Customized computer-based administration of the PCL-5 for the efficient assessment of PTSD

A proof-of-principle study

Finkelman, M.D.; Lowe, S.R.; Kim, W.; Gruebner, O.; Smits, N.; Galea, S.

DOI
10.1037/tra0000226

Publication date
2017

Document Version
Final published version

Published in
Psychological Trauma

Psychological Trauma: Theory, Research, Practice, and Policy
2017, Vol. 9, No. 3, 379–389

1942-9681/17/$12.00 http://dx.doi.org/10.1037/tra0000226

Customized Computer-Based Administration of the PCL-5 for the Efficient Assessment of PTSD: A Proof-of-Principle Study

Matthew D. Finkelman
Tufts University School of Dental Medicine

Sarah R. Lowe
Montclair State University

Wonsuk Kim
Measured Progress, Inc., Dover, New Hampshire

Oliver Gruebner
T.H. Chan School of Public Health, Harvard University

Niels Smits
Research Institute of Child Development and Education, University of Amsterdam

Sandro Galea
School of Public Health, Boston University

Objective: To investigate the potential of customized computer-based testing procedures to reduce the mean test length of the Posttraumatic Stress Checklist for *DSM–5* (PCL-5). ***Method:*** A retrospective analysis was conducted using responses from 942 adults who had completed the full-length (20-item) PCL-5 in the aftermath of Hurricane Sandy. The abilities of 2 testing procedures, curtailment and stochastic curtailment, to lessen the instrument's mean test length while maintaining the same result as the full-length PCL-5 ("positive" or "negative") were evaluated in a post hoc simulation. Curtailment and stochastic curtailment track a respondent's answers as she takes the instrument and stop the test if future items are unable or unlikely to change the result. The performance of each procedure was recorded under 2 scoring methods: a total-score-based method and a cluster-based method. Each procedure's sensitivity, specificity, and overall agreement with the full-length PCL-5 were computed. ***Results:*** Curtailment reduced the mean test length by 40% under the total-score-based method, and by more than 70% under the cluster-based method, while exhibiting 100% sensitivity, specificity, and overall agreement with the full-length PCL-5. Stochastic curtailment reduced the mean test length by up to 88% under the total-score-based method, and up to 84% under the cluster-based method, while always exhibiting at least 92% sensitivity and 99.8% overall agreement, as well as 100% specificity, for the full-length PCL-5. ***Conclusions:*** Curtailment and stochastic curtailment have potential to enhance the efficiency of the PCL-5 when this assessment is administered by computer. The 2 procedures should be evaluated in future prospective studies.

Keywords: curtailment, PCL-5, PTSD, respondent burden, stochastic curtailment

Epidemiologic studies have shown that most people will experience one or more traumatic events over the course of their lifetime (e.g., Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995). Of those who have experienced trauma, it is estimated that 7.8% will go on to develop posttraumatic stress disorder (PTSD), a psychiatric condition characterized by intrusive thoughts, avoidance, negative alterations in cognition and mood, and changes in arousal and reactivity (American Psychiatric Association, 2013; Kessler et al., 1995). PTSD, especially when the disorder takes a chronic course, has been associated with substantial mental and physical health comorbidity and disability (e.g., Dickstein, Suvak, Litz, & Adler, 2010; Sareen et al., 2007). Early interventions for PTSD have shown promise for reducing the likelihood of chronic symptoms and other adverse posttrauma outcomes (Kearns, Ressler, Zatzick, & Rothbaum, 2012), yet their impact relies on providers' ability to accurately screen for PTSD in recent trauma survivors, who often face other pressing concerns, including the financial, legal and social consequences of the events they have endured (e.g., Lowe, Tracy, Cerdá, Norris, & Galea, 2013; Osenbach, Stubbs, Wang, Russo, & Zatzick, 2009). As such, it is important that efforts to screen for PTSD do so efficiently, without posing undue burden to survivors.

Several efforts have been made to date to create self-report screening measures for PTSD. According to a recent review of this literature (Spoont et al., 2015), the two most widely used self-report screeners are the Primary Care PTSD Screen for *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (*DSM–IV*) PC-PTSD;

This article was published Online First November 21, 2016.
Matthew D. Finkelman, School of Dental Medicine, Tufts University; Sarah R. Lowe, Department of Psychology, Montclair State University; Wonsuk Kim, Measured Progress, Inc., Dover, New Hampshire; Oliver Gruebner, T.H. Chan School of Public Health, Harvard University; Niels Smits, Research Institute of Child Development and Education, University of Amsterdam; Sandro Galea, School of Public Health, Boston University.

Oliver Gruebner is now at the Department of Epidemiology and Health Monitoring, Unit Social Determinants of Health, Robert Koch-Institute.

Correspondence concerning this article should be addressed to Matthew D. Finkelman, 1 Kneeland Street, Boston, MA 02111. E-mail: matthew.finkelman@tufts.edu

Prins et al., 2004), which consists of four items, and Posttraumatic Stress Checklist for *DSM–III–R* or *DSM–IV* (PCL; Blanchard, Jones-Alexander, Buckley, & Forneris, 1996; Weathers, Litz, Herman, Huska, & Keane, 1993), which consists of 17 items. Both screeners were shown to have good psychometric properties relative to the Clinician Administer PTSD Scale (CAPS; Blake et al., 1995; Weathers, Blake, et al., 2013), the gold standard interview assessment of PTSD (PC-PTSD: sensitivity = .69, specificity = .92; PCL: sensitivity = .70, specificity = .90), and the diagnostic accuracy of the two scales did not significantly differ.

The clear advantage of the PC-PTSD is its brevity and ease of administration. However, it is thought to carry the potential for misclassifying PTSD in specific subpopulations, including racial and ethnic minorities (e.g., Boscarino et al., 2012). Researchers have attempted to reduce the length of the PCL, and versions with two, four, and six items have been empirically validated (Bliese et al., 2008; Lang & Stein, 2005; Tiet, Schutte, & Leyva, 2013). Notably, however, short versions of the PCL did not cross-validate in a sample of natural disaster survivors (Hirschel & Schulenberg, 2010).

Although both the PC-PTSD and PCL show promise as screeners for PTSD in posttrauma settings, most of the published validation studies to date used the *DSM–III–R* or *DSM–IV* criteria for PTSD. The PTSD criteria were revised in the *DSM–5*, with the elimination of one item from *DSM–IV* and the addition of three novel items, among other changes. It is unclear whether the PC-PTSD is a valid screener for *DSM–5* PTSD. In addition, although the new 20-item version of the PCL, the PCL-5 (Weathers, Litz, et al., 2013), has been shown to have strong psychometric properties (Blevins, Weathers, Davis, Witte, & Domino, 2015), it is unknown whether it could be shortened while maintaining its validity.

Although it is likely that many individuals could complete the PCL-5 without undue burden or stress, certain respondents may be dissuaded by its length. Indeed, the importance of minimizing respondent burden in a general assessment context has been highlighted by the Scientific Advisory Committee of the Medical Outcomes Trust (Aaronson et al., 2002). Reducing the test length may lessen the emotional stress that results from taking an instrument (Kohout, Berkman, Evans, & Cornoni-Huntley, 1993) and may be especially beneficial to respondents who have physical illness and/or difficulty with reading comprehension (Carpenter et al., 1998).

