures of competency and IQ/psychopathology, mediated by the competency status of indi-
individuals, but there should be no direct path between a measure of competency and measures of IQ/psychopathology.

The second relevant aspect of the competency construct is that no fixed set of psychosocial abilities can define it (Drope v. Missouri, 1975; Poythress et al., 1999; Roesch & Golding, 1980). The abilities required to competently stand trial vary, depending on the functional context of the case (e.g., the defendant, the charges, the evidence, available trial strategies; see Golding, 1993; Skeem, Golding, & Emke-Francis, 2004). The open-textured nature of the competency construct raises questions about highly structured competency assessment instruments (e.g., Lipsitt et al., 1971; Wildman et al., 1978). Some competency assessment instruments (e.g., Golding, 1993; Roesch et al., 1998; Rogers et al., 2004) adopt a structured clinical judgment approach, directing evaluators’ attention to the range of psychosocial competency abilities (e.g., appreciation of charges, understanding and reasoned choice of legal options) established by case law and the case context. Though structured, these instruments prescribe case-specific inquiries with defendants, based on their charges and likely legal strategies, and yield idiographic scores based on subjective ratings.

One of the most recently developed instruments, the MacArthur Competence Assessment Tool-Criminal Adjudication (MacCAT–CA; Hoge, Bonnie, Poythress, & Monahan, 1999; Poythress et al., 1999) adopts a newer, standardized approach to assessment in which evaluators ask uniform questions of each defendant that largely reference a hypothetical case vignette. This instrument yields nomothetic, criterion-based scores that may be compared with normative data. The MacCAT–CA developers, however, emphasize that the instrument was designed for use as a tool rather than a test of competence. Scores are meant to be interpreted within the context of the defendant’s case and integrated with more case-specific inquiries.

The third key principle is that standards of competence are one area of inquiry, whereas competence conceptualizations that underlie assessment tools are another. There is no constitutional requirement for a standard of competency to differ across such different adjudicative competency contexts (e.g., competency to stand trial vs. competency to plead guilty; Godinez v. Moran, 1993; see also Cooper v. Oklahoma, 1996). Although the MacCAT–CA, for example, separates competence abilities into two (competence to assist counsel [CAC] and decisional competence [DC]) or three (understanding, reasoning, and appreciation) domains, this is not inconsistent with current competency standards. An important corollary of this principle is that empirical support for the MacCAT–CA’s two- or three-domain conceptualization of competency bears strongly on the instrument’s validity but cannot impart the “truth” about the construct that underlies competency standards (Roesch, Hart, & Zapf, 1996). Tests that operationalize specific conceptualizations of competency assume a particular pattern of relations among items across individuals and are not immune to examination based on the extent of empirical support for their hypothesized factor structure. However, there are clear limits to what that factor structure imparts about the true nature of the competency standard.

The MacCAT–CA

In this paper, these three nomological aspects of the competency construct are considered in light of the results of an analysis of the factor structure and validity of the MacCAT–CA. The MacCAT–CA is a second-generation instrument that was developed to address a number of the limitations of other, first-generation instruments. It was, in part, a response to Grisso’s (1992) call for an “instrument offering standardized administration and scoring (as contrasted with CST interview guides and subjective ratings) to assess the domain of CST-related abilities for the general population of defendants who are referred for CST evaluations” (pp. 366–367). Unlike past measures, the MacCAT–CA includes standardized administration, criterion-based scoring, and a systematic distinction between defendants’ existing legal knowledge and capacity to attain such knowledge. Moreover, the MacCAT–CA was developed on the basis of a strong theoretical foundation (Hoge, Bonnie, et al., 1997).

The MacCAT–CA is based on a comprehensive research instrument (Hoge, Poythress, et al., 1997) that was developed, pilot tested, and refined prior to administration in a large field study (Bonnie et al., 1997; Hoge, Bonnie, et al., 1997; Hoge, Poythress, et al., 1997; Otto et al., 1998). This field study yielded the MacCAT–CA, which consists of 22 items grouped into three subscales of psychosocial abilities: Understanding, Reasoning, and Appreciation. An evaluator begins the MacCAT–CA by reading a hypothetical vignette to the defendant. The first subscale (eight items) assesses the defendant’s ability to understand information about the legal system and the process. For each item, the defendant is asked a question grounded in the vignette (e.g., “What is the job of the attorney for the defense?”) and is awarded 2 points (items are rated 0, 1, 2) if he or she demonstrates full understanding. If the defendant earns less than 2 points, the examiner discloses the answer and asks the defendant to repeat the disclosure in his or her own words, enabling separate assessment of the defendant’s capacity to understand and his or her preexisting understanding.

The second section (eight items) assesses the defendant’s ability to reason. The first five items in this section assess the defendant’s ability to choose from among two options the most legally relevant piece of information that the hypothetical defendant in the vignette should disclose to his lawyer. The last three items require the defendant to evaluate mock legal options for the hypothetical defendant.
The third section (six items) assesses the defendant’s ability to appreciate his or her own legal situation. This section departs from the hypothetical vignette to explore, in a standardized manner, the defendant’s appraisal of how he or she is likely to function and to be treated during the course of adjudication, relative to other defendants (e.g., whether he or she is likely to be treated more, less, or about as fairly as other defendants). These items are scored on the basis of the plausibility (vs. delusionality) of the grounds for the defendant’s judgment.

