NCJFCJ est. 1937 NATIONAL COUNCIL OF JUVENILE AND FAMILY COURT JUDGES WWW.NCJFCJ.ORG

A Roadmap for CANS Validation

By David R. Kraus

ABSTRACT

In nearly every jurisdiction, juvenile or family courts will be using the Child and Adolescent Needs and Strengths (CANS) tool to assess whether children can safely be treated in family-like settings to meet federal Family First Prevention Services Act (FFPSA) requirements. Yet, a number of peer-reviewed publications have raised serious concerns regarding the lack of CANS validation research. Rather than reject CANS, this article provides a roadmap for validating the tool with data that jurisdictions have already collected. Courts should require these three simple analyses before relying on the assessment.

Key words: validity, assessment, child welfare, Family First, CANS.

The Family First Prevention Services Act (FFPSA) requires juvenile or family courts to use an assessment tool to help determine whether children can be safely treated in family-like settings before considering expensive institutionalized treatments such as residential care.

The FFPSA was passed into law in February 2018. Among other reforms of Title IV-E, the law seeks to curtail the use of congregate or group care for children and instead places a new emphasis on family foster homes. With limited exceptions, the federal government will not reimburse states for children placed in group care settings for more than two weeks. Approved settings, known as qualified residential treatment programs, must use a trauma-informed treatment model and employ registered or licensed nursing staff and other licensed clinical staff. The child must be formally assessed within 30 days of placement to determine if his or her needs can be met by family members, in a family foster home, or another approved setting. According to the legislation, "Within 30 days of the start of each placement in such a setting, a qualified individual shall— (i) assess the strengths and needs of the child using an age-appropriate, evidence-based tool."

David R. Kraus, PhD., a Child Clinical Psychologist, is the President and Chief Scientific Officer of Outcome Referrals, Inc., and developer of the Treatment Outcome Package (TOP). Dr. Kraus has over thirty years of experience developing valid, royalty-free tools for adults, adolescents and children, and previously served as an Assistant Professor of Child Psychiatry at the University of Massachusetts Medical School. Outcome Referrals offers real-time processing and scoring of assessments, which in the past included CANS.

© 2020 National Council of Juvenile and Family Court Judges

Most child welfare directors have chosen to use one of the many versions of the Child and Adolescent Needs and Strengths scale (CANS) for this purpose,[1] despite the absence of peer-reviewed validation studies. This oversight should be corrected to prevent children from being misplaced in deep-end, expensive and sometimes harmful institutional settings.

Tool development and validation are guided by the American Psychological Association's ethics (2017) and tool development guidelines (American Educational Research Association, America Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.)., 2014). There is a legitimate place for expert-rated tools, such as CANS[2], to operate alongside consumer-completed tools (assessments completed by the family or child), giving courts the most complete picture of the child and his or her environment. Most domains that CANS attempts to measure are meaningful. In that spirit, this review is designed to help juvenile or family courts evaluate CANS with their jurisdiction's own data to ensure that the revised CANS is ready for its intended purpose.

An initial review of CANS found no published studies on its validity (Winters, Collett & Myers, 2005). During the intervening fifteen years, Lyons[3] and other CANS developers have failed to correct this deficit. A recent peer-reviewed debate with CANS developers has highlighted the ongoing lack of validity (Lyons & Israel, 2017; Kraus, 2017). Instead, CANS developers have engaged in alleged research misconduct that has included making at least 49 false citations to give the appearance of CANS validity. The list of these false citations can be found online and we have reviewed and confirmed each.[4]

The new owner of CANS (the University of Chicago's Chapin Hall[5]) may be working aggressively to correct these flaws. However, it will be a daunting challenge for the University because each state uses a different set of questions, each with a different scoring procedure. As per standard measurement development guidelines (American Educational Research Association, APA, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.), 2014), each unique "CANS" will need its own set of validation studies to demonstrate that these unique questions and scores have merit.

Nevertheless, there is a realistic path forward for CANS. Each jurisdiction's unique set of questions can be validated with the extensive CANS data that most states have

[1] The Praed Foundation's website page of states using CANS has been recently deleted. At one time the site allowed visitors to download the unique version of CANS modified for each state.

