

Development and Validation of a Computerized-Adaptive Test for PTSD (P-CAT)

Susan V. Eisen, Ph.D., Mark R. Schultz, Ph.D., Pengsheng Ni, M.D., M.P.H., Stephen M. Haley, Ph.D., Eric G. Smith, M.D., Ph.D., Avron Spiro, Ph.D., Princess E. Osei-Bonsu, Ph.D., M.S.P.H., Sam Nordberg, Ph.D., Alan M. Jette, Ph.D.

Objective: The primary purpose was to develop, field test, and validate a computerized-adaptive test (CAT) for posttraumatic stress disorder (PTSD) to enhance PTSD assessment and decrease the burden of symptom monitoring.

Methods: Data sources included self-report and interviewer-administered diagnostic interviews. The sample included 1,288 veterans. In phase 1, 89 items from a previously developed PTSD item pool were administered to a national sample of 1,085 veterans. A multidimensional graded-response item response theory model was used to calibrate items for incorporation into a CAT for PTSD (P-CAT). In phase 2, in a separate sample of 203 veterans, the P-CAT was validated against three other self-report measures (PTSD Checklist, Civilian Version; Mississippi Scale for Combat-Related PTSD; and Primary Care PTSD Screen) and the PTSD module of the Structured Clinical Interview for DSM-IV.

Results: A bifactor model with one general PTSD factor and four subfactors consistent with *DSM-5* (reexperiencing, avoidance, negative mood-cognitions, and arousal), yielded good fit. The P-CAT discriminated veterans with PTSD from those with other mental health conditions and those with no mental health conditions (Cohen's *d* effect sizes $>.90$). The P-CAT also discriminated those with and without a PTSD diagnosis and those who screened positive versus negative for PTSD. Concurrent validity was supported by high correlations ($r=.85-.89$) with the validation measures.

Conclusions: The P-CAT appears to be a promising tool for efficient and accurate assessment of PTSD symptomatology. Further testing is needed to evaluate its responsiveness to change. With increasing availability of computers and other technologies, CAT may be a viable and efficient assessment method.

Psychiatric Services 2016; 67:1116–1123; doi: 10.1176/appi.ps.201500382

Computerized-adaptive tests (CATs) based on item response theory (IRT) provide brief but sensitive and accurate assessments of health status (1–4). IRT models estimate how precise test items are in assessing symptom levels at any point along a latent dimension (θ). The result is an efficient algorithm using precisely calibrated items to estimate level of distress, with a stopping point based on uncertainty of the estimate. A CAT first selects a highly discriminating item from an item bank. Given the person's previous response(s), the algorithm then selects the next item with the highest possible information that contributes to a score on the underlying dimension. This adaptive item presentation continues until a predefined measurement precision is reached or a specified number of items have been presented.

CATs have reduced the number of items needed for reliable and valid assessment to 14% of the original number, requiring only 20% of the time needed for completion of the original assessment (5). Median time to complete a 12-item CAT for anxiety was 2.48 minutes (6). This reduced

assessment time can be particularly valuable in fast-paced clinics, where efficiency is a high priority. Additional benefits of CATs include reduced scoring and processing time, increased measurement precision, and more individualized assessment (3,4,7). These benefits can reduce errors and increase accuracy of assessment in both clinical decision making and research (7).

CATs have been developed for several conditions, including anxiety, depression, physical functioning, community integration, and personality disorder (5,6,8–12). However, to date no CAT has been developed for posttraumatic stress disorder (PTSD). A recent Institute of Medicine report on PTSD treatment for service members and veterans noted that lack of routine outcomes measurement made it nearly impossible to determine whether treatment in naturalistic settings was effective (13). The report recommended that the “DoD [Department of Defense] and VA [Department of Veterans Affairs] should develop, coordinate, and implement a measurement-based PTSD

management system that documents patients' progress over the course of treatment . . . with standardized and validated instruments." The VA uses the PTSD Checklist (PCL) and a four-item screener (in primary care). A CAT could increase efficiency and utility over paper-and-pencil measures and facilitate evaluation of treatment outcomes at the provider, facility, and organization levels. Thus the goal of the work reported here was to develop a CAT to facilitate PTSD screening, assessment, and outcomes monitoring. Veterans are an appropriate population for a PTSD CAT (P-CAT) because they are at greater risk of PTSD than the general population (14), although a P-CAT could also be useful in non-VA treatment settings and for the general population.

METHODS

Overview

The study was conducted in two phases. In phase 1, we administered a previously developed item bank to 1,085 veterans to calibrate the items and develop the P-CAT (15). In phase 2, we administered the P-CAT to a separate sample of 203 veterans to test its validity. Data collection began in January 2011 and was completed in September 2013.

