

Mental Health Assessments in ICU and Acute Care
Megan Altom MOT/S
Taylor Harrell MOT/S
Carolynn Keane MOT/S
Erica Smith MOT/S
Danielle Woodward MOT/S
March 2020

Table 10-4 Critically Appraised Topic Template

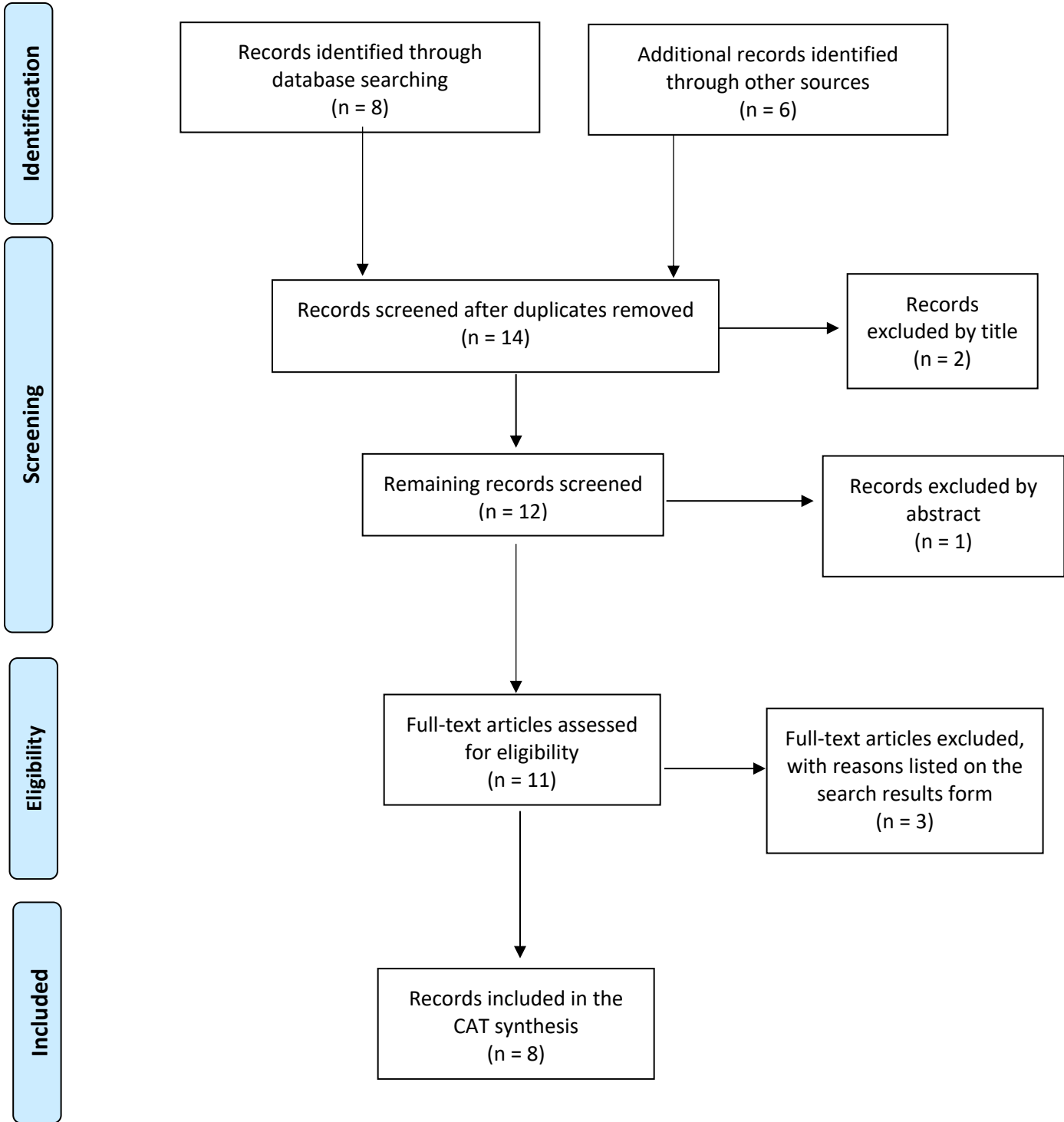
<p>Evidence-Based Practice Question In adults with burn injuries in ICU and acute care settings, which mental health assessments are effective for assessing quality of life during initial evaluation?</p>		
<p>Clinical Scenario Sandy Fletchall, an Occupational therapist at the Firefighters Burn Center-Regional One Health, is looking for assessments in the ICU and acute care settings that are focused on emotional states and quality of life, since current assessments are cognitive-based (i.e. SLUMS, MoCA, and the MMSE). Assessments need to be short, easy to administer assessments that are not expensive or are freely available. Evidence-based assessments are important for accurate identification of emotional and trauma related mental health impairments that can be used to gather baseline data for patients in the ICU and acute care settings.</p>		
<p>Databases and Sources Searched -Google Scholar, Cochrane, PubMed, MEDLINE, OVID, CINAHL, Veterans Affairs website (www.va.gov)</p>	<p>Search Terms -Mental health assessments, Adult with burn injuries, Quality of life, Psychometrics and mental health assessments, Trauma, PTSD</p>	<p>Limits Used -English language, Full-text articles only, Age (18+)</p>
<p>PICO Question Categories</p>		<p>Search Terms Used</p>
<p><u>Population:</u> Adults with burn injuries in ICU and acute care settings</p>		
<p><u>Outcomes:</u> Quality of life, baseline mental health assessment</p>		
<p>Inclusion Criteria for Articles</p> <ul style="list-style-type: none"> ● Full text articles ● English language ● Adults or older adults ● Assessment- short, easy-to-administer 		
<p>Exclusion Criteria for Articles</p> <ul style="list-style-type: none"> ● Pediatric clients ● Adolescents ● Studies occurring more than 20 years ago 		
<p>Review Process Our facility needs valid and quick assessments that they can use upon admission to the burn unit that can help guide in interventions. Abstracts and full-text articles were scanned to determine if they were applicable to our PICO question, useful in the ICU, and relevant to the facilities needs. Assessments did not need to be specifically for burns but more so focused on mental health in the ICU. Data extraction was performed by each member of the team to analyze the articles collected during the initial data search. Mentor and professor review of analyses. Articles were appraised for quality using particular forms from Law & MacDermid, 2014).</p>		