One modern approach to alleviating the respondent burden of a questionnaire is to *customize* the questionnaire to the individual taking it. In this approach, the questionnaire is administered via a computer, which is used to track the respondent's answers as she proceeds through the test. After each item, an internal computer algorithm determines whether the administration of another item is necessary, based on the respondent's set of answers up to that point in the assessment. Because different respondents provide different answers to the various questionnaire items, the test length varies from respondent to respondent. Research has shown that such customization of assessments can improve measurement efficiency (Babcock & Weiss, 2012; Thompson, 2011).

Two techniques for customization that have been studied extensively are curtailment and stochastic curtailment (e.g., Ben-Porath, Slutske, & Butcher, 1989; Finkelman, He, Kim, & Lai, 2011; Finkelman, Smits, Kim, & Riley, 2012; Forbey, Handel, & Ben-Porath, 2000; Roper, Ben-Porath, & Butcher, 1995). These techniques can be employed when the given assessment is being used to produce a categorical (e.g., "positive or negative") result for each respondent. As will be explained, multiple such "positive or negative" decision rules have been suggested for the PCL-5, making this questionnaire a viable candidate for curtailment and stochastic curtailment. For assessments with the two categories "positive" and "negative," both curtailment and stochastic curtailment determine whether to stop testing based on the chance that the given respondent will ultimately be found to be "positive" or "negative" by the full-length questionnaire. In curtailment (which is also referred to as the countdown method; e.g., Butcher, Keller, & Bacon, 1985), the test is stopped early if the chance of one of the outcomes (either "positive" or "negative") reaches 100%, given the respondent's answers up to the current stage of testing. In stochastic curtailment, the criterion for stopping is more liberal: the test is stopped early if the chance of a "positive" result, or the chance of a "negative" result, reaches or exceeds a certain threshold that may be lower than 100% (e.g., 99%).

Previous research has indicated that curtailment and stochastic curtailment have potential to substantially reduce the respondent burden associated with questionnaires (e.g., Ben-Porath et al., 1989; Finkelman et al., 2011; Forbey et al., 2000; Roper et al., 1995). However, to our knowledge, neither of these methods has been studied alongside the PCL-5. The aim of the current research was to fill this gap by examining the extent to which these techniques can reduce the respondent burden of the PCL-5 without compromising its utility as a screener or provisional diagnostic tool.

Method

Participants and Procedures

Participants were recruited as part of a study on psychological resilience in the aftermath of Hurricane Sandy. In total, 1,000 participants were administered the survey, 500 between December 2013 and March 2014, and 500 between January 2015 and March 2015. Adults (18 years and older) were randomly selected from New York City neighborhoods that were most severely affected by the hurricane. Half of the sample was recruited via address-based sampling, and the other half through random digit dialing of cellular phones. Additional information on the sampling frame and participant recruitment can be found elsewhere (Lowe, Sampson, Gruebner, & Galea, 2015).

Trained interviewers conducted telephone surveys in English and Spanish that used a computer-assisted interview system and lasted an average of 20 min. Participants provided oral consent and were compensated with $25. The institutional review board (IRB) of Columbia University approved the study, and the IRB at Tufts Medical Center and Tufts University Health Sciences Campus granted exempt status for this analysis.

Scoring of the PCL-5

Each of the 20 items on the full-length PCL-5 asks participants how much they were bothered by a given problem related to an event (in this case, Hurricane Sandy) in the past month. All items are scored on a scale from 0 to 4 (0 = *not at all*, 1 = *a little bit*, 2 = *moderately*, 3 = *quite a bit*, 4 = *extremely*). The items are grouped into four clusters, referred to as cluster B (Intrusion; items

1–5, e.g., "repeated, disturbing, and unwanted memories of Hurricane Sandy"), cluster C (Avoidance; items 6–7, e.g., "avoiding memories, thoughts, or feelings related to Hurricane Sandy"), cluster D (Negative Alterations in Cognitions and Mood; items 8–14, e.g., "having strong negative feelings, such as fear, horror, anger, guilt, or shame"), and cluster E (Alterations in Arousal and Reactivity; items 15–20, e.g., "being 'super alert,' or watchful or on guard"; Weathers, Litz, et al., 2013).

Weathers, Litz, et al. (2013) described two ways of determining a provisional diagnosis of PTSD using the PCL-5. First, in a *total-score-based method*, symptom severity scores are computed as the sum of the 20 item scores, ranging from 0 to 80. As of January 2016, a cutpoint of 38 was recommended (pending further validation analyses), with scores meeting or exceeding the cutpoint considered "positive" and those below the cutpoint considered "negative" (Weathers, Litz, et al., 2013). Second, a *cluster-based method* has been recommended in which symptoms are considered endorsed if their respective item is rated 2 ("Moderately") or above. A "positive" result occurs if at least one cluster B symptom, one cluster C symptom, two cluster D symptoms, and two cluster E symptoms are endorsed, whereas a "negative" result occurs if these criteria are not satisfied (Weathers, Litz, et al., 2013).

Notably, other means of scoring the PCL-5 are used. For example, total symptom severity scores or symptom cluster severity scores can be useful in assessing the extent of a trauma survivor's symptomatology. However, as described in the Introduction, curtailment and stochastic curtailment are used when the outcome of a questionnaire is categorical (e.g., "positive" or "negative") and therefore are only applicable to the total-score-based and cluster-based methods of determining probable PTSD. The following two sections are devoted to explaining how each stopping rule can be applied to each of these two scoring methods.

Curtailment and Stochastic Curtailment With the Total-Score-Based Method

We begin with curtailment, motivating its use alongside a total-score-based method with a numerical example. Suppose that the PCL-5 is being used alongside a cutpoint of 38 (as preliminarily recommended by Weathers, Litz, et al. [2013] as of January 2016) to produce a "positive" or "negative" result for each respondent. Suppose further that the questionnaire is being administered to a given respondent via computer; therefore, the respondent's answers may be tracked during the assessment, and customized testing is possible. If the respondent's first 10 item scores are 4, 4, 4, 4, 4, 4, 4, 4, 4, and 2, then her cumulative score after these 10 items is 38; hence, regardless of her answers to the remaining 10 items, she will ultimately be given a "positive" result, because her cumulative score has already reached the cutpoint. In this case, administration of the remaining 10 items could not possibly change the "positive" or "negative" result, but it would potentially increase the burden placed on the respondent. A curtailment rule would recognize this fact and stop the test after 10 items in favor of a "positive" result, thus eliminating the increased respondent burden while still arriving at the same "positive or negative" result that would have been made by the full-length PCL-5.

To give a second example, suppose that a respondent's first 11 item scores are 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, and 0, so that her cumulative score after 11 items is 1. Considering that only nine items remain in the assessment, and the maximum possible score on each item is 4, the highest possible cumulative score that the respondent could reach after taking all 20 items is 37 (1 point for her observed cumulative score in the first 11 items, plus a possible 36 points over the last nine items). Thus, it has become mathematically impossible for the respondent to reach the cutpoint of 38. Given this fact, administering the last nine items could not alter the respondent's "positive" or "negative" result, but such administration could add to the respondent's level of burden. A curtailment rule would therefore stop the test after 11 items in favor of a "negative" result, which necessarily matches the result that would have been obtained if the full-length PCL-5 had been given.