Participants

Calibration sample. The calibration sample included 1,085 veterans: 908 were recruited from a national panel maintained by an Internet survey company that provides its members with a computer and Internet access. To ensure inclusion of individuals on the higher end of the PTSD symptom spectrum, this sample was supplemented with 177 veterans who had a diagnosis of and were being treated for PTSD at a VA hospital. Eligibility criteria for both groups included at least one traumatic event (*DSM-IV-TR* PTSD criterion A1 [16]) and consent to complete a one-time survey assessing demographic characteristics, stressful life events, and 89 PTSD symptoms (the item bank). To include veterans from various wartime eras, the sample was stratified by age (50% less than age 45 and 50% age 45 or above). To ensure representation of women, they were oversampled to constitute 20% of the sample. Oversampling segments of the population has been used in previous CAT development work (17,18).

Sample for field test validation. The field test validation enrolled a convenience sample of 203 veterans, similarly stratified by age and gender. Three groups were enrolled. Group 1 (N=91) included veterans who had received a diagnosis of and treatment for PTSD within the past year (at least two visits with a PTSD diagnosis and two visits for PTSD specialty care). Group 2 (N=60) included veterans treated for a different mental health condition (no PTSD but at least two non-PTSD mental health visits in the past year). Group 3 (N=52) included veterans treated for a general medical condition but not a mental health condition (no psychiatric diagnoses or mental health visits but at least two

TABLE 1. Characteristics of 1,288 veterans in two samples

Characteristic	Calibration sample (N=1,085)		Validation sample (N=203)	
	N	%	N	%
Age				
18–30	109	10	29	14
31–45	386	36	44	22
45–60	246	23	72	35
≥61	344	32	58	29
Gender				
Male	851	78	161	79
Female	234	22	42	21
Race ^a				
White	821	76	147	72
African American	91	8	37	18
Other nonwhite	88	8	11	5
Latino ethnicity	78	7	15	7
Education				
Less than high school	35	3	0	—
High school or GED	213	20	31	15
Some college	501	46	114	56
Bachelor's degree or higher	335	31	58	29
Employment ^a				
Employed	573	53	69	34
Student	63	6	24	12
Homemaker	46	4	5	2
Unemployed or disabled	233	21	97	48
Retired	305	28	52	26
Self-reported diagnosis ^a				
PTSD	237	22	103	51
Other mental health condition ^b	404	37	135	67
No mental health condition	646	60	51	25

^a Percentages exceed 100 because multiple responses were possible.

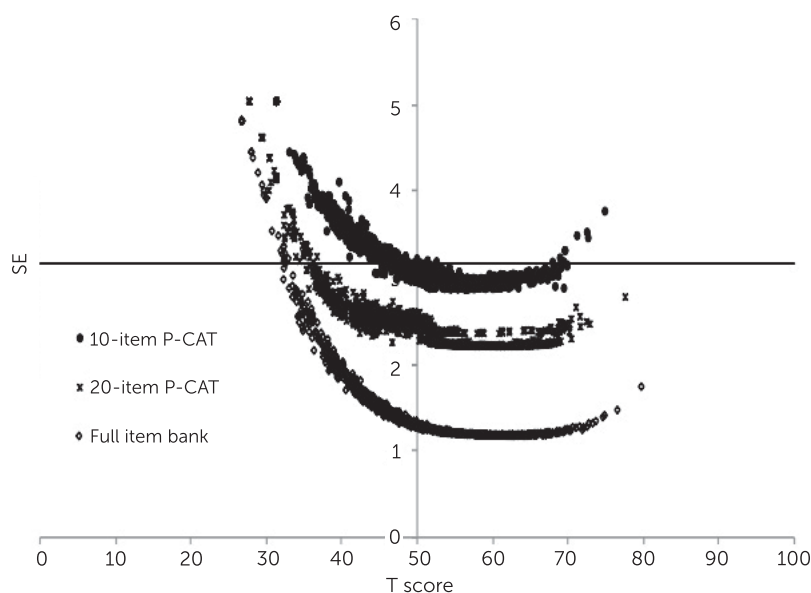
^b Includes alcohol or drug abuse or addiction

general medical visits in the past year). Because of clinician concerns about symptom exacerbation, veterans with a suicide "flag" in the medical record were excluded (<1.3% of mental health service users at participating sites).

Calibration Measures

Traumatic exposure screen. The screening question from the PTSD module of the Structured Clinical Interview for DSM-IV (SCID) (19) was used to screen for traumatic exposure: "Sometimes things happen . . . that are extremely upsetting . . . like being in a life threatening situation . . . major disaster, . . . serious accident, . . . fire, . . . physically assaulted or raped, seeing another person killed, dead, or badly hurt. . . Have any of these . . . things happened to you?" An affirmative answer was followed by the question, "At that time, did you feel intense fear, helplessness or horror?" A second affirmative answer was required for inclusion in the calibration sample.