The psychometric properties of the MacCAT–CA were examined based on a sample of 729 felony defendants in eight different states (Otto et al., 1998; see also Rogers, Grandjean, Tillbrook, Vitacco, & Sewell, 2001). The results indicated that the MacCAT–CA demonstrated good reliability. For each of the three sections, internal consistency ranged from .81 to .88 (α = .81 for Reasoning, .85 for Understanding, and .88 for Appreciation), and interrater reliability ranged from very good to excellent (intraclass R = .75 for Appreciation, .85 for Reasoning, and .90 for Understanding). Otto and colleagues (1998) report that support for the construct validity of the MacCAT–CA was “found in the pattern of correlations between the MacCAT–CA measures and select clinical variables” (p. 439). MacCAT–CA Understanding, Reasoning, and Appreciation scores correlated .41, .34, and .14, respectively, with estimated Wechsler Adult Intelligence Scale—Revised (WAIS–R; Wechsler, 1981) full scale IQ, and –.23, –.29, and –.36, respectively, with Brief Psychiatric Rating Scale (BPRS: Overall & Gorham, 1962) total scores. Whether such relationships with measures of intelligence and psychopathology are theoretically desirable characteristics for a measure of competency is discussed later. The three MacCAT–CA scales correlated moderately with clinicians’ global ratings of competency (r = .36, .42, and .49, respectively) for a subset of defendants who had been hospitalized as incompetent.

**Theoretical Underpinnings of the MacCAT–CA**

**Three factors.** The MacCAT–CA’s three-scale structure is partially rooted in the MacArthur group’s review of the literature on competency to consent to treatment. Despite differences in terminology, four commonly accepted “prongs” of the general construct of competence are apparent (see Appelbaum & Grisso, 1988, 1995; Appelbaum, Lidz, & Meisel, 1987; Appelbaum & Roth, 1982; Roth, Meisel, & Lidz, 1977), including the ability to (a) express a choice, (b) understand relevant information, (c) appreciate the personal importance of the situation, and (d) rationally manipulate information. The MacCAT–CA assesses the three latter prongs of competence—understanding, appreciation, and reasoning.

**Two factors.** Bonnie (1992, 1993) expanded on Winick’s (1987) work to draw parallels between standards of treatment competency and criminal competency. Specifically, he conceptualized adjudicative competence in terms of a foundational construct (CAC) and a related contextualized construct (DC). The foundational construct of CAC refers to “the minimum conditions required for participating in one’s own defense” (p. 297) to preserve the dignity and reliability of the criminal process. It includes the individual’s:

(i) capacity to understand the charges, the purpose of the criminal process and the adversary system, especially the role of the defense counsel; (ii) capacity to appreciate one’s situation as a defendant in the criminal prosecution; and (iii) ability to recognize and relate pertinent information to counsel concerning the facts of the case (Winick, p. 243–287).

In contrast with this foundational construct, the contextualized construct of DC encompasses the defendant’s ability to make decisions and behave appropriately at a trial. According to Bonnie (1992), decisional competence is separable from competence to assist counsel: “although the relevant psychological capacities may overlap, decision-making about defense strategy encompasses conceptual abilities, cognitive skills, and capacities for rational thinking that are not required for assisting counsel” (p. 305). Tests of DC should be contextualized or tailored to the specific situation and the decision that is required. Bonnie’s reformulation of the concept of adjudicative competency was in large part the foundation for the MacCAT–CA. Each section of the MacCAT–CA contains items that tap both CAC and DC.

**Competing or complementary models?** Given the development and theoretical underpinnings of the MacCAT–CA, there appear to be two distinct, perhaps competing, theoretical factor structures for the instrument: the three-factor structure (composed of understanding, appreciation, and reasoning) that corresponds to the MacCAT–CA designed scales, and Bonnie’s two-factor structure (comprising CAC and DC). Although these two models were explicitly viewed as complementary in an early version of the instrument, the relation between them became blurred with the creation of the final MacCAT–CA. The forerunner of the MacCAT–CA blended the three-factor model with Bonnie’s two-factor model. This crossing of three functional ability domains (understanding, reasoning, and appreciation) with two legal domains or levels of competency (CAC, DC) produced six scales (Hoge, Bonnie et al., 1997). Notably, however, an exploratory factor analysis suggested this forerunner instrument was unidimensional (see Cruise & Rogers, 1998). In contrast, the final MacCAT–CA was simply offered as a three-factor instrument. A close examination of the instrument, however, reveals vestiges of the two-factor model. In some way not described in published materials, the MacCAT–CA’s authors appear to have sampled both CAC and DC items from the three functional domains in a manner that conflates the three-factor model with Bonnie’s two-factor model. The authors note only that “the . . . MacCAT–CA will employ items derived from the [research instrument] and [will] be much shorter and user-friendly” (Hoge, Bonnie et al., 1997; see also Otto et al., 1998).