To reiterate the logic of curtailment, this method stops the assessment if a respondent's result has been completely established from her prior answers. For the PCL-5 total-score-based method, unambiguous determination of the respondent's "positive" or "negative" result occurs if either (a) the respondent's cumulative score reaches or exceeds the cutpoint, or (b) the respondent's cumulative score at the given stage of testing is so small that the respondent cannot reach or exceed the cutpoint, regardless of her future answers. Denoting the respondent's cumulative score after k items as X_k, and using a cutpoint of 38, events (a) and (b) above can be summarized mathematically as $\{X_k \geq 38\}$ and $\{X_k + 4(20 - k) < 38\}$, respectively. In the former case, curtailment terminates the test with a "positive" result, whereas in the latter case, curtailment terminates the test with a "negative" result. In the formula for "negative" stopping, the number 4 appears because it is the maximum possible score for each item; the number 20 appears because it is the number of items on the full-length PCL-5.

In stochastic curtailment, the test is stopped every time that a curtailment rule calls for termination ($\{X_k \geq 38\}$ or $\{X_k + 4(20 - k) < 38\}$, in the example above), but stopping may occur in other cases as well. Specifically, the test is also stopped if the *probability* of obtaining a "positive" result from the full-length PCL-5, or the *probability* of obtaining a "negative" result from the full-length PCL-5, meets a prespecified level γ (or higher). As will be explained, these probabilities are computed based on the respondent's cumulative score at the current stage of testing, using a statistical model. If the probability of a "positive" outcome becomes $\geq \gamma$, stochastic curtailment stops in favor of a "positive" result; if the probability of a "negative" outcome becomes $\geq \gamma$, stochastic curtailment stops in favor of a "negative" result (the value of γ is set at a value higher than 50% so that these two events cannot occur simultaneously). Because stochastic curtailment's rule for stopping is more liberal than that of curtailment, the former generally achieves greater reductions in respondent burden than the latter. However, it is possible for the result of stochastic curtailment (either "positive" or "negative") not to match that of the full-length PCL-5, whereas curtailment's result is guaranteed to match that of the full-length test.

Figure 1 provides graphical depictions of the logic of stochastic curtailment. The two panels of the figure show, for two different hypothetical respondents, the probability of a "positive" result plotted against the number of items administered. After the presentation of each item, a computer algorithm updates the probability that a given respondent will receive a "positive" result from the full-length PCL-5, conditional on her answers to that point (note that the probability of a "negative" result is equal to 100%

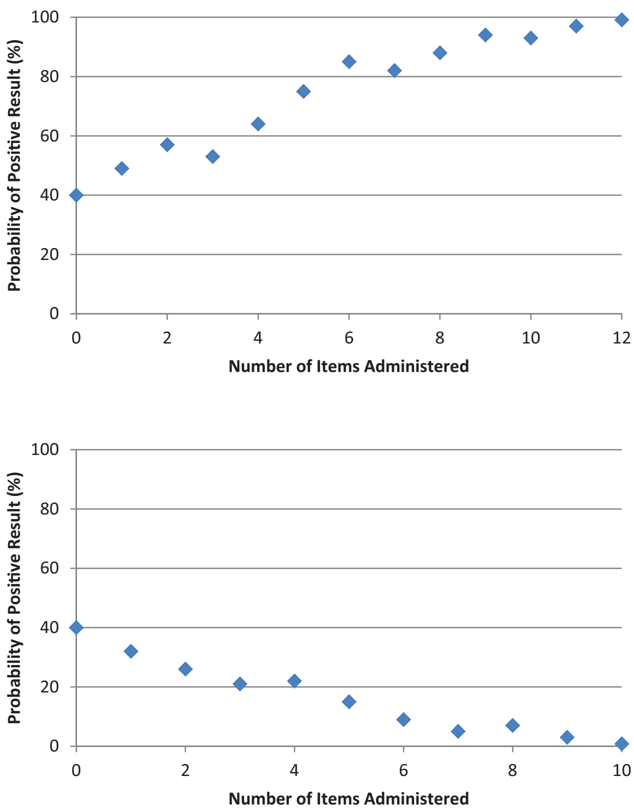

Figure 1. Schematic of stochastic curtailment (γ = 0.99). Top panel: Results for a hypothetical respondent with a "positive" result. Bottom panel: Results for a hypothetical respondent with a "negative" result. See the online article for the color version of this figure.

minus the probability of a "positive" result). If the probability of a "positive" result, or the probability of a "negative" result, becomes greater than or equal to γ after any item, the assessment is terminated; γ was set to 99% in Figure 1. For the respondent whose probabilities are depicted in the top panel of Figure 1, stopping occurred after 12 items, when her probability of a "positive" result was equal to 99.1%. For the respondent whose probabilities are depicted in the bottom panel of Figure 1, stopping occurred after 10 items, when her probability of a "negative" result was equal to 99.2%; the probability of a "positive" result, 0.8%, is depicted in the figure.

For stochastic curtailment to be used operationally, a method to obtain the probability of a "positive" result at a given stage of testing is clearly necessary. As mentioned previously, all probabilities are based on a statistical model that takes into account the respondent's cumulative score at the current stage of assessment. Specifically, at stage k of testing, the probability that a given respondent will ultimately receive a "positive" result from the full-length PCL-5 is calculated via a simple logistic regression model (Finkelman et al., 2012). The independent variable in the logistic regression model is the respondent's cumulative score after k items, and the dependent variable is the result ("positive" or "negative") from the full-length PCL-5. To fit the model, stochastic curtailment requires data from respondents who have completed the PCL-5 (curtailment, however, does not require such data). See Finkelman et al. (2012) for details about the logistic regression model.

We note that conducting a logistic regression analysis after each item would be computationally burdensome; therefore, all calculations involved in the logistic regression are performed prior to the use of stochastic curtailment in practice. In particular, the results of the logistic regression analysis can be used to determine which cumulative scores should result in stopping at each stage of testing. These scores are then listed in a simple "look-up table" that can be used in practice to govern when a new respondent's test is terminated (Finkelman et al., 2012).

Curtailment and Stochastic Curtailment With the Cluster-Based Method

When the cluster-based method is used, the logic behind curtailment and stochastic curtailment is the same as described above; however, the specifics of these techniques are adapted. In particular, the use of curtailment alongside the cluster-based method results in two different situations in which items are skipped: (a) the skipping of items *within* a cluster, and (b) the termination of the entire assessment *following* (or during) a cluster. The skipping of items within a cluster occurs if the result of that cluster has been determined for a given respondent, based on her answers in that cluster, prior to all of the items in the cluster being administered (e.g., if a cluster only requires one item to be endorsed, and the respondent endorses the first item in the cluster, the remaining items in that cluster are skipped). The termination of the entire assessment occurs if the ultimate result of the full-length PCL-5 has been unambiguously determined (e.g., if the respondent does not endorse enough items in a given cluster, the remaining clusters are skipped, because it is known that a "negative" result is given by the cluster-based method if any cluster is not adequately endorsed). Further description of curtailment for the cluster-based method is provided in Appendix 1.