PTSD item bank. The item bank consisted of 89 items described previously (15). Briefly, we used the *DSM-IV* PTSD diagnostic criteria as a conceptual framework, conducted a systematic review of PTSD instruments, and created a database of items from existing instruments. Although *DSM-5* was not

FIGURE 1. Standard error (SE) estimates for simulated ten- and 20-item P-CATs and the full item bank^a

^aP-CAT, computerized-adaptive test for PTSD. The solid horizontal line represents the SE level of 3.16 on the T scale, where the score reliability equals .90 (corresponding to .316 SE on the Z scale) (36). Points at or below this line indicate reliability $\geq .90$. Points above the line indicate reliability $< .90$.

completed when we began the study, we followed its progress and included items from new domains that were likely to emerge. The time frame for all items was the past month.

Validation Measures

Three self-report measures were used to validate the P-CAT: the civilian version of the PCL (PCL-C), Mississippi Scale for Combat-Related PTSD (M-PTSD), and Primary Care PTSD Screen (PC-PTSD) (20–22). To minimize burden and distress for participants, we used the SCID PTSD module (19) to obtain a standardized clinical diagnosis rather than the Clinician-Administered PTSD Scale. The PCL-C is a 17-item PTSD symptom scale with high reliability (.96) and validity (23). The M-PTSD is a 35-item instrument with high sensitivity (.93) and specificity (.89) (21). The PC-PTSD is a four-item screener with high sensitivity ($> .75$) and specificity ($> .86$) (22) that has been implemented in VA primary care. Validation measures were based on *DSM-IV* because no *DSM-5* measures were available at the time.

Procedures

Item calibration. All calibration sample participants ($N=1,085$) completed demographic questions and the 89-item PTSD item bank. The Internet sample was contacted and screened for eligibility by the Internet survey company; participants then completed the survey via the Internet. The VA sample completed a mailed survey. Written consent for this phase of the study was waived by the institutional review boards.

Field test item validation. All validation sample participants ($N=203$) provided written informed consent. They were then

administered the P-CAT on a tablet computer and also completed a paper-and-pencil packet containing the validation measures. A trained interviewer administered the SCID PTSD module. All procedures were approved by the institutional review boards of the participating hospitals.

Data Analysis

Using the calibration sample, we conducted a factor analysis to assess dimensionality of the P-CAT items. We first examined a unidimensional one-factor model by using confirmatory factor analysis. Anticipating that PTSD symptoms might encompass multiple domains paralleling the diagnostic criteria, we also explored a bifactor model, which allowed for a primary dimension (PTSD) and multiple subdomains (6,8). Model fit was examined by fit indices; model comparison was based on the likelihood ratio chi-square test. On the basis of the fitted model, we applied a multidimensional graded-response IRT model to estimate the item parameters and examine item fit on the basis of the Z

score (standardized difference between the observed and expected log likelihood of response patterns). We used a one-sided test; under the null hypothesis, the z scores were normally distributed. A value less than -1.645 indicated misfit (24,25). Analyses were conducted using Mplus, IRTPRO, and SAS, version 9.2 (26,27).

P-CAT Development and Psychometric Properties

To create the P-CAT, we used CAT software developed by the Health and Disability Research Institute (11,28,29). Using a maximum a posteriori estimation procedure (30), the software selects successive items that maximize precision of the respondent's score estimate, based on responses to each previous item. We used Newton-Raphson iterative methods (31) to compute initial estimates of PTSD severity; the score was updated after each successive response. The first P-CAT item presented was selected on the basis of the highest information function across the score range and on the basis of item content considered central to PTSD. We set 12 items as the program stopping rule to allow scores for all four *DSM-5* domains. To ensure content balancing, we required the first four items to come from each of the four PTSD domains, after which the program selected subsequent items from those remaining in the item bank. To optimize item selection, we selected items with the first-, second-, and third-highest determinants of the posteriori information at the respondent's score level (32). If the second or third item's general PTSD factor discrimination parameter was greater than 2.7 (median value of the general PTSD factor discrimination parameter), we randomly selected one of them; otherwise we selected the first item as the next administered.

To examine P-CAT score accuracy and precision, we generated the score and standard error (SE) estimates for simulated ten- and 20-item P-CATs and compared them through real-data simulation with those generated from the full item bank, a method used to reduce the length of a conventional test and determine the minimum number of items needed to achieve acceptable accuracy and precision (32). To evaluate accuracy, we calculated the correlation between scores generated by the P-CATs and those for the full item bank. Precision (reliability) was assessed by calculating the SEs across the range of scores for the P-CATs and for the full item bank. Reliability was defined as $\rho = 1 - SE(\theta)^2$. Better accuracy and precision mean higher reliability and lower SEs.

Validation

Using the validation sample, we assessed concurrent validity by examining correlations between the P-CAT, the M-PTSD, and the PCL-C. We assessed discriminant validity by examining differences in P-CAT scores on the basis of clinically diagnosed and treated PTSD, a SCID PTSD diagnosis, and a positive PTSD screen. Finally, we examined sensitivity and specificity of the P-CAT against the SCID PTSD diagnosis by using receiver operating characteristics (ROCs).