Search Results by Level of Evidence

Level of Evidence	Study Design	# of Articles Included
1	Systematic Review	2
3	Longitudinal study	1
4	Comparative Analysis (1), Correlational Analysis (2), Methodological (2)	5



Modified PRISMA 2009 Flow Diagram (awm 2018)



Article	Level of Evidence	Quality Score	Name of Assessment	Purpose of Assessment	Reliability	Validity	Limitations
Wei, L. A., Fearing, M. A., Sternberg, E. J., & Inouye, S. K. (2008). The Confusion Assessment Method: A Systematic Review of Current Usage. <i>Journal of the American Geriatrics Society</i> , 56(5), 823–830. Doi:10.1111/j.1532-5415.2008.01674.x	I	88.9%	The Confusion Assessment Method	Assesses presence, severity, and fluctuation of 9 delirium features: acute onset, inattention, disorganized thinking, altered level of consciousness, disorientation, memory impairment, perceptual disturbances, psychomotor agitation or retardation, and altered sleep-wake cycle	-7 studies- sensitivity rate of 94% (95% CI 91-97%) -combined sensitivity rate of 89% (95% CI 85-94%)	-N/A	-Future work needed to validate summary scores or adaptations to measure delirium severity
Wang, Y. P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: a comprehensive review. <i>Brazilian Journal of Psychiatry</i> , 35(4), 416-431.	I	92%	Beck Depression Inventory-II	BDI-II is a self-report measure of depression in a variety of settings and populations.	- internal consistency around 0.9. -retest reliability ranged from 0.73 to 0.96	-criterion-based validity showed good sensitivity and specificity for detecting depression in comparison to the adopted gold standard. -factor analysis showed a robust dimension of general depression composed by two constructs: cognitive-affective and somatic-vegetative.	-Spectrum bias -Self-report
Augustin, M., Conde Montero, E., Zander, N., Baade, K., Herberger, K., Debus, E. S., ...Blome, C. (2017, September). Validity and feasibility of the wound-QoL questionnaire on health-related quality of life in chronic wounds.	III	62.5%	Wound-QoL	Wound-QoL (Quality of Life) is a patient-reported assessment used to determine the quality of life of a patient who is experiencing a chronic wound.	N/A	-high convergent validity, -high construct validity, -high longitudinal validity,	-patient-reported
Amoyal, N. R., Mason, S. T., Gould, N. F., Corry, N., Mahfouz, S., Barkey, A., & Fauerbach, J. A. (2011). Measuring coping behavior in patients with major burn injuries: a psychometric evaluation of the BCOPE. <i>Journal of Burn Care & Research</i> , 32(3), 392-398.	IV	87.5%	BCOPE	Evaluate the BCOPE in measuring coping behaviors, to establish factors underlying coping behaviors in the trauma population and evaluate construct validity.	-Item scale correlation (r = .23 to .76) -Test-retest reliability (baseline to 6 months) ranged from r = .352 -Self-distraction to r = .855 for humor.	-Exploratory factor analysis yielded two factors: active coping and avoidance coping.	-potential differences in coping strategies used in a controlled environment vs. natural setting.
Mouthaan, J., Sijbrandij, M., Reitsma, J. B., Gersons, B. P. R., & Olf, M. (2014). Comparing Screening Instruments to Predict Posttraumatic Stress Disorder. <i>PLoS ONE</i> , 9(5). Doi: 10.1371/journal.pone.0097183	IV	91%	SPAN, TSQ, IES-R	SPAN - Assesses frequency and severity of PTSD symptoms over the past week TSQ - Assesses the presence of 5 intrusion items and 5 hyperarousal items over the past week ISE-R - Measures intrusion, avoidance, and hyperarousal	-SPAN - Specificity of 0.64 -TSQ - Specificity of 0.59 -IES-R - Specificity of 0.72 -SPAN (ICC = 0.98) -TSQ (ICC = 0.82) -IES-R (ICC = 0.83)	N/A	-Screening instruments were administered at around 3 weeks following injury, which limited generalizability towards the immediate aftermath of traumatic events -Researchers were unable to collect a clinical PTSD diagnosis of 211 patients at 6 months, whose results could not be included in the index tests.