The use of stochastic curtailment alongside the cluster-based method is similar to that of curtailment: stochastic curtailment terminates the test whenever curtailment does so, but it may terminate the test in other instances as well. Specifically, stochastic curtailment also calls for stopping if the probability of obtaining a "positive" provisional PTSD diagnosis (all clusters adequately endorsed), or the probability of a "negative" provisional PTSD diagnosis (at least one cluster not adequately endorsed), reaches a prespecified level γ. Again, a simple logistic regression model can be used to obtain the required probabilities. One detail of note is that the application of stochastic curtailment would become complex after a given respondent has skipped one or more questionnaire items within a cluster. Such skipping would result in missing data, which could cause issues in both statistical modeling (Little & Rubin, 2002) and the logistical administration of stochastic curtailment. In particular, if separate stopping rules were to be written for every possible combination of skip patterns, these stopping rules would quickly become unwieldy. Therefore, it may be prudent to utilize the additional stopping of stochastic curtailment only until an item is skipped, and then revert to the simple curtailment rule afterward. Although this rule may result in less early stopping, it is a conservative approach that also offers the advantage of easier practical usage. This approach was therefore used in the simulation study described below.

Data Analysis

Curtailment and stochastic curtailment were compared to the full-length PCL-5 using post hoc simulation, which is a standard approach to evaluating assessment forms retrospectively (e.g., Ben-Porath et al., 1989; Forbey et al., 2000). In post hoc simulation, existing data are analyzed to determine what the performance of each technique under study (here, curtailment and stochastic curtailment) *would have been,* if these techniques had been employed prospectively in a computer-based test. In particular, for each respondent included in the analysis, the test length that would have been observed for that respondent using curtailment and stochastic curtailment was determined. Separate analyses were conducted for the total-score-based and cluster-based methods. For the total-score-based method, a cutpoint of 38 was used. Summary statistics regarding the respondent burden of each technique (mean test length, standard deviation of test length, and percentage of tests shortened by at least one item) were calculated. Additionally, the result that would have been obtained using each technique (either "positive" or "negative") was found and compared with the result obtained by the full-length PCL-5. The sensitivity and specificity of each technique with respect to the full-length PCL-5 was calculated, as was the overall percentage of "positive or negative" results matching those of the PCL-5. Three different versions of stochastic curtailment, corresponding to different values of the parameter γ, were evaluated. The three values of γ used were 99%, 99.5%, and 99.9%; the versions of stochastic curtailment corresponding to these values will be referred to as SC_{99}, $SC_{99.5}$, and $SC_{99.9}$, respectively, in the sequel. A computer program using the statistical software package R (Version 3.1.2; R Core Team, 2015) was written to conduct the analysis.

Results

Nine hundred forty-two of the 1000 respondents (94.2%) completed the full-length PCL-5 and were included in the analysis. Five hundred eighty-nine of the 942 were female (63%). Of the 933 respondents with information on marital status, 354 were married (38%). Of the 929 respondents with information on age, the mean (*SD*) age was 50.2 (17.4). Of the 840 respondents reporting race, the majority (65.0%) identified as white, 20.2% as Black or African American, 5.6% as Asian, 6.6% as "other," and 2.6% as multiracial. Of the 929 respondents reporting ethnicity, 19.8% identified as Hispanic and 80.2% as non-Hispanic. The mean (*SD*) total score on the full-length PCL-5 among the 942 included respondents was 6.0 (10.9).

Results for the Total-Score-Based Method

Under the total-score-based method, 25 of the 942 respondents (3%) received a positive result from the full-length PCL-5 when a cutpoint of 38 was used. Table 1 presents the specific stopping rules of each technique under study (curtailment, $SC_{99.9}$, $SC_{99.5}$, and SC_{99}), arranged from the most conservative technique (curtailment) to the most liberal (SC_{99}). The stopping rules are presented as simple look-up tables showing the set of cumulative scores resulting in early stopping at each stage of testing. For example, at stage 13 of testing (i.e., after the thirteenth item has been administered), curtailment stops in favor of a negative result if the respondent's cumulative score ("CS") is ≤ 9; it stops in favor of a positive result if the respondent's cumulative score is $\geq$ 38. The stopping rules for the other techniques are as follows: for $SC_{99.9}$, ≤ 11 and $\geq$ 38; for $SC_{99.5}$, ≤ 15 and $\geq$ 37; and for SC_{99}, ≤ 16 and $\geq$ 36, as displayed in Table 1. Appendix 2

Table 1
Stopping Rules of Curtailment and Stochastic Curtailment for the Total-Score-Based Method, Using a Cut Point of 38 (N = 942)

	Curtailment		$SC_{99.9}$		$SC_{99.5}$		SC_{99}	
Stage of testing	Stop: Negative result	Stop: Positive result	Stop: Negative result	Stop: Positive result	Stop: Negative Result	Stop: Positive result	Stop: Negative result	Stop: Positive result
1	NA[a]	NA	NA	NA	CS[b] = 0	NA	CS = 0	NA
2	NA	NA	NA	NA	CS = 0	NA	CS ≤ 1	NA
3	NA	NA	NA	NA	CS = 0	NA	CS ≤ 1	NA
4	NA	NA	NA	NA	CS ≤ 1	NA	CS ≤ 3	NA
5	NA	NA	NA	NA	CS ≤ 2	NA	CS ≤ 4	NA
6	NA	NA	CS = 0	NA	CS ≤ 3	NA	CS ≤ 5	NA
7	NA	NA	CS ≤ 1	NA	CS ≤ 5	NA	CS ≤ 6	NA
8	NA	NA	CS ≤ 2	NA	CS ≤ 6	CS ≥ 32	CS ≤ 8	CS ≥ 31
9	NA	NA	CS ≤ 5	NA	CS ≤ 8	CS ≥ 33	CS ≤ 10	CS ≥ 31
10	NA	CS ≥ 38	CS ≤ 7	CS ≥ 36	CS ≤ 10	CS ≥ 33	CS ≤ 11	CS ≥ 31
11	CS ≤ 1	CS ≥ 38	CS ≤ 8	CS ≥ 38	CS ≤ 11	CS ≥ 35	CS ≤ 13	CS ≥ 33
12	CS ≤ 5	CS ≥ 38	CS ≤ 8	CS ≥ 38	CS ≤ 12	CS ≥ 37	CS ≤ 14	CS ≥ 35
13	CS ≤ 9	CS ≥ 38	CS ≤ 11	CS ≥ 38	CS ≤ 15	CS ≥ 37	CS ≤ 16	CS ≥ 36
14	CS ≤ 13	CS ≥ 38	CS ≤ 14	CS ≥ 38	CS ≤ 17	CS ≥ 37	CS ≤ 18	CS ≥ 36
15	CS ≤ 17	CS ≥ 38	CS ≤ 17	CS ≥ 38	CS ≤ 19	CS ≥ 37	CS ≤ 20	CS ≥ 36
16	CS ≤ 21	CS ≥ 38	CS ≤ 21	CS ≥ 38	CS ≤ 21	CS ≥ 37	CS ≤ 22	CS ≥ 36
17	CS ≤ 25	CS ≥ 38	CS ≤ 25	CS ≥ 38	CS ≤ 25	CS ≥ 38	CS ≤ 25	CS ≥ 37
18	CS ≤ 29	CS ≥ 38	CS ≤ 29	CS ≥ 38	CS ≤ 29	CS ≥ 38	CS ≤ 29	CS ≥ 38
19	CS ≤ 33	CS ≥ 38	CS ≤ 33	CS ≥ 38	CS ≤ 33	CS ≥ 38	CS ≤ 33	CS ≥ 38
20	CS ≤ 37	CS ≥ 38	CS ≤ 37	CS ≥ 38	CS ≤ 37	CS ≥ 38	CS ≤ 37	CS ≥ 38

[a] NA = Not applicable (no early stopping can occur). [b] CS = Cumulative score.

provides a numerical example of how the stopping rules of Table 1 were derived for the stochastic curtailment techniques.