RESULTS

Sample Characteristics

Both the calibration and validation samples were predominantly white (76% and 72%, respectively) and male (78% and 79%) (Table 1). Most had graduated from high school, and about a third had graduated from college (31% for the calibration sample and 29% for the validation sample). In the calibration sample, 22% reported a PTSD diagnosis, and 37% reported another mental health condition; 60% reported no psychiatric diagnosis. Reported PTSD and other mental health conditions were substantially higher in the validation sample (51% and 67%, respectively).

Factor Analysis and Item Calibration

A bifactor model (33,34) with one general PTSD factor and four subfactors, consistent with the four *DSM-5* (35) domains (reexperiencing, avoidance, negative mood-cognitions, and arousal), yielded much better fit than a unidimensional model (comparative fit index=.958,

TABLE 2. Discriminant validity of the P-CAT among 203 veterans, by clinical group^a

P-CAT total and domain	PTSD (N=91)		Other mental health condition (N=60)		No mental health condition (N=52)		F ^b
	M	SD	M	SD	M	SD	
Total	62.98	4.41	57.17	7.39	52.93	7.82	43.75
Reexperiencing	63.29	4.79	57.09	7.70	52.78	8.11	43.73
Avoidance	62.19	4.13	57.44	6.68	53.85	7.83	33.07
Negative mood-cognitions	62.49	4.47	56.99	7.54	52.09	7.65	45.64
Arousal	62.71	4.24	56.55	7.36	52.58	8.19	44.41

^a P-CAT, computerized-adaptive test for PTSD. Possible P-CAT scores range from 1 to 100, with higher scores indicating greater symptom severity.

^b df=2 and 200; $p < .001$ for all comparisons

Tucker-Lewis index=.956, root mean square error of approximation=.058; $\chi^2=4,821.30$, $df=89$, $p < .001$, for the likelihood ratio test between models). All items fitted the model with z greater than -1.645 .

P-CAT Scoring and CAT Simulations

P-CAT scores were standardized (mean=0 and SD=1). To ease interpretation, we converted the z scores to T scores, with a mean of 50 and an SD of 10. Higher scores indicate greater PTSD symptom severity. The score correlations of the full item bank (89 items), with simulated ten- and 20-item CATs were .94 and .98, respectively. Figure 1 presents the score SE estimate distributions, which show that P-CAT precision was higher at moderate and severe PTSD symptom levels; precision generally decreased at mild or minimal PTSD symptom levels. Regarding score reliability $>.90$, the 20-item CAT worked about as well as the full item bank (36). When the score exceeded the mean of 50, the ten-item CAT reliability was $>.90$, indicating that the ten-item worked better in cases of moderate or severe PTSD symptom severity.

Concurrent and Discriminant Validity

The Pearson correlation of the P-CAT with the PCL-C was $r=.88$, and with the M-PTSD, it was $r=.85$, indicating strong concurrent validity. P-CAT scores also showed highly

TABLE 3. Concurrent and discriminant validity of the P-CAT among 203 veterans, by SCID diagnosis and PTSD screen^a

P-CAT total and domain	PTSD diagnosis					PTSD screen				
	Yes (N=129)		No (N=71)		F ^{b,c}	Positive (N=140)		Negative (N=63)		F ^{b,d}
	M	SD	M	SD		M	SD	M	SD	
Total	62.18	5.10	52.04	6.97	138.5	62.35	4.86	50.55	6.07	218.1
Reexperiencing	62.26	5.67	52.09	7.31	119.5	62.54	5.28	50.38	6.37	202.3
Avoidance	61.55	4.70	53.17	7.16	99.0	61.65	4.40	51.98	6.97	143.0
Negative mood-cognitions	61.77	5.24	51.39	6.76	145.4	61.85	5.04	50.09	6.09	206.9
Arousal	61.80	5.15	51.58	7.06	137.7	61.98	4.93	50.10	6.19	214.1

^a P-CAT, computerized-adaptive test for PTSD. SCID, Structured Clinical Interview for DSM-IV. Possible P-CAT scores range from 1 to 100, with higher scores indicating greater symptom severity. The SCID was not completed for 3 veterans.

^b $p < .001$ for all comparisons

^c $df=1$ and 199

^d $df=1$ and 202

TABLE 4. P-CAT items and scores for two sample respondents with low and high PTSD symptom severity