<p>Aaron, L. A., Patterson, D. R., Finch, C. P., Carrougher, G. J., & Heimbach, D. M. (2001). The utility of a burn specific measure of pain anxiety to prospectively predict pain and function: a comparative analysis. <i>Burns</i>, 27(4), 329-334.</p>	IV	68%	BSPAS	Designed to assess burn-related anxiety	-Internal (α -coefficient = 0.90)	-High predictive validity in predicting decreased physical role functioning at time of discharge	-need a larger sample of burn patients and the long-term consequences of reduced physical functioning at discharge should be evaluated both in terms of future function, and psychological adjustment following discharge from hospital
<p>Guenther, U., Popp, J., Koecher, L., Muders, T., Wrigge, H., Ely, E. W., & Putensen, C. (2010). Validity and reliability of the CAM-ICU Flowsheet to diagnose delirium in surgical ICU patients. <i>Journal of critical care</i>, 25(1), 144-151.</p>	IV	73%	CAM-ICU Flowsheet	The CAM-ICU Flowsheet is a practical, time-sparing algorithm to assess delirium criteria in intubated patients.	-Very high interrater reliability (κ , 0.96; 0.87-1.00).	-investigators had sensitivities of 88% (95% confidence interval, 69%-98%) and 92% (74%-99%); and specificities of 100% (85%-100%).	-has not been validated as severity scales to distinguish between high and low severity states -included a modest number of intubated patients.
<p>Wu, C.-Y., Lee, M.-B., Lin, Y.-Y., & Liao, S.-C. (2019). Development and validation of the 9-item Concise Mental Health Checklist (CMHC-9) for suicide risk assessment. <i>Journal of the Formosan Medical Association</i>, 118(9), 1308–1316. doi: 10.1016/j.jfma.2019.05.025</p>	IV;	68%	CMHC-9	Suicide risk assessment with a three factor structure of demographics, psychopathology and suicidality	-CMHC-9 with AUCs (recent SI 92.9%: 95% CI= 91.5-94.3; lifetime SI 75.9%: 95% CI= 73.9-77.9)	-Cronbach's α values for total sample: (alpha= 0.79); three sub-samples: (alpha= 0.76 psychiatric; 0.67 community subjects; and 0.69 for medical outpatient subjects)	-use of suicidal ideation rather than other risk factors due to study duration and low prevalence of completed suicide

Bottom Line and Recommendations: There is limited research available specifically for mental health assessments in burn ICU. However, many of the assessments we found, look at aspects that are relevant to the burn population. These assessments can be modified to fit the needs of each client seen in the burn ICU.

REFERENCES

- Aaron, L. A., Patterson, D. R., Finch, C. P., Carrougher, G. J., & Heimbach, D. M. (2001). The utility of a burn specific measure of pain anxiety to prospectively predict pain and function: a comparative analysis. *Burns*, 27(4), 329-334.
- Amoyal, N. R., Mason, S. T., Gould, N. F., Corry, N., Mahfouz, S., Barkey, A., & Fauerbach, J. A. (2011). Measuring coping behavior in patients with major burn injuries: a psychometric evaluation of the BCOPE. *Journal of Burn Care & Research*, 32(3), 392-398.
- Augustin, M., Conde Montero, E., Zander, N., Baade, K., Herberger, K., Debus, E. S., ... Blome, C. (2017, September). Validity and feasibility of the wound-QoL questionnaire on health-related quality of life in chronic wounds. *Wound Repair Regeneration* 25(5):852-857
- Guenther, U., Popp, J., Koecher, L., Muders, T., Wrigge, H., Ely, E. W., & Putensen, C. (2010). Validity and reliability of the CAM-ICU Flowsheet to diagnose delirium in surgical ICU patients. *Journal of Critical Care*, 25(1), 144-151.
- Hissong A.N., Lape, J.E., & Bailey, D.M. (2015). *Bailey's research for the health professional* (3rd ed.), Philadelphia: FA Davis.
- Law, M. & McDermid, J., (Eds.). (2014). Evidence-based rehabilitation: A guide to practice (3rd ed.). Thorofare, NJ: SLACK. Inc.
- Mouthaan, J., Sijbrandij, M., Reitsma, J. B., Gersons, B. P., & Olf, M. (2014). Comparing screening instruments to predict posttraumatic stress disorder. *PloS one*, 9(5), e97183. <https://doi.org/10.1371/journal.pone.0097183>
- Wang, Y. P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: a comprehensive review. *Brazilian Journal of Psychiatry*, 35(4), 416-431.
- Wei, L. A., Fearing, M. A., Sternberg, E. J., & Inouye, S. K. (2008). The Confusion Assessment Method: A Systematic Review of Current Usage. *Journal of the American Geriatrics Society*, 56(5), 823–830. doi: 10.1111/j.1532-5415.2008.01674.x
- Wu, C.-Y., Lee, M.-B., Lin, Y.-Y., & Liao, S.-C. (2019). Development and validation of the 9-item Concise Mental Health Checklist (CMHC-9) for suicide risk assessment. *Journal of the Formosan Medical Association*, 118(9), 1308–1316. doi: 10.1016/j.jfma.2019.05.025