[2] CANS was developed to measure treatment outcomes (Lyons, Weiner & Lyons, 2004). Without a citation, Lyons claims that anyone with a BA degree can be trained to complete CANS reliably (praedfoundation.org).

[3] Key players in this article: John S. Lyons, Ph.D. is the principal developer of CANS. He set up a business controlled by him and his wife, called the Praed Foundation through which he sells his services that include CANS training. In or around 2017 he sold or donated the rights to CANS to Chapin Hall, a nonprofit affiliate of the University of Chicago, where Dr. Lyons was employed.

[4] See: www.kickthecans.net/whistleblower

[5] "Last year, Praed granted Chapin Hall (also a 501.c.3) with exclusive licensing rights." (Lyons & Israel, 2017, p. 3)

collected. Below are specific recommended steps for creating the necessary reliability, construct, and concurrent validity publications based on industry-standard guidelines.

RELIABILITY

Reliability assesses the consistency of a measure. There are three different ways of measuring reliability. Test-retest reliability assesses whether the measure is consistent over time. In the context of children seen by juvenile or family courts, we expect, or hope, they will change over time, so this is not the best subtype to evaluate. Internal consistency assesses whether the items related to a construct are consistently changing together. Measures of Construct validity (as reviewed in the validity section) can be maximized with factor analytic procedures (cf., Kraus, Boswell, Wright, Castonguay & Pincus, 2010).

Inter-rater reliability measures the extent to which two or more raters (or observers, coders, examiners) agree. A judge, for example, needs to know that when a clinician completes the CANS and rates the child with a specific severity score, that a different clinician would make the same rating. In other words, it is important to ensure that the rating is not just subjective. The judiciary needs to rely upon the information presented, regardless of who completes the tool.

CANS developers claim that inter-rater reliability is the most important for CANS (Anderson, Lyons, Giles, Price & Estes, 2003), and we concur. An ideal study of inter-rater reliability would take a group of recently trained CANS raters and ask them to read relevant case documentation and to interview about fifty families. Each family would be interviewed by two different raters, each blind to the other's ratings and each completing the *exact* version of CANS to be used in that state's jurisdiction. Inter-rater reliability is essentially measured by making a comparison of the agreement or disagreement between raters.

The CANS developers claim that completing CANS only takes fifteen minutes (Anderson et al., 2003), but an independent assessment of the amount of time comes from a Boston group which reported that the process takes about two hours (Kisiel, Blaustein, Fogler, Ellis & Saxe, 2009). The 200 hours of work to document the reliability of CANS is insignificant compared to the hundreds of millions of dollars spent by federal and state governments on child placements.

Using existing historical data, almost every jurisdiction can estimate the results these ideal studies would reveal. Almost every state has thousands of CANS administrations in a database[6]. Some of these CANS are completed by two different agencies treating the same child during the same thirty-day window. They are likely blind ratings (meaning each rater does not have access to the other rater's results before completing CANS) and these independent ratings can be compared for a solid reliability estimate. Courts should demand that these analyses be conducted and that the results be shared with the courts before proceeding.

[6] c.f., https://www.mass.gov/child-and-adolescent-needs-and-strengths-cans

VALIDITY

There are four types of validity that fall into two categories, according to the guidelines that govern the development of tools and measures (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.)., 2014).

The first category of validity – Content related validity – ensures that the tool has the appropriate content. Face validity is a non-psychometric way of assessing the content by looking at the items and personally agreeing that the items appear to assess the concepts in question. Construct validity psychometrically tests whether the questions are appropriately associated with the underlying theoretical concepts.

The second category of validity – Criterion related validity – assesses the relationship to other measures either at the same point in time (Concurrent) or at a time in the future (Predictive). A few years from now, it would be prudent to conduct a Predictive Validity study using the various tools used to comply with FFPSA and determine which best predicted the success of children placed in family-like settings or institutions. For now, this review will focus on Construct and Concurrent validity.