Respondent and item	Domain	Response	Score ^a	SE
Low symptom severity				
I felt upset when I was reminded of the trauma	Reexperiencing	Never	45.3	.66
I had sleep problems	Arousal	Sometimes	46.9	.52
I avoided situations that might remind me of something terrible that happened to me	Avoidance	Never	46.1	.48
I felt distant or cut off from people	Negative mood-cognitions	Never	44.9	.44
I found myself remembering bad things that happened to me	Reexperiencing	Rarely	45.9	.36
I lost interest in social activities	Negative mood-cognitions	Never	47.0	.36
I felt jumpy or easily startled	Arousal	Rarely	47.0	.31
I had bad dreams or nightmares about the trauma	Reexperiencing	Never	46.9	.30
I felt that if someone pushed me too far, I would become angry	Arousal	Rarely	47.1	.28
I had flashbacks (sudden, vivid, distracting memories) of the trauma	Reexperiencing	Never	47.0	.28
I had trouble concentrating	Arousal	Rarely	47.4	.26
I felt that I had no future	Negative mood-cognitions	Never	47.3	.26
High symptom severity				
I felt upset when I was reminded of the trauma	Reexperiencing	Often	62.5	.50
I had sleep problems	Arousal	Often	62.1	.45
I tried to avoid activities, people, or places that reminded me of the traumatic event	Avoidance	Often	62.7	.37
I lost interest in social activities	Negative mood-cognitions	Sometimes	61.9	.33
I had flashbacks (sudden, vivid, distracting memories) of the trauma	Reexperiencing	Sometimes	61.7	.31
I felt emotionally numb	Negative mood-cognitions	Often	62.5	.30
Memories of the trauma kept entering my mind	Reexperiencing	Sometimes	62.3	.29
I felt distant or cut off from people	Negative mood-cognitions	Often	62.5	.28
I had bad dreams about terrible things that have happened to me	Reexperiencing	Sometimes	62.4	.28
I felt jumpy or easily startled	Arousal	Often	63.0	.25
Any reminder brought back feelings about the trauma	Reexperiencing	sometimes	63.4	.24
I felt that I had no future	Negative mood-cognitions	Rarely	63.3	.24

^a P-CAT, computerized-adaptive test for PTSD. Possible P-CAT scores range from 1 to 100, with higher scores indicating greater symptom severity.

significant differences among the three clinical groups in the validation sample (PTSD, other mental health condition, and no mental health condition) (Table 2), with Cohen's *d* effect sizes $>.90$. Differences were also significant between participants who did or did not meet SCID criteria for PTSD and between those who did or did not screen positive for PTSD (Table 3).

Sensitivity and Specificity

The ROC analysis indicated that the area under the curve (which quantifies the P-CAT's ability to discriminate between participants meeting SCID criteria for PTSD from those without PTSD) was .88. A cutoff score of 58 provided sensitivity of 82% and specificity of 80%. For the PCL, a cutoff score ≥ 50 yielded sensitivity of .70 and specificity of .87. For the PC-PTSD screener, a cutoff score ≥ 3 yielded sensitivity of .90 and specificity of .71.

Example of P-CAT Administration

The first P-CAT item is "I felt upset when I was reminded of the trauma." Table 4 presents response profiles, P-CAT scores, and SEs for two individuals—one with low and one with high PTSD symptom severity. As shown in Table 4, different responses to the same item resulted in markedly different P-CAT scores and different subsequent items.

DISCUSSION

Accurate assessment of mental health is critical for referring patients to appropriate services and monitoring outcomes. CATs can facilitate this process and reduce the burden of assessment on patients and health systems. The P-CAT adds another psychiatric condition for which a CAT is now available, in addition to depression and anxiety. Although PTSD is not as prevalent in the general population as depression or anxiety, PTSD can be highly disabling and expensive to treat (37–39). Thus a briefer, more efficient, and cost-effective measure of PTSD can be valuable to patients,

clinicians, and researchers. Average completion time for the P-CAT was 116 seconds, with 89% of participants completing it in less than three minutes (data not shown). This completion time is similar to that reported for a seven-item screener and less than the five to ten minutes for the PCL-C and other instruments of similar length (40).

The P-CAT showed strong concurrent and discriminant validity, and sensitivity (.82) and specificity (.80) were comparable with those of other measures based on the same cut points (sensitivity ranging from .70 to .78 and specificity from .85 to .92 for the PC-PTSD screen, and sensitivity from .21 to .82 and specificity from .84 to .99 for the PCL-C) (14,41). The fact that a simulated ten-item P-CAT was not as reliable at low levels of PTSD suggests that some patients who do not have PTSD may test as false positives. Further clinical evaluation would be needed to make this

determination. Test-retest reliability, predictive validity, and responsiveness to change were not assessed, which is a limitation of this study.

The stopping rule used for the P-CAT was 12 items to allow scoring of the four PTSD domains and an overall score. However, the two sample administrations shown in Table 4 indicate that for low PTSD symptom severity, only eight items were needed to reach an SE of .30 (which implies reliability $>.90$ [42]); for high PTSD symptom severity, only six items were needed to reach this SE. Thus it is likely that the P-CAT could be substantially shorter than 12 items without significant loss of precision while still allowing for assessment of the four domains. However, even the 12-item P-CAT is shorter, less burdensome, more individualized, and more efficiently scored and processed than the 35-item M-PTSD or the new 20-item PCL-5 (43). A P-CAT shorter than four items would not allow for reliable domain scores.