Name of Appraiser (s): Megan Altom MOT/S Taylor Harrell MOT/S Carolynn Keane MOT/S Erica Smith MOT/S Danielle Woodward MOT/S	Date Completed: 3/4/2020
---	-----------------------------

Critically Appraised Papers
Team #8
March 2020

Critically Appraised Paper #1
**Data Extraction Form for Studies Evaluating the
Clinical Measurement Properties of Outcome Measures**

Authors: Aaron, L. A., Patterson, D. R., Finch, C. P., Carrouger, G. J., & Heimbach, D. M.

Year: 2001

Rater: Danielle Woodward, MOT/S

DATA EXTRACTED	
Population studied	
Population	27 consecutively admitted patients, 24 men and 3 women. The mean age of the group was 35 years of age. Participants were mostly White but included one African-American and one Hispanic. All were admitted to a major medical burn trauma center.
Intervention	N/A, since the study looked only at different assessments.
Reliability	
Reliability (relative)	α -coefficient for the abbreviated Burn Specific Pain Anxiety Scale = 0.90
Reliability (absolute)	N/A
Minimum detectable change	N/A
Content/Structural validity	
Internal consistency	N/A
Content validity	Results in the study concluded that the Burn Specific Pain Anxiety Scale (BSPAS) and its contents confidently predict how physical health will affect daily activities.
Floor-ceiling effects	Floor-ceiling effects were not discussed in this study.
Factorial validity	The purpose of the study was to see if the BSPAS had predictive validity, and after conducting the assessments on the patients, the researchers confirmed that the assessment was a very good predictor of physical health outcomes.
Item response/Rasch analyses	This study did not have items that were ranged in difficulties
Construct/Criterion validity	
Known groups	The researchers already knew about the high reliability of the BSPAS and wanted to test for its predictive validity.
Convergent	Researchers discussed how the BSPAS compares to the Profile of Mood States (POMS) and State-Trait Anxiety Inventory (STAI) in assessing/measuring anxiety and tension.
Divergent	The BSPAS, POMS, and STAI measured anxiety, while the SF-36 was used to measure physical and emotional functioning and the visual analogue scale (VAS) was used to measure pain.
Longitudinal validity	The assessments were used to establish a baseline for patients' anxiety and pain levels when they are admitted to the burn center.

Concurrent criterion	The researchers discussed how the BSPAS shows incredibly high levels of both criterion and concurrent validity, especially when it came to being correlated with a patient's procedural pain and burn severity.
Predictive criterion	Results from the study showed the BSPAS being a reliable tool to predict how physical dysfunction will affect performance in daily activities.
Responsiveness/Clinical change	
Responsiveness	The assessments studied were not used to measure change over time
Minimally clinical important difference	CID/MID was not discussed in the study
Usefulness/practicality	
Readability	The patients who were given the assessments were asked to give consent and were able to understand the assessments' words/phrases
Interpretability	Researchers discussed that the scoring of the results shows a high correlation between procedural pain and anxiety, while also showing a connection between physical dysfunction and daily activity performance.
Time to administer	Less than 10 minutes
Administration burden	Researchers found the results to be significant, but recommend the study be replicated with a larger sample size to gain more differences in assessments.
Cultural applicability	Suffering from burns does not discriminate, and thus, the applicability of this study to many different cultures/backgrounds is very relevant, as burns can happen to anyone, and returning to normalcy can be very difficult for anyone.

Rationale: The reason I feel this article works for our PICO is because the assessment specifically targets burn patients and how their anxiety could potentially affect their function to perform daily activities in the future.

SLACK Incorporated, 2014. Law, M., & MacDermid, J.C. (2014). *Evidence-Based Rehabilitation: A Guide to Practice, Third Edition*. Thorafare, NJ: SLACK Incorporated.

Critically Appraised Paper #1
Quality Appraisal for Clinical Measurement
Research Reports Evaluation Form

Authors: Aaron, L. A., Patterson, D. R., Finch, C. P., Carrouger, G. J., & Heimbach, D. M.