CONSTRUCT VALIDITY

Construct validity tells whether the tool in question actually measures the concepts it purports to assess (e.g., CANS "Risk Behaviors"). If a construct is labeled "Suicide," we need to know that it is measuring that construct and not something else (e.g., anxiety or hopefulness). CANS developers admitted that these construct validity studies should have been completed (Sieracki, Leon, Miller & Lyons, 2008). The developers previously tried, and the results were unacceptable (Miller, Leon & Lyons, 2007). In this unpublished conference presentation, the developers even recommended revising CANS because of the poor results, but they have not done so.

These validation failures are not trivial. They are fatal, requiring CANS to be reconstructed before it is used. Each of the CANS domains (e.g., "Strengths", "Risk Behaviors", "Functioning") do not include questions that belong together, nor do they measure anything meaningful. Each version of CANS must be re-constructed and re-evaluated for validity.

Each jurisdiction is in a strong position to assess and revise the construct validity of its unique version of CANS. Its large database of CANS administrations should be mined, and an expert in *exploratory* and *confirmatory* factor analysis should be hired to create a valid set of CANS constructs. Exploratory factor analytic (EFA) studies help the developer find the best possible set of questions for the underlying constructs being measured. If these EFA studies meet industry standards, then, on a different sample, these constructs are tested with different patients (*confirmatory* analyses) to make sure they are stable and not just unique to the smaller EFA study sample. With each state's large

datasets, three datasets could be randomly selected. One is for the exploratory analyses and the other two for confirmation. The hard work has already been done – the data have already been collected. Given the results of previous attempts, jurisdictions may not like the results and a valid version of CANS may look very different than the CANS versions they are using today. Nevertheless, working to create one unified version of CANS would allow for useful interstate benchmarking. Courts, the children they serve, and the communities they protect, deserve valid data at their fingertips.

CONCURRENT VALIDITY

Once a revised version of CANS is shown to have construct validity, the new CANS tool then needs to be compared to one or more gold standards of measurement. This is called concurrent validity. For example, a gold-standard assessment of depression is the Beck Depression Inventory (BDI; Beck, Steer, & Ranieri, 1988) and most concurrent validity studies testing the validity of a new measure of depression test the new scale's relationship or correlation to the BDI.

Here again, CANS developers tried to conduct this analysis with the very poor results shown below (Dilley & Lyons, 2003). Again, this failure is fatal, if not corrected, and must be replicated with each state's unique set of CANS questions.

In a recent review of CANS (Kraus, Baxter, Alexander & Bentley, 2015), this limitation was noted. The review pointed out, "for example, that CANS' depression and anxiety scale has a correlation coefficient of 0.18 with the Child And Adolescent Functional Assessment Scale (CAFAS) Moods/Emotions scale and the CANS family functioning scale has a correlation coefficient of 0.26 with the home performance scale of the CAFAS" (p. 3). Correlations should typically be well above 0.50 to show *minimal* validity.

It may be helpful to explain what these correlation coefficients mean. A correlation ranges from –1.0 to +1.0, with a score of zero meaning that there is no relationship between the two scores. A 1.0 (either negative or positive) means there is a complete one-to-one relationship between the scores. A perfect correlation would mean that if we knew the answer from one test, we would know the exact score on the other. For example, if we weigh an object on a scale calibrated to kilograms, there should be a perfect correlation of 1.00 with a scale calibrated to pounds.

To understand how poor these CANS' results are, we square the correlation coefficients (e.g., 0.18 x 0.18 = 0.03), to estimate the amount of overlapping variance between these measures. A well-validated measure would have a correlation in the 0.8 to 0.9 range indicating that there is a large amount of overlap between the two measures.

CANS' results, however, indicate that CANS anxiety and depression scales account for only 3% of the variance with the CAFAS. In other words, the CANS' measure of anxiety and depression tells us almost nothing about the CAFAS Moods/Emotions scale. They are measuring different things.

These correlations are devastating. Since CANS developers got to choose the gold-standard – a tool they thought was the most similar to CANS – it can be concluded that CANS is not validly measuring anxiety and depression. The same is true for almost all

other constructs within CANS. This failure could lead to many children receiving inappropriate medications and incorrect levels of care determinations. Remember, with FFPSA requirements, we are using CANS to decide if a child can have the love and support of a family-like setting, or be placed in an institution.