In summary, during the development of the MacCAT–CA, the authors strongly emphasized the importance of producing a measure with a clear conceptual structure (e.g., Hoge, Bonnie et al., 1997; Otto et al., 1998). Although the research instrument explicitly crossed three functional abilities and two legal domains, there was a conceptual leap to the MacCAT–CA, which is offered as tapping three functional abilities (without the relation of these abilities to the two legal domains). The MacCAT–CA may have a three-factor structure, a two-factor structure, or a structure that crosses both models. One such possible “crossed” structure involves construing understanding, appreciation, and reasoning as types of abilities that may each be exercised within two content domains—competency to assist counsel and decision competence (see Bonnie & Grisso, 2000).
Investigating the MacCAT–CA’s Factor Structure

Rogers and his colleagues (2001) completed exploratory factor analyses of the MacCAT–CA, based on 149 mentally disordered offenders, and found a two- or three-factor solution. The three-factor solution grossly conformed to the designed three-factor model (understanding, reasoning, and appreciation), whereas the two-factor solution combined the instrument’s understanding and reasoning scales. The authors raised the possibility that the latter finding was based on method variance, given that the understanding and reasoning items are based on the hypothetical case vignette, whereas the appreciation scale is more case specific. Although this study is a welcome investigation of the factor structure of the MacCAT–CA, it is based on two small samples of convenience and does not apply confirmatory factor analytic techniques, which are most appropriate given the MacCAT–CA’s theoretical basis.

The chief purpose of this article is to examine the relative fit of two theoretical factor structures (separately and in combination) to the MacCAT–CA normative data. Our secondary aim is to assess aspects of the MacCAT–CA’s construct validity by estimating the extent to which it taps aspects of severe mental illness and disability that do not overlap with competence deficits per se.

Method

The dataset used in this study was obtained from the researchers who completed the original MacCAT–CA validation and normative study. The method of the original MacCAT–CA validation study is described in detail elsewhere (Otto et al., 1998; Poythress et al., 1999). Briefly, the MacCAT–CA was normed on a sample of 729 felony defendants in eight different states. Three subgroups of defendants were recruited, including Jail Unscreened (JU), Jail Treated (JT), and Hospitalized Incompetent (HI) groups. The JJ subgroup consisted of 197 randomly selected jail inmates who were presumed to be competent to proceed (in that the issue of adjudicative competency had not been raised in their cases). The Jail Treated (JT) subgroup consisted of 249 jail inmates presumed to be competent but who were receiving mental health services for reasons unrelated to competency. These two groups of jail inmate participants (JU and JT) were recruited from lists of defendants provided by the public defenders’ offices in each of the eight states. The refusal rate was 9% for each of the two jail inmate subgroups. The HI subgroup consisted of a total of 283 defendants who had been admitted to forensic psychiatric units for competency restoration after having been adjudicated incompetent to proceed. Potential HI participants were identified as they entered the facility for restoration and were recruited within 14 days of admission. The refusal rate was 22% for the HI subgroup.

The HI group was significantly older than the participants in the JU and JT samples and Caucasian and non-Caucasian participants were “some-what disproportionately distributed across the three samples” (p. 437). In addition, the HI group obtained higher scores on the Minnesota Multiphasic Personality Inventory—2 Psychoticism scale and on the BPRS. The three groups were similar in gender, education, socioeconomic status, amount and type of involvement with the criminal justice system, and estimated current cognitive-intellectual functioning.

The study procedure consisted of a one-time interview in which defendants completed a variety of measures that included the MacCAT–CA, the BPRS (Overall & Gorham, 1962), and the Information and Picture Completion subtests of the WAIS-R (Wechsler, 1981). These subtests were used to estimate full scale WAIS-R intelligence scores.

The MacCAT–CA consists of 22 items with each item being scored as 0, 1, or 2. The scores for the subscales, therefore, range from 0–16 for understanding, from 0–16 for reasoning, and from 0–12 for appreciation.

Results

The aims of this study were addressed in three stages. First, confirmatory factor analyses (CFAs, [maximum likelihood estimation1]) performed in LISREL 8.30 were completed to test the relative fit of the theoretical three-factor model, two-factor model, and combined two- and three-factor models. Second, modification indices from the CFAs were used to determine the best fitting, theoretically logical model for the data. Third, supplementary scale analyses were performed to examine the construct validity of the MacCAT–CA, including the nature and degree of association between MacCAT–CA scale scores and measures of intelligence and psychopathology. The results of these three stages of analyses are presented below. In each section, the results of the analyses for the combined sample are presented first, followed by the results for the JU, the JT, and HI subgroups.

Confirmatory Factor Analyses (A Priori Models)

First, we tested the two-factor, three-factor, and combined two- and three-factor models using CFA. For these analyses, we report several relevant indices of goodness of fit for each model because each index addresses a slightly different issue, and good-fitting models will produce consistent results across many indices (Ullman, 2001). Specifically, for each model, we report (a) absolute fit indices (χ², χ²/df, Goodness of Fit Index; GFI), (b) relative fit indices (Comparative Fit Index [CFI], Normed Fit Index [NFI]), (c) parsimoniousness fit indices (Akaike Information Criterion [AIC]; Consistent Akaike Information Criterion [CAIC]), and (d) non-centrality-based indices (root-mean-square error of approximation; RMSEA—for overviews, see Bentler, 1995; Church & Burke, 1993; Kline, 1998; MacCallum & Austin, 2000; Ullman, 2001). Because the CFI and NFI indices are not easy to interpret we note “models with overall fit indices of less than .90 can usually be improved substantially” (Bentler & Bonett, 1980, p. 600). In other words, minimum values of .90 for CFI and NFI statistics are conventionally required to indicate even “adequate,” “satisfactory,” or “acceptable” fit (Byrne, 1994; Kline, 1998; Ullman, 2001). In fact, Hu and Bentler (1999) recently analyzed the ability of various statistics to distinguish between true population and misspecified models at different thresholds and recommended a higher cutoff of .95 for the CFI. To ensure consistency in this article, we used the following heuristic labels to describe fit: inadequate when most values of the GFI, CFI, and NFI statistics are conventionally required to indicate even “adequate,” “satisfactory,” or “acceptable” fit (Byrne, 1994; Kline, 1998; Ullman, 2001).