Year: 2001

Rater: Danielle Woodward, MOT/S

Evaluation Criteria	Score		
Study Question	2	1	0
1. Was the relevant background work cited to define what is currently known about the measurement properties of measures under study, and the potential contributions of the current research question to informing the knowledge base?		1	
Study Design			
2. Were appropriate inclusion/exclusion criteria defined?		1	
3. Were specific clinical measurement questions/hypotheses identified?	2		
4. Was an appropriate scope of measurement properties considered?		1	
5. Was an appropriate sample size used?	2		
6. Was appropriate retention/follow-up obtained? (For studies involving retesting; otherwise n/a)	N/A	N/A	N/A
Measurements			
7. Were specific descriptions provided of the measure under study and the method (s) used to administer it?	2		
8. Were standardized procedures used to administer all study measures?		1	
Analyses			
9. Were analyses conducted for each specific hypothesis or purpose?	2		
10. Were appropriate statistical tests used to obtain point estimates of the measurement of properties?		1	
11. Were appropriate ancillary analyses done to quantify in the estimates of the clinical measurement property or the confidence in the point estimate (confidence intervals, benchmark comparisons/ROC curves, alternate forms of analysis like SEM/MID, etc.)?			0
Recommendations			
12. Were clear, specific, and accurate conclusions made about the clinical measurement properties and supported by the study objectives, analysis, and results?	2		
Subtotals (of columns 1 and 2)	10	5	
Total score (sum of subtotals divided by 24 x 100). If for a specific paper or topic an item is deemed inappropriate then you can sum of items / 2 x number of items x 100	68%		

SLACK Incorporated, 2014. Law, M., & MacDermid, J.C. (2014). *Evidence-Based Rehabilitation: A Guide to Practice, Third Edition*. Thorafare, NJ: SLACK Incorporated.

Critically Appraised Paper #2
**Data Extraction Form for Studies Evaluating the
Clinical Measurement Properties of Outcome Measures**

Authors: Amoyal, N.R., Mason, S.T., Gould, N. F., Corry, N., Mahfouz, S., Barkey, A., & Fauerbach, J.A.

Year: 2011

Rater: Taylor Harrell, MOT/S

DATA EXTRACTED	
Population studied	
Population	362 participants admitted to Johns Hopkins Bayview Medical Center Burn Unit. Primarily men (75%), white (65%), employed at time of injury (72%). The most common types of burn injuries were flame (53%) and scald (17%). Participants had an average TBSA (total burn surface area) burned of 15% and an average TBSA grafted of 7%. Participants met American Burn Association criteria for a major burn.
Intervention	Participants completed the BCOPE survey at discharge from an acute stay hospital and then 24 months after discharge. Also, the Davidson trauma scale, satisfaction with appearance scale, and short-form health survey (SF-12). The BCOPE is a 28-item measure. The participants are asked to rate their agreeance one a 4-point scale.
Reliability	
Reliability (relative)	Results of the aforementioned studies yielded five or more factors across solutions, under scoring the need for a better understanding of coping behaviors in specific populations
Reliability (absolute)	DTS: test retest reliability with a coefficient of 0.86 ($P < .01$) Acceptance was positively associated with less affective distress at discharge, pre-burn, and 12 months after discharge. Test-retest reliability (from baseline to 6 months) ranged from $r.352$ for self-distraction to $r.855$ for humor (Table 1).
Minimum detectable change	N/A
Content/Structural validity	
Internal consistency	α coefficients for the scales ranging from 0.50 to 0.9 DTS: high internal consistency ($\alpha = 0.99$) for the frequency and severity items. SWAP: Good internal consistency has been reported for the SWAP, internal consistency (Cronbach's $\alpha = 0.87$)
Content validity	N/A
Floor-ceiling effects	N/A
Factorial validity	Exploratory factorial analysis was conducted using maximum likelihood extraction method.