Each jurisdiction should select several gold-standard assessments and pay a subset of children and families (about 100) to complete these additional tools. This concurrent validity data would not cost much to collect and would provide dispositive validity data.

CONCLUSION

FFPSA requires that the bench determine whether a child's needs (short- and long-term goals) *cannot* be met through a family-like setting. The use of CANS or other tools are designed to assist the courts in making this determination. It is essential that the data are reliable and valid.

Since each jurisdiction uses a different set of CANS items, each jurisdiction must demonstrate to the juvenile or family courts that it is using a valid, customized measure. In most states, it would be prudent for the child welfare department that selected the unvalidated set of CANS questions to hire researchers and incur the expense to validate CANS before it is presented as accurate information to the judiciary. This roadmap can assist in creating a useful and valid measure of children's risk and functioning within a short period of time.

REFERENCES

Anderson, R. L., Lyons, J. S., Giles, D. M., Price, J. A., Estes, G. (2003). Examining the Reliability of the Child and Adolescent Needs and Strengths-Mental Health (CANS-MH) Scale from two perspectives: A comparison of clinician and researcher ratings. *Journal of Child and Family Studies*, *12*, 279–289.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). Standards for educational and psychological testing.

American Psychological Association. (2017). Ethical principles of psychologists and code of conduct.

Beck, A. T., Steer, R. A., & Ranieri, W. F. (1988). Scale for suicide ideation: Psychometric properties of a self-report version. *Journal of Clinical Psychology*, *44*, 499–505.

Dilley, J. B., & Lyons, J. S. (2003). *The validity of the Child and Adolescent Needs and Strengths Assessment*. Toronto, Ontario: Poster presentation at the American Psychological Association.

Kisiel, C. L., Blaustein, M., Fogler, J., Ellis, H., & Saxe, G. N. (2009). Treating children with traumatic experiences: Understanding and assessing needs and strengths. In J. S. Lyons & D. A. Weiner (Eds.), *Behavioral health care: Assessment, service planning, and total clinical outcomes management* (pp. 17.1–17.15). Kingston, NJ: Civic Research Institute.

Kraus, D. R. (2017). Rebuttal to Lyons & Israel. *Children and Youth Services Review*, *74*, 133–138.

Kraus, D. R., Baxter, E. E., Alexander, P. C., & Bentley, J. H. (2015). The Treatment Outcome Package (TOP): A multi-dimensional level of care matrix for child welfare. *Children and Youth Services Review*, *57*, 171–178.

Kraus, D. R., Boswell, J. F., Wright, A., Castonguay, L. G., & Pincus, A. L. (2010). Factor structure of the Treatment Outcome Package for children. *Journal of Clinical Psychology*, *66*, 1–14.

Lyons, J. S. & Israel, N. (2017) Reply to Kraus, et. al. *Children and Youth Services Review*, *74*, 129–132.

Lyons, J. S., Weiner, D. A., & Lyons, M. B. (2004). Measurement as communication. The Child and Adolescent Needs and Strengths tool. In M. Mariush (Ed.) *The use of psychological testing for treatment planning and outcome assessment*. 3rd Edition, volume *2*, Lawrence Erlbaum Associates, Inc, Mahwah, New Jersey pp 461–476.

Miller, A. S., Leon, S. C., & Lyons, J. S.(2007). The Child and Adolescent Needs and Strengths scale: Factor analytic investigations. Poster presented at The 2007 Meeting of the Midwestern Psychological Association, Chicago, Illinois.

Sieracki, J. H., Leon, S. C., Miller, S. A., & Lyons, J. S. (2008). Individual and provider effects on mental health outcomes in child welfare: A three level growth curve approach. *Children and Youth Services Review*, *30*, 800–808.

Winters, N. C., Collett, B. R., & Myers, K. M. (2005). Ten-year review of rating scales, VII: scales assessing functional impairment. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*, 309–38.