Three-factor model. First, we tested the fit of the three-factor model, given that this model is explicitly set forth for the MacCAT–CA in publications (Otto et al., 1998; Poythress et al., 1999). We tested a correlated three-factor model composed of

---

1 Given the ordinal nature of the items, we attempted to determine whether maximum likelihood estimation would provide estimates that would differ significantly from more robust methods. We ran the bootstrapping technique available in Amos 5.0 (Assessment Systems; St. Paul, MN) on these data, with 1,000 samples, in an attempt to correct for non-normally distributed data. With this procedure, we found the exact same results as reported using maximum likelihood estimation.
three latent variables that directly corresponded to the three sections of the MacCAT–CA. The first latent variable (understanding) consisted of Items 1–8, the second (reasoning) consisted of Items 9–16, and the third (appreciation) consisted of Items 17–22. This three-factor model appeared to be a reasonably good fit of the data for the full sample (see Figure 2; Table 1). When this same model was tested using each of the three subsamples, its fit was found to be inadequate for the JU and JT subgroups but adequate for the HI subgroup.

An examination of the correlations between the three factors reveals strong relationships between understanding and reasoning \((r = .92)\), reasoning and appreciation \((r = .73)\), and understanding and appreciation \((r = .62)\). This suggests the presence of a strong second-order factor that runs through the entire instrument. Given that respecifying the model in terms of three first-order factors that load on a single second-order factor would result in identical fit indices, we instead tested for the presence of a single second-order factor by testing the difference between a correlated-factors model and an orthogonal-factors model \((Gorsuch, 1983)\). Specifically, we specified the three first-order factors in two models—one in which the correlations between the first-order factors are left free to vary and one in which the correlations are set to zero. The correlated factors model fit significantly better than the orthogonal factors model, suggesting the existence of a higher order structure, \(\chi^2\) (3) = 876.94, \(p < .0001\) (see Table 1). In addition, we also tested three variations of a correlated-factors model (wherein we sequentially set to 1 the correlation between each of the three possible pairs of first-order factors) to test for the independent presence of each of the three first-order factors. The results suggest that each first-order factor is necessary to avoid a decrement in fit, difference in \(\chi^2\) (1) = 1449.75, 57.29, and 606.27, all \(p s < .001\). Taken together, these results suggest the presence of three distinct first-order factors and an overarching second-order factor.

**Bonnie’s two-factor model.** The MacCAT–CA’s primary author provided us with a description of the MacCAT–CA items that relate to Bonnie’s DC and CAC factors (Norman Poythress, personal communication, December 5, 2001). This two-factor model (see Figure 3) was composed of two latent variables. The first factors model fit significantly better than the orthogonal factors model, suggesting the existence of a higher order structure, \(\chi^2\) (3) = 876.94, \(p < .0001\) (see Table 1). In addition, we also tested three variations of a correlated-factors model (wherein we sequentially set to 1 the correlation between each of the three possible pairs of first-order factors) to test for the independent presence of each of the three first-order factors. The results suggest that each first-order factor is necessary to avoid a decrement in fit, difference in \(\chi^2\) (1) = 1449.75, 57.29, and 606.27, all \(p s < .001\). Taken together, these results suggest the presence of three distinct first-order factors and an overarching second-order factor.

**Bonnie’s two-factor model.** The MacCAT–CA’s primary author provided us with a description of the MacCAT–CA items that relate to Bonnie’s DC and CAC factors (Norman Poythress, personal communication, December 5, 2001). This two-factor model (see Figure 3) was composed of two latent variables. The first

---

2 Although comparison of an oblique solution with an orthogonal one does not technically prove the existence of a single higher order factor, we will use the terms first-order and second-order factors as a heuristic convenience in the remainder of this article.
variable (DC) was composed of six items—two (Items 7 and 8) from the Understanding section, three (Items 14, 15, and 16) from the Reasoning section, and one (Item 22) from the Appreciation section. The second factor (CAC) consisted of all 16 remaining items on the MacCAT-CA. This two-factor model inadequately fit the data for the full sample (both when the correlation between the factors was left free to vary as well as when it was set to zero) as well as data for each of the three subgroups (see Figure 3 and Table 1).

**Combined two- and three-factor models.** Given that the fore-runner to the MacCAT-CA explicitly crossed two- and three-factor models, we also wished to simultaneously test the fit of these two models. Thus, we did this in three different ways. First, we partially crossed the three functional abilities (understanding, reasoning, appreciation; see Figure 2) with the legal domains (CAC, DC; see Figure 3) to simultaneously test the fit of these models (in a correlated five-factor model). That is, we allowed each of the 22 items to cross-load on two factors—one of the three functional abilities (understanding, reasoning, appreciation, or reasoning) and one of the two legal domains (CAC, DC). This model is consistent with construing understanding, appreciation, and reasoning as types of abilities that may each be exercised within the content domains of competency to assist counsel and decision competence (see Bonnie & Grisso, 2000).