	<p>Solutions were evaluated using Kaiser's criterion, 24 scree plot analysis, and cohesion of the factors on conceptual grounds. Exploratory factor analysis yielded a seven-factor solution that accounted for 51% of the total variance.</p> <p>Correlational analyses revealed that interitem correlations ranged from 0.01 to 0.76</p>
Item response/Rasch analyses	N/A
Construct/Criterion validity	
Known groups	N/A
Convergent	<p>Spearman correlations were used to determine the relationship between the BCOPE and the SF-12, DTS, and SWAP. Results indicated that the active and avoidance factors were each positively associated with total DTS scores at 6 and 12 months after discharge.</p> <p>Heightened post-traumatic stress disorder and distress and highly linked to ambivalent coping.</p>
Divergent	<p>Spearman correlations were used to determine the relationship between the BCOPE and the SF-12, DTS, and SWAP. Acceptance was negatively associated with SWAP at discharge. Using either approach or avoidant coping— but not both—has been shown to be related to less distressing outcome when compared to ambivalent coping.</p>
Longitudinal validity	N/A
Concurrent criterion	N/A
Predictive criterion	<p>Many of the assessments were done up to 12 months after discharge. Results indicated that the active and avoidance factors were each positively associated with total DTS scores at 6 and 12 months after discharge. Regarding subscale scores, active coping was positively associated with DTS intrusion at 1 and 6 months; DTS avoidance at 1, 6, and 12 months; and hyperarousal at 1, 6, and 12 months. Avoidance coping was significantly correlated with DTS intrusion at 1 and 12 months and DTS avoidance at all time points</p>
Responsiveness/Clinical change	
Responsiveness	Measured from time of discharge in acute care (baseline) to 6 months.
Minimally clinical important difference	N/A
Usefulness/practicality	
Readability	<p>The study was designed to assess the BCOPE, Davidson trauma scale, satisfaction with appearance scale, and short-form health survey (SF-12). The study looked at the different items on each test and compared if they would be relevant for the acute care settings to use with burn victims.</p>

Interpretability	The results were interpreted, and the study found certain parts of each test proved to distinguish different factors. Item scale correlations ranged from $r = .23$ to $.76$
Time to administer	The study talks about the need for a quick, valid and reliable assessment to use on burn patients in the acute care settings to aid in intervention and coping strategies.
Administration burden	N/A
Cultural applicability	N/A

SLACK Incorporated, 2014. Law, M., & MacDermid, J.C. (2014). *Evidence-Based Rehabilitation: A Guide to Practice, Third Edition*. Thorafare, NJ: SLACK Incorporated.

Critically Appraised Paper #2
Quality Appraisal for Clinical Measurement
Research Reports Evaluation Form

Authors: Amoyal, N.R., Mason, S.T., Gould, N. F., Corry, N., Mahfouz, S., Barkey, A., & Fauerbach, J.A.

Year: 2011

Rater: Taylor Harrell, MOT/S

Evaluation Criteria	Score		
	2	1	0
Study Question	2	1	0
1. Was the relevant background work cited to define what is currently known about the measurement properties of measures under study, and the potential contributions of the current research question to informing the knowledge base?	2		
Study Design			
2. Were appropriate inclusion/exclusion criteria defined?	2		
3. Were specific clinical measurement questions/hypotheses identified?	2		
4. Was an appropriate scope of measurement properties considered?	2		
5. Was an appropriate sample size used?	2		
6. Was appropriate retention/follow-up obtained? (For studies involving retesting; otherwise n/a)	2		
Measurements			
7. Were specific descriptions provided of the measure under study and the method (s) used to administer it?		1	
8. Were standardized procedures used to administer all study measures?		1	
Analyses			
9. Were analyses conducted for each specific hypothesis or purpose?	2		
10. Were appropriate statistical tests used to obtain point estimates of the measurement of properties?	2		
11. Were appropriate ancillary analyses done to quantify in the estimates of the clinical measurement property or the confidence in the point estimate (confidence intervals, benchmark comparisons/ROC curves, alternate forms of analysis like SEM/MID, etc.)?		1	
Recommendations			
12. Were clear, specific, and accurate conclusions made about the clinical measurement properties and supported by the study objectives, analysis, and results?	2		
Subtotals (of columns 1 and 2)	18	3	
Total score (sum of subtotals divided by 24 x 100). If for a specific paper or topic an item is deemed inappropriate, then you can sum of items / 2 x number of items x 100	87.5%		

SLACK Incorporated, 2014. Law, M., & MacDermid, J.C. (2014). *Evidence-Based Rehabilitation: A Guide to Practice, Third Edition*. Thorafare, NJ: SLACK Incorporated.

Critically Appraised Paper #3
**Data Extraction Form for Studies Evaluating the
Clinical Measurement Properties of Outcome Measures**

Authors: Augustin, M., Conde Montero, E., Zander, N., Baade, K., Herberger, K., Debus, E. S., Blome, C.

Year: 2017

Rater: Megan Altom, MOT/S

DATA EXTRACTED	
Population studied	
Population	227 Chronic Wound patients
Intervention	N/A
Reliability	
Reliability (relative)	This assessment was administered as a baseline and then two more times throughout an 8-week span.
Reliability (absolute)	N/A
Minimum detectable change	N/A
Content/Structural validity	
Internal consistency	Internal consistency was high in 3 time points; Cronbach's alpha=0.928, Cronbach's alpha=0.937, Cronbach's alpha=0.947
Content validity	Adult patients (age 18 years) with an existing chronic wound by definition were included at any chosen time point within the therapeutic program. Patients with a lack of mental, physical, or linguistic ability were excluded. Written consent was provided, and written data protection policy was available.
Floor-ceiling effects	A low floor effect was observed in T1: 0.5%, T2: 1%, and T3: 4%. Ceiling effects were also low (0%, 1%, 0.5%, respectively)
Factorial validity	As a statistically significant change in a PRO (patient-reported outcomes) score does not necessarily represent a clinically important improvement, and as it can be difficult to know if a PRO score is acceptable from the patient's point of view.
Item response/Rasch analyses	The wound-QoL instrument is a patient reported assessment. It gives them the chance to answer the questions on a range from none at all, a little, moderately, quite a bit, and very much. Item selectivity ranges were very similar over time. In T1, item selectivity ranged from 0.358 to 0.834; T2: 0.441 to 0.821, and T3: 0.538 to 0.828.
Construct/Criterion validity	
Known groups	N/A
Convergent	Correlation between FLQA- wk and Wound-QoL was high for both global score and subscales. This was observed in all 3 time points (higher than 0.8 <0.001).
Divergent	N/A
Longitudinal validity	Wound-QoL was found satisfactory in longitudinal validity since there were significantly ($p \leq 0.001$) good correlations with the EQ-