The top panel of Table 2 presents statistics regarding each technique's level of concordance with the full-length PCL-5's result ("positive" or "negative") under the total-score-based method as well as statistics regarding the number of items administered by each technique. By definition, curtailment always matches the result of the full-length assessment for every respondent; therefore, its sensitivity for the full-length PCL-5, specificity for the full-length PCL-5, and overall percentage of results matching the full-length PCL-5 were 100%. Moreover, for 99.5% of respondents, curtailment resulted in a reduction of the number of items administered; its mean (*SD*) number of items administered was 11.9 (1.7). Turning to $SC_{99.9}$, this technique is not guaranteed to be perfectly concordant with the result of the full-length PCL-5 in every dataset, but the two were perfectly concordant in the current dataset of 942 respondents. Additionally, $SC_{99.9}$ made further improvements in test length compared to curtailment: $SC_{99.9}$'s mean (*SD*) number of items administered was 7.9 (3.2), with the percentage of respondents with shortened tests remaining at 99.5%. The two most liberal techniques, $SC_{99.5}$ and SC_{99}, did not always match the result of the full-length PCL-5: $SC_{99.5}$'s sensitivity, specificity, and overall percentage of results matching the full-length PCL-5 were 96%, 100%, and 99.9%, respectively, whereas these numbers were 92%, 100%, and 99.8% for SC_{99}. These two techniques exhibited the greatest reduction in the number of items administered, with a mean (*SD*) test length of 3.1 (4.3) items for $SC_{99.5}$ and 2.4 (3.6) for SC_{99}. Each of the latter two techniques shortened the test for at least 99.6% of respondents.

Results for the Cluster-Based Method

Thirty-seven of the 942 respondents (4%) received a positive result from the full-length PCL-5 under the cluster-based method. The bottom panel of Table 2 shows the results of each stopping technique under this scoring method. Curtailment's sensitivity for the full-length PCL-5, specificity for the full-length PCL-5, and overall percentage of results matching the full-length PCL-5 were 100%, as guaranteed by this technique. Its mean number of items administered was 5.3 ($SD = 1.6$), and it shortened the test for 100% of respondents. The $SC_{99.9}$ and $SC_{99.5}$ stopping rules were identical to curtailment under the cluster-based method (i.e., the stochastic curtailment component of these techniques did not call for additional early stopping beyond what was already provided by curtailment); hence, the statistics presented in the bottom panel of Table 2 for $SC_{99.9}$ and $SC_{99.5}$ are the same as those presented for curtailment. For SC_{99}, the rule for stopping was identical to curtailment, with the exception that SC_{99} stopped early for a "negative" result under three additional conditions: (a) the respondent's cumulative score after two items was not more than 0; (b) the respondent's cumulative score after four items was not more than 1; and (c) the respondent's score after 19 items was not more than 10. Because of these additional stopping conditions, the results for SC_{99} were different from those of curtailment, $SC_{99.9}$, and $SC_{99.5}$. In particular, SC_{99} exhibited a sensitivity of 94.6% for the full-length PCL-5, a specificity of 100% for the full-length PCL-5, and an overall percentage of results matching the full-length PCL-5 of 99.8%. The additional stopping conditions of the SC_{99} also resulted in this technique having the smallest mean number of items administered (3.1), along with a standard deviation of 2.3 items. It shortened the test for 100% of respondents.

Discussion

The administration of assessments via computer has been a prominent research topic in the health literature over the last decade (e.g., Cella et al., 2007; Fries, Cella, Rose, Krishnan, & Bruce, 2009; Hung et al., 2014). One advantage of computer-based testing is that it has the potential to enhance the efficiency of an instrument. Curtailment and stochastic curtailment are two techniques that have been studied as a means to improve the efficiency of other assessments (Ben-Porath et al., 1989; Finkelman et al., 2011, 2012; Forbey et al., 2000; Roper et al., 1995); to our knowledge, however, no previous research has explored the application of these techniques to the PCL-5.

Results of the study suggest that both curtailment and stochastic curtailment have potential to maintain the same "positive or negative" result as the full-length version of the PCL-5 in all (or nearly all) cases, while substantially reducing its respondent burden. By definition, curtailment matched the full-length PCL-5's result in 100% of cases; it also shortened the test in 99.5% of cases under the total-score-based method and 100% of cases under the cluster-

Table 2
Performance of Curtailment and Stochastic Curtailment (N = 942)

Method	Concordance with Full-Length PCL-5			Number of Items Administered		
	Sensitivity	Specificity	Overall % matching	Mean test length	*SD* of test length	% of tests shortened
Total-Score-Based Method						
Curtailment	100.0	100.0	100.0	11.9	1.7	99.5
$SC_{99.9}$	100.0	100.0	100.0	7.9	3.2	99.5
$SC_{99.5}$	96.0	100.0	99.9	3.1	4.3	99.6
SC_{99}	92.0	100.0	99.8	2.4	3.6	99.8
Cluster-Based Method						
Curtailment	100.0	100.0	100.0	5.3	1.6	100.0
$SC_{99.9}$	100.0	100.0	100.0	5.3	1.6	100.0
$SC_{99.5}$	100.0	100.0	100.0	5.3	1.6	100.0
SC_{99}	94.6	100.0	99.8	3.1	2.3	100.0

based method. Moreover, it reduced the mean test length by 40% under the former scoring method and by over 70% under the latter. Stochastic curtailment made further improvements in lessening the respondent burden (reducing the mean test length by up to 88% under the total-score-based method and up to 84% under the cluster-based method) while matching the result of the full-length PCL-5 in 99.8% of cases or higher. Such reductions in test length may benefit not only respondents, but also providers, given the need for efficiency in health care delivery (Dugdale, Epstein, & Pantilat, 1999).

As mentioned previously, curtailment and stochastic curtailment can only be used when a categorical result is produced by the assessment in question, as opposed to a continuous score. Although this aspect of the stopping techniques means that they can only be applied in certain administrations of the PCL-5, it is nevertheless worthwhile to study their potential to enhance the efficiency of those administrations. Extensions to the case of continuous PCL-5 scoring methods (total symptom severity scores or symptom cluster severity scores) could prove to be fruitful.

One limitation of the current study is that no "gold standard" diagnosis of PTSD was available in our dataset, such as a diagnosis based on the CAPS. However, the purpose of the study was not to examine the concordance between the PCL-5 and a gold standard, but to compare curtailment and stochastic curtailment to the full-length PCL-5. Given that curtailment is guaranteed to provide the same result ("positive" or "negative") as the full-length PCL-5, the sensitivity and specificity of curtailment for predicting any gold standard measure are necessarily identical to the sensitivity and specificity of the full-length PCL-5. The high degree of concordance between stochastic curtailment and the full-length PCL-5 implies that this stopping technique also has similar sensitivity and specificity to the full-length assessment, while garnering the potential to reduce respondent burden considerably.