Second, we allowed each of the 22 items to load onto one of six first-order factors (understanding CAC, understanding DC, appreciation CAC, appreciation DC, reasoning CAC, reasoning DC) and then each of these first-order factors was allowed to load on one of the three second-order functional ability factors (understanding, appreciation, or reasoning) and one of the two second-order legal domains (CAC, DC). Finally, given that item-level analyses can sometimes be problematic, we formed six scales from the 22 items as per the direction of the developers of the MacCAT-CA (UCAC = Items 1–6, UDC = Items 7–8, RCAC = Items 9–13, RDC = Items 14–16, ACAC = Items 17–21, ADC = Item 22) and then allowed each of these six scales to load on one of the three functional abilities and one of the two legal domains. None of these three different models fit the data. That is, there was either a problem with the identification of the model and thus it did not converge, or if the number of iterations was increased to beyond what is considered appropriate so as to allow for convergence (see Jöreskog & Sörbom, 1993), the solutions did not make sense as evidenced by loadings greater than 1.

**Comparing the relative fit of the theoretical models.** Given these problems with the combined two- and three-factor model specifications, we were unable to compare them with the individual two- and three-factor models. The difference in fit between the three-factor model and Bonnie’s two-factor model suggested that the three-factor model fit the data better for both the full sample (an adequate vs. inadequate fit) and the three subgroups. Unfortunately, because these models were not nested, we could not statistically compare their relative fit.

Notably, although the three-factor model adequately fit the data for the HI subgroup, even this model inadequately fitted the JU and JT subgroups. This difference does not appear attributable to sample size differences, given that each subgroup consisted of approximately 200 participants (HI = 283, JU = 197, JT = 249), which is sufficient for small to medium models (Ullman, 2001). This difference in fit may, however, be attributable to differences in the distributions of MacCAT-CA scores across subgroups. Specifically, there was more variability in MacCAT-CA scores for the HI subgroup (i.e., the “incompetent” or “more severely ill” group) than for the JT and JU subgroups (i.e., the “presumed competent” samples). MacCAT-CA scores approximated a normal distribution in the HI subgroup, but were substantially more negatively skewed and kurtotic in the JT and JU subgroups. The HI subgroup’s closer approximation to a normal distribution than the JT and JU subgroups mainly reflects the fact that a greater number of individuals in the former group scored at the lower end of the distribution, although the overall range of scores for each of the three groups did not appear to be glaringly restricted.
Confirmatory Factor Analyses & Schmid–Leiman Transformation (Modified Models)

In summary, the first stage of analysis suggested that a hierarchical three-factor model fit the data better than the two-factor model. To determine whether a particularly good fit could be obtained with alternative models, the second stage of analysis involved examining modification indices from the three-factor CFA in an attempt to develop a better-fitting model. The models were developed and tested in a stepwise manner.

**Collapsed two-factor model.** The results of the correlated three-factor CFA suggested that the understanding and reasoning factors were very strongly correlated ($r = .92$). Because these two factors appeared virtually redundant, we combined them into a single factor (i.e., fixed the correlation between factors to 1) and tested the fit of this new two-factor model. Although the results (see Table 1) indicated that this model (understanding/reasoning and appreciation) was a *reasonably good* fit to the data, a chi-square comparison of this model’s fit with that of the original three-factor model suggested that combining the understanding and reasoning factors significantly reduced the model’s fit, $\Delta \chi^2(2) = 57.29, p < .001$.

**Modified three-factor model.** An examination of the items’ factor loadings suggested that Items 9–13 tended to load by themselves and apart from the other items on the understanding/reasoning factor. Thus, we constructed a modified three-factor model: the first factor included Items 1–8 and 14–16 (i.e., the original Understanding subscale plus Items 14–16), the second factor included Items 9–13, and the third factor included Items 17–22 (i.e., the original Appreciation subscale). In this model, we allowed the error terms for Items 2 and 3 and Items 15 and 16 to correlate with each other, given that these two item pairs were highly similar in nature and format to one other. Specifically, Items 2 and 3 ask about charges (i.e., aggravated and simple assault, respectively), and Items 15 and 16 ask about advantages and disadvantages for pleading one way or another. Given our earlier finding that there did appear to be evidence to suggest a second-order factor (in addition to the three first-order factors), we included this second-order factor in the representation of this model (see Figure 4 and Table 1). This modified three-factor model provided the best overall fit to the data.

This modified three-factor model differs from the theoretical three-factor model in two respects. First, the model collapses items...
from the Understanding section and Items 14–16 from the Reasoning section (i.e., the items that tap decisional competence) into one factor, places Items 9–13 from the Reasoning section into a second factor, and keeps all items from the Appreciation section as the third factor. Second, the model allows error terms for two item pairs to correlate. To determine the extent to which the better fit of the modified than the theoretical three-factor model was attributable to substantive item changes or to permitting error terms to correlate, we conducted a CFA of the theoretical three factor model in which we permitted error terms for Items 2 and 3 and Items 15 and 16 to correlate. The results (see Table 1; three-factor model with error terms correlated) suggest that the fit of the theoretical three-factor model (e.g., CFI = .96) manifests modest improvement when error terms are permitted to correlate (e.g., CFI = .97) and additional modest improvement if substantive item changes are made (e.g., CFI = .98). In short, both aspects of the new model contribute modestly to improved fit.