Outcome evaluation of evidence-based treatments may be possible with the P-CAT, although further work to assess the instrument's responsiveness to change is needed. Two evidence-based psychotherapies for PTSD, cognitive processing (44) and prolonged exposure (45), require weekly monitoring of PTSD symptoms. The P-CAT allows items to vary from one administration to the next, which can reduce both practice and memory effects that may occur with paper-and-pencil measures, but does not allow for monitoring change in the same items over time, which may be a limitation of adaptive testing. However, even with varying items, the P-CAT provides scores for all four PTSD symptom domains, allowing assessment of change in each domain. Further testing of the P-CAT at multiple time points is needed to assess responsiveness to change, which has been shown for other mental health CATs used in clinical settings (46,47).

The P-CAT was developed among U.S. military veterans, for whom trauma is likely related to military service, although other types of trauma are also possible. Among civilians, military service is not a source of trauma, and the proportion of women is much higher in the civilian population. Consequently, future work should test the P-CAT in broader samples to understand how it can best be used in clinical practice and research. Calibration of items in other samples (for example, women and college students) may identify different items with high information value.

This study excluded veterans at high risk of suicide because of clinical concerns about potential symptom exacerbation; however, $<2\%$ of mental health service users were excluded. The study included veterans with comorbid diagnoses (for example, depression and substance abuse), which commonly co-occur with PTSD.

During the course of this study, *DSM-5* (43) replaced *DSM-IV*, and each has different diagnostic criteria for PTSD. With *DSM-5*, criterion A1, exposure to trauma, now must be a more direct experience; criterion A2 (reaction with intense fear, helplessness, or horror) is eliminated; and symptoms have been added to cover revamped symptom clusters.

Recent research suggests that the prevalence of PTSD among nonveterans is somewhat lower when measured by *DSM-5* than when measured by *DSM-IV* (4.2% and 5.1%, respectively), but prevalence rates are almost identical among veterans with high PTSD prevalence (38.8% when measured by *DSM-5*, and 39.9% by *DSM-IV*) (48,49). Because revised diagnostic criteria may alter the comparison groups used to validate the P-CAT, future work is warranted to test our results.

CONCLUSIONS

The P-CAT adds to the growing collection of CATs available for screening and symptom monitoring of general medical and mental health conditions. In 2008, the question was raised, "Are we ready for computerized adaptive testing?" (3). Concerns were expressed in regard to technology and infrastructure limitations in health care facilities. Now, eight years later, these technologies have proliferated, including tablet computers and smart phones. Many health care systems offer patients secure Internet portals for communication of sensitive health information (including VA through its electronic patient portal, MyHealtheVet). The P-CAT can be installed on any of these media, thus enabling implementation in many mental health care settings.

AUTHOR AND ARTICLE INFORMATION

When this work was done, Dr. Eisen was with the Center for Healthcare Organization and Implementation Research at the Edith Nourse Rogers Memorial Veterans Hospital, Bedford, Massachusetts, where Dr. Schultz and Dr. Smith are affiliated. Dr. Eisen was also formerly with the Department of Health Policy and Management, Boston University School of Public Health, Boston (e-mail: seisen@bu.edu). Dr. Smith is also with the Department of Psychiatry, University of Massachusetts Medical School, Worcester. Dr. Ni and Dr. Jette are with the Health and Disability Research Institute, Boston University School of Public Health, Boston, where the late Dr. Haley was affiliated. Dr. Spiro is with the Massachusetts Veterans Epidemiology Research and Information Center, Jamaica Plain Campus, Department of Veterans Affairs (VA) Boston Health Care System, and with the Department of Psychiatry, Boston University School of Medicine, Boston. Dr. Osei-Bonsu is with the Center for Chronic Disease Outcomes Research, Minneapolis VA Health Care System, Minneapolis, Minnesota. Dr. Nordberg is with Atrius Health, Harvard Vanguard Medical Associates, Boston. Some of the results were presented at the annual meeting of the International Society for Traumatic Stress Studies, November 7–9, 2013.

This research was supported by grant IIR 09-342 from the VA Health Services Research Development Service (HSR&D). Dr. Spiro was supported by a Senior Research Career Scientist award from the VA Clinical Sciences Research and Development Service. Dr. Smith was supported by a Career Development Award from HSR&D. The authors thank Eve Davison, Ph.D., Rani Elwy, Ph.D., Lawrence Herz, M.D., Terry Keane, Ph.D., Brian Marx, Ph.D., Karen Ryabchenko, Ph.D., Dawne Vogt, Ph.D., and the veterans who served as expert panel members for this study. They also thank Patrick Furlong, Alexandra Howard, Jenniffer Leyson, M.A., and Linda McCoy, M.S., for assistance with data collection and administrative support for the study. The views expressed in this article are those of the authors and do not necessarily represent the views of the VA.

Dr. Eisen reports receipt of a proportion of licensing fees collected from users of an outcome measure for which she is the primary author. The

measure was not used in this study. Dr. Jette holds stock in a small business that licenses and disseminates outcome instruments unrelated to this work.

Received September 4, 2015; revision received January 28, 2016; accepted February 26, 2016; published online June 1, 2016.