Another limitation is that the results were obtained from one dataset and cannot be assumed to generalize to other contexts. In particular, the respondents were limited to one geographic location, completed the PCL-5 in reference to a particular event (Hurricane Sandy), and did so over a year after that event took place. Results found retrospectively based on telephone interviews may also not generalize to settings in which the screener is given prospectively and via computer. Furthermore, although the curtailment stopping rules defined in Table 1 are applicable to any population for which a cutpoint of 38 is appropriate, the stopping rules of stochastic curtailment are population-specific and must therefore be defined based on data from the same population in which the stopping technique will be used. Finally, the large reduction in mean test length provided by both curtailment and stochastic curtailment might not be found among populations with a higher percentage of individuals with scores closer to the cutpoint. Nevertheless, the results indicate the existence of populations for which the stopping rules may prove to be of value.

Given the promising results obtained in this retrospective study, the next step is to test the curtailment and stochastic curtailment stopping rules prospectively. Because computerized testing is now commonplace in the field of assessment, the development of a computer-based version of the PCL-5 alongside a simple look-up table for stopping is relatively straightforward. This operational computer-based version can then be used to investigate the improvement in respondent burden provided via curtailment and stochastic curtailment in different populations. Such future work will be valuable in promoting the efficient assessment of PTSD.

References

Aaronson, N., Alonso, J., Burnam, A., Lohr, K. N., Patrick, D. L., Perrin, E., . . . the Scientific Advisory Committee of the Medical Outcomes Trust. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 11,* 193–205. http://dx.doi.org/10.1023/A:1015291021312

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Babcock, B., & Weiss, D. J. (2012). Termination criteria in computerized adaptive testing: Do variable-length CATs provide efficient and effective measurement? *Journal of Computerized Adaptive Testing, 1,* 1–18. http://dx.doi.org/10.7333/jcat.v1i1.16

Ben-Porath, Y. S., Slutske, W. S., & Butcher, J. N. (1989). A real-data simulation of computerized adaptive administration of the MMPI. *Psychological Assessment, 1,* 18–22. http://dx.doi.org/10.1037/1040-3590.1.1.18

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a clinician-administered PTSD scale. *Journal of Traumatic Stress, 8,* 75–90. http://dx.doi.org/10.1002/jts.2490080106

Blanchard, E. B., Jones-Alexander, J., Buckley, T. C., & Forneris, C. A. (1996). Psychometric properties of the PTSD Checklist (PCL). *Behaviour Research and Therapy, 34,* 669–673. http://dx.doi.org/10.1016/0005-7967(96)00033-2

Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The posttraumatic stress disorder checklist for *DSM–5* (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress, 28,* 489–498. http://dx.doi.org/10.1002/jts.22059

Bliese, P. D., Wright, K. M., Adler, A. B., Cabrera, O., Castro, C. A., & Hoge, C. W. (2008). Validating the primary care posttraumatic stress disorder screen and the posttraumatic stress disorder checklist with soldiers returning from combat. *Journal of Consulting and Clinical Psychology, 76,* 272–281. http://dx.doi.org/10.1037/0022-006X.76.2.272

Boscarino, J. A., Kirchner, H. L., Hoffman, S. N., Sartorius, J., Adams, R. E., & Figley, C. R. (2012). The New York PTSD risk score for assessment of psychological trauma: Male and female versions. *Psychiatry Research, 200,* 827–834. http://dx.doi.org/10.1016/j.psychres.2012.04.022

Butcher, J. N., Keller, L. S., & Bacon, S. F. (1985). Current developments and future directions in computerized personality assessment. *Journal of Consulting and Clinical Psychology, 53,* 803–815. http://dx.doi.org/10.1037/0022-006X.53.6.803

Carpenter, J. S., Andrykowski, M. A., Wilson, J., Hall, L. A., Rayens, M. K., Sachs, B., & Cunningham, L. L. (1998). Psychometrics for two short forms of the Center for Epidemiologic Studies-Depression Scale. *Issues in Mental Health Nursing, 19,* 481–494. http://dx.doi.org/10.1080/016128498248917

Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., . . . the PROMIS Cooperative Group. (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap cooperative group during its first two years. *Medical Care, 45,* S3–S11. http://dx.doi.org/10.1097/01.mlr.0000258615.42478.55

Dickstein, B. D., Suvak, M., Litz, B. T., & Adler, A. B. (2010). Heterogeneity in the course of posttraumatic stress disorder: Trajectories of symptomatology. *Journal of Traumatic Stress, 23,* 331–339.

Dugdale, D. C., Epstein, R., & Pantilat, S. Z. (1999). Time and the patient-physician relationship. *Journal of General Internal Medicine, 14,* S34–S40. http://dx.doi.org/10.1046/j.1525-1497.1999.00263.x

Finkelman, M. D., He, Y., Kim, W., & Lai, A. M. (2011). Stochastic curtailment of health questionnaires: A method to reduce respondent burden. *Statistics in Medicine, 30,* 1989–2004. http://dx.doi.org/10.1002/sim.4231

Finkelman, M. D., Smits, N., Kim, W., & Riley, B. (2012). Curtailment and stochastic curtailment to shorten the CES-D. *Applied Psychological Measurement, 36,* 632–658. http://dx.doi.org/10.1177/0146621612451647

Forbey, J. D., Handel, R. W., & Ben-Porath, Y. S. (2000). A real data simulation of computerized adaptive administration of the MMPI-A. *Computers in Human Behavior, 16,* 83–96. http://dx.doi.org/10.1016/S0747-5632(99)00053-9

Fries, J. F., Cella, D., Rose, M., Krishnan, E., & Bruce, B. (2009). Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. *The Journal of Rheumatology, 36,* 2061–2066. http://dx.doi.org/10.3899/jrheum.090358

Hirschel, M. J., & Schulenberg, S. E. (2010). On the viability of PTSD Checklist (PCL) short form use: Analyses from Mississippi Gulf Coast Hurricane Katrina survivors. *Psychological Assessment, 22,* 460–464. http://dx.doi.org/10.1037/a0018336

Hung, M., Franklin, J. D., Hon, S. D., Cheng, C., Conrad, J., & Saltzman, C. L. (2014). Time for a paradigm shift with computerized adaptive testing of general physical function outcomes measurements. *Foot & Ankle International, 35,* 1–7. http://dx.doi.org/10.1177/1071100713507905

Kearns, M. C., Ressler, K. J., Zatzick, D., & Rothbaum, B. O. (2012). Early interventions for PTSD: A review. *Depression and Anxiety, 29,* 833–842. http://dx.doi.org/10.1002/da.21997

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry, 52,* 1048–1060. http://dx.doi.org/10.1001/archpsyc.1995.03950240066012

Kohout, F. J., Berkman, L. F., Evans, D. A., & Cornoni-Huntley, J. (1993). Two shorter forms of the CES-D (Center for Epidemiological Studies Depression) depression symptoms index. *Journal of Aging and Health, 5,* 179–193. http://dx.doi.org/10.1177/089826439300500202

Lang, A. J., & Stein, M. B. (2005). An abbreviated PTSD checklist for use as a screening instrument in primary care. *Behaviour Research and Therapy, 43,* 585–594. http://dx.doi.org/10.1016/j.brat.2004.04.005