	5D (range 5 0.5–0.7) and FLQA-wk global score ($r > 0.8$) at every time point throughout the 8 weeks.
Concurrent criterion	<p>The correlation of the change in Wound-QoL subscales with the change in FLQA-wk subscales was also highly significant, ranging from $r = 0.514$ (psyche) to $r = 0.661$ (body) for T1-T2 and from $r = 0.577$ (psyche) to $r = 0.698$ (body) for T1-T3</p> <p>These correlations show concurrent validity with the Wound-QoL and the FLQA-wk assessment.</p>
Predictive criterion	N/A
Responsiveness/Clinical change	
Responsiveness	High correlation was found for the change in Wound-QoL global score with change in the wound size. The Wound-QoL global score was significantly better among those with a healed wound compared to the others (0.5 vs. 1.6; $p = 0.002$).
Minimally clinical important difference	The newly developed Wound-QoL was found to be valid and responsive and is useful as a short instrument for assessing health-related quality of life.
Usefulness/practicality	
Readability	N/A
Interpretability	N/A
Time to administer	Brief questionnaire, 17 questions. Time depends upon client answering and comprehending the questions
Administration burden	N/A; Article didn't specifically address administrator burden; however, it is a short, patient-reported assessment.
Cultural applicability	As a limitation, the current data were derived from specialized centers with a high expertise in wound care. There may thus be a selection of hard-to-heal wounds. Nevertheless, the study included patients from hospitals as well as from office-based physicians and a large variety of indications, thus providing a certain level of variability.

SLACK Incorporated, 2014. Law, M., & MacDermid, J.C. (2014). *Evidence-Based Rehabilitation: A Guide to Practice, Third Edition*. Thorafare, NJ: SLACK Incorporated.

Critically Appraised Paper #3
Quality Appraisal for Clinical Measurement
Research Reports Evaluation Form

Authors: Augustin, M., Conde Montero, E., Zander, N., Baade, K., Herberger, K., Debus, E. S., Blome, C.

Year: 2017

Rater: Megan Altom, MOT/S

Evaluation Criteria	Score		
	2	1	0
Study Question	2	1	0
13. Was the relevant background work cited to define what is currently known about the measurement properties of measures under study, and the potential contributions of the current research question to informing the knowledge base?	2		
Study Design			
14. Were appropriate inclusion/exclusion criteria defined?		1	
15. Were specific clinical measurement questions/hypotheses identified?	2		
16. Was an appropriate scope of measurement properties considered?			0
17. Was an appropriate sample size used?		1	
18. Was appropriate retention/follow-up obtained? (For studies involving retesting; otherwise n/a)			0
Measurements			
19. Were specific descriptions provided of the measure under study and the method (s) used to administer it?	2		
20. Were standardized procedures used to administer all study measures?		1	
Analyses			
21. Were analyses conducted for each specific hypothesis or purpose?	2		
22. Were appropriate statistical tests used to obtain point estimates of the measurement of properties?	2		
23. Were appropriate ancillary analyses done to quantify in the estimates of the clinical measurement property or the confidence in the point estimate (confidence intervals, benchmark comparisons/ROC curves, alternate forms of analysis like SEM/MID, etc.)?			0
Recommendations			
24. Were clear, specific, and accurate conclusions made about the clinical measurement properties and supported by the study objectives, analysis, and results?	2		
Subtotals (of columns 1 and 2)	12	3	
Total score (sum of subtotals divided by 24 x 100). If for a specific paper or topic an item is deemed inappropriate then you can sum of items / 2 x number of items x 100	62.5%		

SLACK Incorporated, 2014. Law, M., & MacDermid, J.C. (2014). *Evidence-Based Rehabilitation: A Guide to Practice, Third Edition*. Thorafare, NJ: SLACK Incorporated.

Critically Appraised Paper #4
**Data Extraction Form for Studies Evaluating the
Clinical Measurement Properties of Outcome Measures**

Authors: Guenther, U., Popp, J., Koecher, L., Muder, T., Wrigger, H., Ely, E. W., & Putensen, C.