REFERENCES

- Hays RD, Morales LS, Reise SP: Item response theory and health outcomes measurement in the 21st century. *Medical Care* 38(suppl): II28–II42, 2000
- Revicki DA, Cella DF: Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Quality of Life Research* 6:595–600, 1997
- Unick GJ, Shumway M, Hargreaves W: Are we ready for computerized adaptive testing? *Psychiatric Services* 59:369, 2008
- Gibbons RD, Weiss DJ, Kupfer DJ, et al: Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatric Services* 59:361–368, 2008
- Haley SM, Raczek AE, Coster WJ, et al: Assessing mobility in children using a computer adaptive testing version of the Pediatric Evaluation of Disability Inventory. *Archives of Physical Medicine and Rehabilitation* 86:932–939, 2005
- Gibbons RD, Weiss DJ, Pilkonis PA, et al: Development of the CAT-ANX: a computerized adaptive test for anxiety. *American Journal of Psychiatry* 171:187–194, 2014
- Kraemer HC, Freedman R: Computer AIDS for the diagnosis of anxiety and depression. *American Journal of Psychiatry* 171: 134–136, 2014
- Gibbons RD, Weiss DJ, Pilkonis PA, et al: Development of a computerized adaptive test for depression. *Archives of General Psychiatry* 69:1104–1112, 2012
- Resnik L, Borgia M, Ni P, et al: Reliability, validity and administrative burden of the Community Reintegration of Injured Service Members Computer Adaptive Test (CRIS-CAT). *BMC Medical Research Methodology* 12:145, 2012
- Simms LJ, Goldberg LR, Roberts JE, et al: Computerized adaptive assessment of personality disorder: introducing the CAT-PD project. *Journal of Personality Assessment* 93:380–389, 2011
- McDonough CM, Tian F, Ni P, et al: Development of the computer-adaptive version of the Late-Life Function and Disability Instrument. *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 67:1427–1438, 2012
- Achtyes ED, Halstead S, Smart L, et al: Validation of computerized adaptive testing in an outpatient nonacademic setting: the VOCATIONS Trial. *Psychiatric Services* 66:1091–1096, 2015
- Institute of Medicine: Treatment for Posttraumatic Stress Disorder in Military and Veteran Populations: Final Assessment. Washington, DC, National Academies Press, 2014
- Gates MA, Holowka DW, Vasterling JJ, et al: Posttraumatic stress disorder in veterans and military personnel: epidemiology, screening, and case recognition. *Psychological Services* 9:361–382, 2012
- Del Vecchio N, Elwy AR, Smith E, et al: Enhancing self-report assessment of PTSD: development of an item bank. *Journal of Traumatic Stress* 24:191–199, 2011
- Diagnostic and Statistical Manual of Mental Disorders, 4th ed, Text Revision. Washington, DC, American Psychiatric Association, 2000
- Rose M, Bjorner JB, Gandek B, et al: The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. *Journal of Clinical Epidemiology* 67:516–526, 2014
- Eisen SV, Normand SLT, Belanger A, et al: The Revised Behavior and Symptom Identification Scale (BASIS-R): reliability and validity. *Medical Care* 42:1230–1241, 2004
- First MB, Spitzer RL, Gibbon M, et al: Structured Clinical Interview for DSM-IV Axis I Disorders: Research Version, Patient Edition. New York, New York State Psychiatric Institute, Biometrics Research Department, 1996
- Weathers FW: PCL-C for DSM-IV. Boston, National Center for PTSD–Behavioral Science Division, 1991
- Keane TM, Caddell JM, Taylor KL: Mississippi Scale for Combat-Related Posttraumatic Stress Disorder: three studies in reliability and validity. *Journal of Consulting and Clinical Psychology* 56: 85–90, 1988
- Prins AP, Ouimette R, Kimerling RP, et al: The Primary Care PTSD screen (PC-PTSD): development and operating characteristics. *Primary Care Psychiatry* 9:9–14, 2004
- Norris F, Hamblen JL: Standardized self-report measures of civilian trauma and PTSD; in *Assessing Psychological Trauma and PTSD*, 2nd ed. Edited by Wilson JP, Keane TM. New York, Guilford, 2004
- Drasgow F, Levine M, Williams E: Appropriateness measurement with polytomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology* 38: 67–86, 1985
- Ackerman T, Hombro C, Neustel S: Evaluating indices used to assess the goodness-of-fit of the compensatory multidimensional item response theory model. Presented at the National Council on Measurement in Education, New Orleans, April 2–4, 2002
- Muthén B, Muthén L: *Mplus Statistical Analysis With Latent Variables: User's Guide*. Los Angeles, Muthén & Muthén, 2010
- Thissen D (ed): The MEDPRO project: an SBIR project for a comprehensive IRT and CAT software system—IRT software; in *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. Minneapolis, Minn, Graduate Management Admissions Council, 2009
- Haley SM, Ni P, Hambleton RK, et al: Computer adaptive testing improved accuracy and precision of scores over random item selection in a physical functioning item bank. *Journal of Clinical Epidemiology* 59:1174–1182, 2006
- Haley SM, Ni P, Ludlow LH, et al: Measurement precision and efficiency of multidimensional computer adaptive testing of physical functioning using the Pediatric Evaluation of Disability Inventory. *Archives of Physical Medicine and Rehabilitation* 87:1223–1229, 2006
- Wang SD, Wang TY: Precision of Warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing. *Applied Psychological Measurement* 25:317–331, 2001
- Lord FM: *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ, Erlbaum, 1980
- Segall DO: Multidimensional adaptive testing. *Psychometrika* 61: 331–354, 1996
- Gibbons RD, Hedeker D: Full-information item bi-factor analysis. *Psychometrika* 57:423–436, 1992
- Gibbons RD, Bock RD, Hedeker D, et al: Full-information item bi-factor analysis of graded response data. *Applied Psychological Measurement* 31:4–19, 2007
- Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Arlington, Va, American Psychiatric Association, 2013
- Wainer H (ed): *Computerized Adaptive Testing: A Primer*, 2nd ed. Mahwah, NJ, Erlbaum, 2000
- Amir M, Kaplan Z, Efroni R, et al: Suicide risk and coping styles in posttraumatic stress disorder patients. *Psychotherapy and Psychosomatics* 68:76–81, 1999
- Davidson JRT, Stein DJ, Shalev AY, et al: Posttraumatic stress disorder: acquisition, recognition, course, and treatment. *Journal of Neuropsychiatry* 16:135–147, 2004
- Kessler RC, Zhao S, Katz SJ, et al: Past-year use of outpatient services for psychiatric problems in the National Comorbidity Survey. *American Journal of Psychiatry* 156:115–123, 1999
- Orsillo SM: Measures for acute stress disorder and posttraumatic stress disorder; in *Practitioner's Guide to Empirically Based Measures of Anxiety*. New York, Kluwer Academic/Plenum, 2001