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/9781119013563

Lowe, S. R., Sampson, L., Gruebner, O., & Galea, S. (2015). Psychological resilience after Hurricane Sandy: The influence of individual- and community-level factors on mental health after a large-scale natural disaster. *PLoS ONE, 10,* e0125761. http://dx.doi.org/10.1371/journal.pone.0125761

Lowe, S. R., Tracy, M., Cerdá, M., Norris, F. H., & Galea, S. (2013). Immediate and longer-term stressors and the mental health of Hurricane Ike survivors. *Journal of Traumatic Stress, 26,* 753–761. http://dx.doi.org/10.1002/jts.21872

Osenbach, J. E., Stubbs, J., Wang, J., Russo, J., & Zatzick, D. (2009). Legal events as predictors of posttraumatic stress in injured trauma survivors. *Psychiatry: Interpersonal and Biological Processes, 72,* 70–78. http://dx.doi.org/10.1521/psyc.2009.72.1.70

Prins, A., Ouitmette, P., Kimerling, R., Cameron, R. P., Hugelshofer, D. S., Shaw-Hegwer, J., . . . Sheikh, J. I (2004). The primary care PTSD screen (PC-PTSD): Development and operating characteristics. *Primary Care Psychiatry, 9,* 9–14. http://dx.doi.org/10.1185/135525703125002360

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Roper, B. L., Ben-Porath, Y. S., & Butcher, J. N. (1995). Comparability and validity of computerized adaptive testing with the MMPI-2. *Journal of Personality Assessment, 65,* 358–371. http://dx.doi.org/10.1207/s15327752jpa6502_10

Sareen, J., Cox, B. J., Stein, M. B., Afifi, T. O., Fleet, C., & Asmundson, G. J. (2007). Physical and mental comorbidity, disability, and suicidal behavior associated with posttraumatic stress disorder in a large community sample. *Psychosomatic Medicine, 69,* 242–248. http://dx.doi.org/10.1097/PSY.0b013e31803146d8

Spoont, M. R., Williams, J. W., Jr., Kehle-Forbes, S., Nieuwsma, J. A., Mann-Wrobel, M. C., & Gross, R. (2015). Does this patient have posttraumatic stress disorder? Rational clinical examination systematic review. *JAMA: Journal of the American Medical Association, 314,* 501–510. http://dx.doi.org/10.1001/jama.2015.7877

Thompson, N. A. (2011). Termination criteria for computerized classification testing. *Practical Assessment, Research & Evaluation, 16*. Retrieved from http://pareonline.net/pdf/v16n4.pdf

Tiet, Q. Q., Schutte, K. K., & Leyva, Y. E. (2013). Diagnostic accuracy of brief PTSD screening instruments in military veterans. *Journal of Substance Abuse Treatment, 45,* 134–142. http://dx.doi.org/10.1016/j.jsat.2013.01.010

Weathers, F. W., Blake, D. D., Schnurr, P. P., Kaloupek, D. G., Marx, B. P., & Keane, T. M. (2013). *The Clinician-Administered PTSD Scale for DSM–5 (CAPS-5)*. Interview available from the National Center for PTSD at www.ptsd.va.gov

Weathers, F., Litz, B., Herman, D., Huska, J., & Keane, T. (1993). The PTSD checklist (PCL): Reliability, validity, and diagnostic utility. Paper presented at the Annual Convention of the International Society for Traumatic Stress Studies, San Antonio, TX.

Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2013). The PTSD Checklist for *DSM–5* (PCL-5). Scale available from the National Center for PTSD at www.ptsd.va.gov

Appendix 1

Description of Curtailment with the Cluster-Based Method of Provisional PTSD Diagnosis

Consider a respondent who has just begun Cluster B (Items 1-5), and suppose that the respondent answers "Not at all" (score = 0) to Item 1 and "Moderately" (score = 2) to Item 2. Because the cluster-based method prescribes that a response of "Moderately" (or above) constitutes endorsement of any given item, the respondent has now endorsed an item (namely, Item 2) within Cluster B. Moreover, because a provisional PTSD diagnosis only requires one item from Cluster B to be endorsed, it is not necessary to present the remainder of the items in Cluster B (since the result of this cluster has been unambiguously determined to be "adequate endorsement" following the endorsement of Item 2). Therefore, a curtailment stopping rule would skip the respondent past the remaining items in Cluster B (namely, Items 3, 4, and 5), and move the respondent directly to Item 6 (the first item of Cluster C). Next, consider a different respondent who has also just begun to take the PCL-5. Suppose that this second respondent answers either "Not at all" (score 0) or "A little bit" (score 1) to all five items in Cluster B. Because the respondent has not endorsed any item in Cluster B (no items answered "Moderately" or above), it has become impossible for her to receive a provisional PTSD diagnosis under the cluster-based method, regardless of her answers to the remaining PCL-5 items. Therefore, a curtailment rule would stop the test in favor of a "negative" result after five items.

Figure A1 shows that the above logic, whereby items may be skipped within Cluster B or the assessment may be terminated based on responses to Cluster B, applies to other clusters as well. In Cluster C (which only contains Items 6 and 7), a respondent's endorsement of Item 6 constitutes adequate endorsement of the cluster, which implies that Item 7 may be skipped. If neither Item 6 nor Item 7 is endorsed, then Cluster C does not exhibit adequate endorsement, so the test is terminated after seven items in favor of a "negative" result. In Cluster D (Items 8-14), *two* items must be endorsed in order for the cluster to be considered "adequately endorsed" by the cluster-based method. Thus, as soon as *two* Cluster D items have been endorsed, the remaining Cluster D items may be skipped. Additionally, if no item between Item 8 and Item 13 is endorsed, then it becomes impossible for two Cluster D items to be endorsed (since only one item, Item 14, remains in the cluster); hence, not only can Item 14 be skipped, but also the entire test may be terminated in favor of a "negative" result. If the respondent is administered all Cluster D items, and fewer than two are endorsed, the test may similarly be terminated in favor of a "negative" result following Item 14 (the end of Cluster D). Finally, Cluster E (Items 15-20) also requires the endorsement of two items. Analogous to the process for Cluster D, as soon as two Cluster E items have been endorsed, the remaining items from that cluster can be skipped. Note that under a curtailment rule, a respondent can only enter Cluster E if she has already exhibited adequate endorsement of Clusters B, C, and D; therefore, if Cluster E is stopped early due to the endorsement of two items, then all four clusters have been adequately endorsed, and a "positive" result is called for. If no item between Item 15 and Item 19 is endorsed, then it becomes impossible for two Cluster E items to be endorsed (since only Item 20 remains in the cluster); hence, the test may be terminated after Item 19 in favor of a "negative" result. If the test proceeds all the way to Item 20, then the usual rule for the cluster-based method is applied: a "positive" result is obtained if all four clusters have been adequately endorsed, and a "negative" result is obtained otherwise. Because curtailment prescribes that a respondent can only reach Cluster E by adequately endorsing Clusters B-D, a respondent who reaches Cluster E will receive a "positive" result from the PCL-5 if and only if she endorses two items from that cluster.