Schmid-Leiman transformation. Given that a hierarchical three-factor structure appeared to best fit the data, an exploratory higher order factor analysis was performed using the Schmid and Leiman (1957) orthogonalization procedure to more closely explore the measure’s hierarchical factor structure. The results provide additional evidence suggestive of a strong second-order factor (accounting for 72.3% of the variance) and three first-order factors, as well as the best fitting modified three-factor model. With respect to the latter point, Items 14–16 loaded on the Understanding subscale of the MacCAT–CA in this analysis (factor loadings are provided in Table 2).

Association of the MacCAT–CA With Measures of Intelligence and Psychopathology

In summary, the second stage of analysis involved developing a modified three-factor model and determining that it, like the theoretical three-factor model, is a good fit to the data. The results provide additional evidence suggestive of a strong second-order factor (accounting for 72.3% of the variance) and three first-order factors, as well as the best fitting modified three-factor model. With respect to the latter point, Items 14–16 loaded on the Understanding subscale of the MacCAT–CA in this analysis (factor loadings are provided in Table 2).
oretical three factor model, provided a reasonably good fit to the data. In a third stage of analysis, we performed supplementary analyses to determine the degree of association between scores on scales constructed from the theoretical three-factor model and measures of intelligence and psychopathology. As explained in the introduction, the degree of association between the MacCAT–CA and measures of IQ and psychopathology are high enough to justify further investigation of the scale’s discriminant validity.

**Basic relation to intelligence and psychopathology.** To estimate the basic relations between the MacCAT–CA factors on one hand and intelligence and psychopathology on the other, two regression analyses were performed. First, all three MacCAT–CA subscales were entered in a single step as predictors of WAIS IQ scores. The squared semipartial correlation for each subscale was used to estimate the variance uniquely accounted for by each first-order factor (understanding, reasoning, and appreciation), and the difference between the sum of these correlations and the increment in variance accounted for on that step of the regression was used to estimate the variance uniquely accounted for by the second-order factor.\(^5\) Second, this process was repeated, using BPRS total scores as the criterion. The results, shown as semipartial correlations in Table 3, suggest a moderate relation between the MacCAT–CA and intelligence and psychopathology.

**Relation to competence-relevant intelligence and psychopathology.** Examining correlations between the MacCAT–CA and measures of psychopathology (BPRS total scores) and intelligence (WAIS IQ estimate) estimates Areas 2 and 5 of Figure 1 and thus confounds covariation that reflects valid covariance (Area 5) with covariation that is undesirable (Area 2). For this reason, we applied Baron and Kenny’s (1986) framework to examine the size of Area 2. Again, there should be an indirect path between the MacCAT–CA and intelligence/psychopathology, mediated by the competency status of individuals. There should, however, be no direct path between the MacCAT–CA and measures of intelligence/psychopathology. Area 2 represents the direct path between the MacCAT–CA and measures of IQ/psychopathology, whereas Area 5 represents the indirect path between these measures mediated by competency status. We used group status\(^6\) (HI vs. JT/JU) as an indicator of differences in psychological abilities related to competency status.

The mediation strategy was applied as a method for estimating the direct and indirect relation between the MacCAT–CA and intelligence or psychopathology rather than as a test of a causal model. A series of six regression analyses were performed (see Baron & Kenny, 1986) to yield the path coefficients \((a, b, c, and c')\) necessary to test MacCAT–CA (total scores) mediation models for intelligence (as indexed by WAIS IQ estimates) and psychopathology (as indexed by BPRS total scores). Where the MacCAT–CA is X, competency is M, and intelligence or psychopathology is Y, the traditional paths are \(a\) (relation between X and M), \(b\) (relation between M and Y, controlling for X), \(c\) (direct relation between X and Y), and \(c'\) (relation between X and Y, controlling for M). These coefficients were checked against the results of a model assessed with Amos 5.0.

**Intelligence (WAIS IQ).** MacCAT–CA total scores and scores of intelligence were moderately related (path \(c = .35\)). This relationship was weakly, but significantly, mediated by competence status \((c-c' = .08, p < .001; Sobel, 1982). As shown in Figure 5, the mediated effect was small (path \(b = -.15\), particularly relative to the direct relation between the MacCAT–CA and intelligence (path \(c' = .43\)). Once competency status is taken into

---

\(^5\) We remind the reader here that we are using the term second-order factor as a heuristic convenience, as we have not technically proven the existence of a single, second-order factor but the existence of a higher order factor structure has been suggested by the results of our analyses.

\(^6\) Competence status (HI v. JT/JU) is an imperfect measure of true competence. It was, however, the best measure available for this study and represents judicial decisions made (vs. not raised) about these defendants.
account, there is a sizable independent relation between the MacCAT–CA and WAIS-IQ scores.

Psychopathology (BPRS total). MacCAT–CA total scores and scores of psychopathology were moderately and inversely related (path \( c = -0.24 \)). This relationship was weakly, but significantly, mediated by competence status (\( \text{path } c' = 0.09, p < 0.01; \text{Sobel, 1982} \)). As shown in Figure 6, the mediated effect (path \( b = 0.17 \)) was weaker than the direct relation between the MacCAT–CA and psychopathology (path \( c' = -0.24 \)). Again, once competency status is taken into account, there is a significant independent relation between the MacCAT–CA and psychopathology.

Discussion

This study was designed to examine the factor structure of the MacCAT–CA to determine which theoretical or empirically derived model best fits its normative data and to explore this measure’s construct validity by assessing the extent to which it taps severe mental illness and disability rather than competence deficits per se. Each of these purposes is discussed in turn below, followed by an analysis of adjudicative competence as a nomothetic construct.