41. Spoont M, Arbisi P, Fu S, et al: Screening for Post-Traumatic Stress Disorder (PTSD) in Primary Care: A Systematic Review. VA-ESP Project 09-009. Washington, DC, US Department of Veterans Affairs, 2013. www.ncbi.nlm.nih.gov/books/NBK126691/
42. Gibbons RD, Weiss DJ, Pilkonis PA, et al: Computerized adaptive test-depression inventory not ready for prime time-reply. *JAMA Psychiatry* 70:763–765, 2013
43. Weathers FW, Litz BT, Keane TM, et al: The PTSD Checklist for DSM-5 (PCL-5). Washington, DC, US Department of Veterans Affairs, National Center for PTSD, 2013
44. Resick PA, Schnicke MK: Cognitive processing therapy for sexual assault victims. *Journal of Consulting and Clinical Psychology* 60:748–756, 1992
45. Foa EB, Hembree EA, Rothbaum BO: Prolonged Exposure Therapy for PTSD: Emotional Processing of Traumatic Experiences. New York, Oxford University Press, 2007
46. Devine J, Fliege H, Kocalevent R, et al: Evaluation of computerized adaptive tests (CATs) for longitudinal monitoring of depression, anxiety, and stress reactions. *Journal of Affective Disorders* 190: 846–853, 2016
47. Wagner LI, Schink J, Bass M, et al: Bringing PROMIS to practice: brief and precise symptom screening in ambulatory cancer care. *Cancer* 121:927–934, 2015
48. Kilpatrick DG, Resnick HS, Milanak ME, et al: National estimates of exposure to traumatic events and PTSD prevalence using DSM-IV and DSM-5 criteria. *Journal of Traumatic Stress* 26:537–547, 2013
49. Miller MW, Wolf EJ, Kilpatrick D, et al: The prevalence and latent structure of proposed DSM-5 posttraumatic stress disorder symptoms in US national and veteran samples. *Psychological Trauma* 5: 501–512, 2013

Submissions Invited for Datapoints Column

Datapoints encourages the rapid dissemination of relevant and timely findings related to clinical and policy issues in psychiatry. National or international data, especially from large representative databases, are preferred. The editors are particularly interested in data that can be accessed by other researchers. Topics may include differences or trends in diagnosis and practice patterns or in treatment modalities, especially across different care settings or in the context of new policies or payment sources. The analyses should be straightforward, so that the data displayed tell a clear story. The text should follow the standard research format and include a brief introduction, description of the methods and data set, description of the results, and comments on the implications or meanings of the findings.

Datapoints columns must include one figure or table, and because the column is limited to one printed page, it is therefore limited to 350–400 words. Submissions with multiple authors are discouraged because of space constraints.

Inquiries or submissions should be directed to the column editors: Amy M. Kilbourne, Ph.D., M.P.H. (amykilbo@umich.edu), or Tami L. Mark, Ph.D. (tami.mark@truvenhealth.com).