(*Appendices continue*)

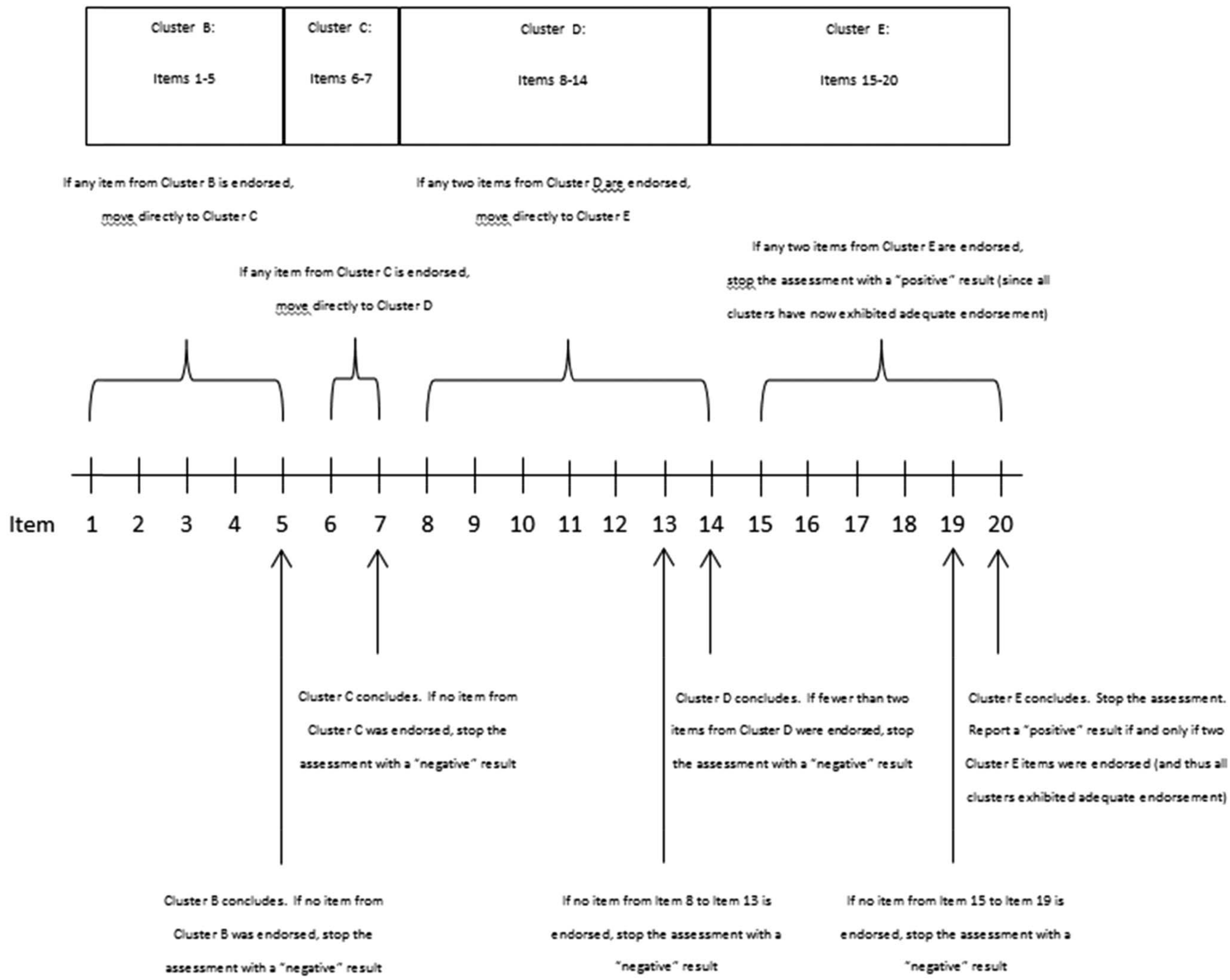

Figure A1. Schematic of Curtailment for the Cluster-Based Method.

(*Appendices continue*)

Appendix 2

Numerical Example of the Derivation of Stochastic Curtailment's Stopping Rules under the Total-Score-Based Method

Table A1
Estimated Probabilities of a "Positive" Result, and Stochastic Curtailment's Stopping Rules, after Two Items (Total-Score-Based Method)

Score	Estimated Probability of "Positive" Result	Stopping Rule: $SC_{99.9}$	Stopping Rule: $SC_{99.5}$	Stopping Rule: SC_{99}
0	0.4%	Continue Testing	Stop: Negative Result	Stop: Negative Result
1	0.9%	Continue Testing	Continue Testing	Stop: Negative Result
2	2.3%	Continue Testing	Continue Testing	Continue Testing
3	5.8%	Continue Testing	Continue Testing	Continue Testing
4	13.7%	Continue Testing	Continue Testing	Continue Testing
5	29.0%	Continue Testing	Continue Testing	Continue Testing
6	51.4%	Continue Testing	Continue Testing	Continue Testing
7	73.2%	Continue Testing	Continue Testing	Continue Testing
8	87.6%	Continue Testing	Continue Testing	Continue Testing

To derive the stopping rules of stochastic curtailment under the total-score-based method and a cut point of 38, a simple logistic regression model was constructed for every stage of testing. The outcome (dependent variable) of each logistic regression was the binary result of the full-length PCL-5 (coded "negative" = 0 and "positive" = 1); the variable predicting this outcome was the cumulative score at the given stage of testing. Table A1 provides information about the stochastic curtailment stopping rules corresponding to the case where two items have been presented. At this stage of testing, the lowest possible cumulative score is 0 and the highest possible cumulative score is 8. The estimated intercept of the logistic regression model was −5.633, and the estimated slope was 0.948. Based on these parameters, after two items the estimated probability of a "positive" result was 0.4% for a respondent with a cumulative score of 0, 0.9% for a respondent with a cumulative score of 1, and so forth, up to an estimated probability of 87.6% for a respondent with a cumulative score of 8 (Table A1). These estimated probabilities were examined in order to define the stopping rules $SC_{99.9}$, $SC_{99.5}$, and SC_{99}. For instance, $SC_{99.9}$ only terminates testing if an estimated probability falls below 0.1% ("negative" result) or above 99.9% ("positive" result); as none of the estimated probabilities in Table A1 satisfies either condition, it was determined that $SC_{99.9}$ "continue testing" after two items for any possible cumulative score 0-8. $SC_{99.5}$ only terminates testing if an estimated probability falls below 0.5% ("negative" result) or above 99.5% ("positive" result). The former condition is satisfied for a cumulative score of 0 (estimated probability = 0.4%); therefore, $SC_{99.5}$ stops early after two items, in favor of a "negative" result, for respondents with a cumulative score of 0. Finally, SC_{99} only terminates testing if an estimated probability falls below 1% ("negative" result) or above 99% ("positive" result). The former condition is satisfied for a cumulative score of 0 (estimated probability = 0.4%) and for a cumulative score of 1 (estimated probability = 0.9%); therefore, SC_{99} stops early after two items, in favor of a "negative" result, for respondents with either a cumulative score of 0 or a cumulative score of 1. In no case was the estimated probability high enough that any of the stochastic curtailment techniques would stop in favor of a "positive" result after two items. The stopping rules derived herein are consistent with those presented in the second row of Table 1.

Received February 20, 2016
Revision received October 1, 2016
Accepted October 11, 2016 ■