Which Factor Structure Best Fits the Data?

With respect to the question regarding which of the theoretical factor structures best fit the MacCAT–CA normative data, it appears that the three-factor model better fits these data than Bonnie’s two-factor model. In addition, the results of both exploratory factor analyses (Schmid–Leiman transformation) and confirmatory factor analyses (comparison of correlated and orthogonal three-factor models) indicate the possibility of a second-order general factor running through the MacCAT–CA such that its structure is hierarchical.

Although the original three-factor model that reflects the three sections of the MacCAT–CA fits the data reasonably well, we found the best overall fit in a modified three-factor model in which all items from the Understanding section and Items 14–16 from the Reasoning section are collapsed into a first factor, Items 9–13 from the Reasoning section compose the second factor, and all items from the Appreciation section make up the third factor. The Schmid–Leiman results suggested a similar solution. Careful examination of the items composing each of the factors indicates that MacCAT–CA items that are similar in nature and format appear to cluster together, suggesting that the method of measurement is somewhat confounded with the facet of competency measured. Specifically, the first factor in the modified model was composed of items that addressed vignette-based understanding (including legal options); the second factor, case-specific items that addressed specific appreciation; and the third factor, vignette-based items that required identification of the more relevant of two pieces of information. Our finding that method variance may be confounded with construct variance is consistent with that of Rogers and colleagues’ (2001) finding that the MacCAT–CA items based on the hypothetical case vignette tended to load separately from those that involved more case-specific inquiries.

As discussed later, if the MacCAT–CA confounds construct variance with method variance, this may have important implications. However, the modifications made to arrive at the modified three-factor model (and the exploratory Schmid–Leiman results) risk capitalizing on chance fluctuations that are specific to the normative sample and may not be replicable. The original three-factor model, having been specified prior to examining the data, does not capitalize on sampling error. Future research with an independent sample is needed to test the modified model in an a priori manner to truly evaluate its fit.

Bonnie’s two-factor model did not adequately fit the data. This may be attributable partially to the way in which items for the MacCAT–CA were chosen from the research instrument. That is, although Bonnie’s model apparently provided strong theoretical guidance for developing items for the research instrument, this theoretical grounding may have become somewhat diluted as decisions were made about which items to keep as part of the more practical MacCAT–CA. The exact method for item selection has not been addressed in published materials. Although the principle of assessing decisional competence separate from foundational competence is fine in theory (and doesn’t necessarily contradict Godinez v. Moran, 1993), the MacCAT–CA does not appear to achieve this. Notably, the authors of the MacCAT–CA may not have intended to assess decisional competence separately given that they abandoned this idea somewhere between the forerunner to the MacCAT–CA and the actual MacCAT–CA. Nevertheless, it appears that there is no discernible distinction between CAC and DC in the MacCAT–CA (i.e., \( r = 1.00 \) between CAC and DC in Figure 3).
Distinguishing Competency From Other Constructs

The MacCAT–CA (scales/higher order factor) was moderately related to intelligence \((r = .42)\) and psychopathology \((r = -.36)\) as seen in Table 3. Mediation analyses of the relation between MacCAT–CA total scores and measures of IQ and psychopathology indicate a direct association between these measures independent of whether a defendant is competent. Although a defendant’s competence status partially mediates the association between these measures, the indirect effect is weak. Thus, the association between the MacCAT–CA and intelligence/psychopathology is not an artifact of the sampling procedure used in the normative study, as an adjudicated incompetent group (HI) was mixed with groups assumed competent (JT and JU). As noted by Meehl (1995), mixing such groups, when they differ on mean levels of the IQ and psychopathology, can create covariation between MacCAT–CA scores and measures of IQ/psychopathology that are ambiguous in their meaning. Here, by distinguishing direct and indirect paths, we were able to show that the direct paths between the MacCAT–CA and psychopathology and intelligence are nontrivial (see Area 2; Figure 1). This is particularly the case for intelligence, which was moderately \((\text{path } = .43)\) associated with MacCAT–CA scores, independent of whether a defendant was incompetent. The MacCAT–CA may test intellectual abilities (e.g., abstract reasoning) that are not essential to the functional capacities needed for competence.

Assuming that the moderate to strong covariation with intelligence and psychopathology is replicable, an issue of the construct validity of the MacCAT–CA potentially arises. Although Otto and his colleagues (Otto et al., 1998) may be correct in noting that such covariation supports the construct validity of the MacCAT–CA, we read the theoretical, empirical and jurisprudential literature in the opposite fashion. Throughout the past 3 decades of research and scholarship on the competency construct, a common theme has been the inappropriateness of confounding psychopathology or retardation with incompetency. The level of direct relation between the MacCAT–CA and measures of psychopathology and intelligence approach the range of the empirical correlations between the MacCAT–CA subscales and clinician global ratings of competency previously reported by Otto and colleagues (1998). We believe that this pattern of covariation has important implications for the discriminant validity of the MacCAT–CA. This is in part because clinical judgments have historically been quite strongly associated with psychopathology and intelligence (see Nicholson, Briggs, & Robertson, 1988; Nicholson & Kugler, 1991). We recognize that these data may reasonably be construed differently by others. In any event, an informed debate about the nomological aspects of the competency construct seems warranted.

Valuing Adjudicative Competence an Idiographic Construct

The results of this study raise a second issue about the extent to which norm-referenced or nomothetic assessment approaches appropriately capture an individual’s adjudicative competence. As suggested earlier, our findings and those of Rogers and his colleagues (2001) suggest that MacCAT–CA items that reference the hypothetical case vignette tend to load separately from those that address the defendant’s own case. Hypothetical inquiries may reveal different response patterns than case-specific ones. Although the MacCAT–CA represents a significant advance in standardized administration, criterion-based scoring, and normative data, it seems worth exploring in future research whether these advances come at the cost of diluting focus on the defendant’s particular case and competency. Less than one third of the MacCAT–CA items are case specific, and these items merely ask the defendant to make judgments about how equitably he will be treated in the criminal justice system, relative to other defendants. Although it is an open empirical question, it seems unlikely that this defendant’s rational and factual understanding of the proceedings against him and an ability to consult with his own attorney (Rogers et al., 2001) can be assessed fully by the MacCAT–CA’s six relativistic judgments about, for example, how fairly he will be treated compared with other defendants. Although the MacCAT–CA’s operational definition of competency abilities is nomothetic, the measure is intended to inform idiographic decisions about a particular defendant. The extent to which this nomothetic-idiographic translation is successful is unclear.

It may be that case-specific inquiries are needed to assess accurately some of the general psycholegal abilities encompassed by the MacCAT–CA. An open-ended conversation with a defendant about his or her case, recent experiences with the criminal justice system, and planned legal strategy may reveal irrational or delusional thought processes that are not apparent during brief, highly structured inquiries about relative judgments and their bases. A defendant may answer the appreciation section items well in an abstract sense but hold specific delusions about whether or not he will even be tried that become evident after the MacCAT–CA booklet is put away. The incremental utility of the MacCAT–CA to structured, case-specific inquiries (and vice versa) is an important issue to address in future research.

In a related sense, future work must establish how examiners should integrate MacCAT–CA normative scores with contextual issues involved in the defendant’s specific case. The MacCAT–CA’s authors explicitly acknowledge that competency cannot be reduced to an inflexible list of requisite abilities (Poythress et al., 1999) and recommend that the MacCAT–CA be supplemented with case-specific inquiries and consideration of the case context (Otto et al., 1998). It remains for future research, however, to provide examiners with clear guidance about how best to integrate the results of the MacCAT–CA with the rest of the competency assessment. In the absence of such guidance, decision makers may be apt to rely heavily on MacCAT–CA scores, which may be quite compelling when considered alongside less scientific-looking idiographic considerations. Just as the presence of mental illness is only a starting point in assessing competency, a high or low score on the MacCAT–CA does not resolve the issue.

Final Conclusions

Despite some of the structural weaknesses identified in this study, the MacCAT–CA appears to represent a psychometrically sound normative approach to measuring the generalized aspects of understanding, reasoning, and appreciation—as well as an overarching component—as applied to adjudicative competence (represented by the suggestion of a strong second-order factor, in addition to three distinct first-order factors). The MacCAT–CA is unique in its attempt to assess systematically reasoning and decisional capacities relevant to competency. As a normative tool, the
measure’s primary conceptual weakness concerns its inability to link normative scores to an individual, contextualized decision. Paradoxically, alternative devices, including structured interviews, being developed under a different conceptual approach have, as their strength, a direct focus on the linkage of an individual’s psychosocial abilities to a contextualized decision and have, as their weakness, an inability to normatively reference an individual’s psychosocial abilities. It would thus seem obvious that future research should focus on the comparative validity, reliability, and qualitative strengths and weaknesses of different assessment approaches (see generally, Hase & Goldberg, 1967; Knudson & Golding, 1974). To our knowledge, only one study has ever examined the comparative validity of comprehensive interview-based competency assessment approaches in the United States (Golding, Roesch, & Schreiber, 1984; Schreiber, Roesch, & Golding, 1987). A few others have examined the comparative validity of competency assessment instruments in Canada (Zapf & Roesch, 2001) or of noncomprehensive interview-based instruments such as the Georgia Court Competency Test and the Competency Screening Test (Nicholson, Briggs, & Robertson, 1988; Ustad, Rogers, Sewell, & Guarnaccia, 1996).

An additional important direction would involve comparing the different approaches in different contexts. There are both rational and empirical reasons to believe that different approaches to competency assessment, in routine practice, will often exhibit approximately equal validities (Golding, Roesch, & Schreiber, 1984). The primary reason for such approximate equivalency is that most jurisdictions have few controls over the granting of petitions evaluate defendants’ competency, and hence the referral process often results in upward of 70% of defendants recommended as competent (Roesch & Golding, 1987; Roesch, Zapf, Golding, & Szymon, 1996). In addition, many of the incompetent defendants are grossly psychotic and cognitively disorganized. Thus, in a typical referral stream, most assessment approaches may easily reach roughly equivalent results. A more strenuous testing ground would involve so-called “gray area” cases in which there is dispute, disagreement, and/or controversy surrounding the defendants’ competencies.

References


Wildman, R. W., Batchelor, E. S., Thompson, L., Nelson, F. R., Moore, I. T., Patterson, M. E., & deLaosa, M. (1978). *The Georgia Court Competency Test: An attempt to develop a rapid, quantitative measure of fitness for trial*. Unpublished manuscript, Forensic Services Division, Central State Hospital, Milledgeville, GA.