Year: 2010

Rater: Erica Smith, MOT/S

DATA EXTRACTED	
Population studied	
Population	<p>Surgical ICU patients.</p> <p>Minimum sample size of 41 patients.</p> <p>Patients' enrollment. One hundred two patients were screened, 48 were excluded (24 for being in a comatose state, acute stroke, or non-German speaking) with another 23 immediate postoperative patients still under residual effects of anesthesia. One was unwilling to complete the assessment, rendering 54 to enroll who were then evaluated by the reference rater using DSM-IV criteria and the CAM-ICU Flowsheet raters.</p>
Intervention	N/A
Reliability	
Reliability (relative)	Delirium prevalence of 40% derived from preliminary studies in the ICU; and a 95% confidence interval.
Reliability (absolute)	The agreement (interrater reliability) between the 2 CAM-ICU Flowsheet investigators was very high (Cohen κ , 0.96; 95% confidence interval, 0.87-1.00).
Minimum detectable change	N/A
Content/Structural validity	
Internal consistency	N/A
Content validity	Delirium can be subdivided by assessing motoric symptoms: (1) hyperactive or "agitated" delirium with positive symptoms; (2) hypoactive or "quiet" delirium with negative symptoms; and (3) mixed type, if both subtypes appear alternately over time.
Floor-ceiling effects	N/A
Factorial validity	N/A
Item response/Rasch analyses	N/A
Construct/Criterion validity	
Known groups	Analyzed with the Mann-Whitney U test, except for the rate of mechanical ventilation and sex analysis, which were done with the Fisher exact test. Tests were 2-sided, and a $P < .05$ was considered statistically significant.

Convergent	N/A
Divergent	N/A
Longitudinal validity	N/A
Concurrent criterion	N/A
Predictive criterion	<p>The CAM-ICU Flowsheet has high sensitivity and high specificity.</p> <p>The CAM-ICU Flowsheet investigators had sensitivities of 88% (95% confidence interval, 69%-98%) and 92% (74%-99%), specificities of 100% (85%-100%).</p> <p>Screened for delirium (1) by a psychiatrist as the reference rater using the 4 delirium criteria of the Diagnostic and Statistical Manual of Mental Diseases, Fourth Edition (DSM-IV).</p> <p>The “CAM-ICU Flowsheet,” derived from the Confusion Assessment Method for Intensive Care Units (CAM-ICU), provides an algorithm by which to assess the 4 delirium criteria of the DSM-IV in a standardized fashion in intubated patients.</p> <p>Sensitivity (true positives [TP]/true positives [TP] + false negatives [FN]), specificity (true negatives [TN]/true negatives [TN] + false positives [FP]), positive predictive values (TP/TP + FP), negative predictive values (TN/TN + FN), and overall accuracies (TP + TN/TP + FP + TN) were calculated from 2 × 2 tables (Prism4 Software for Macintosh, GraphPad Software Inc, San Diego, Calif).</p>
Responsiveness/Clinical change	
Responsiveness	N/A
Minimally clinical important difference	N/A
Usefulness/practicality	
Readability	N/A
Interpretability	<p>Accuracies of 94% to 96%</p> <p>It is important to note that the CAM-ICU Flowsheet switches the original numbering of features 3 and 4 for simplicity because most ICU patients with delirium are positive in the order of the flow sheet, thus allowing the CAM-ICU Flowsheet to be completed in just 3 features and only needing to include the fourth feature in a minority of patients.</p> <p>Delirium subtypes were classified into a motoric subtype grouping according to the Richmond Agitation Sedation Scale (RASS), which was rated for the feature 3 (“altered level of consciousness”) of the CAM- ICU Flowsheet.</p>

	<p>Patients were deemed as having hypoactive delirium if they were DSM positive by the reference rater and had RASS -3 to 0, or deemed as having hyperactive delirium if their RASS was between +1 to +4. Mixed type is defined by alternating between either state, but because we only evaluated each patient once to avoid repeat-observer bias, mixed-type delirium was not diagnosed in this investigation. Patients with a RASS of -4 or -5 were considered comatose and were excluded from the study because comatose patients cannot be assessed for delirium.</p>
Time to administer	<p>Quick assessment instrument.</p> <p>Median assessment times of less than a minute</p> <p>The median time to complete the CAM-ICU Flowsheet assessment was 45 seconds (IQR, 40-75 seconds) in patients without delirium vs 50 seconds (IQR, 40-120 seconds; maximum, 180 seconds) in patients with delirium.</p>
Administration burden	N/A
Cultural applicability	<p>The CAM-ICU Flowsheet is measured similarly across cultures.</p> <p>The German CAM-ICU Flowsheet was translated according to the Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient- Reported Outcomes Measures into German language [31,32]. This included the “forward translation” of the original English CAM-ICU Flowsheet, the revision by a geriatric psychiatrist, and the back-translation into English by 2 physicians who were unaware of the original.</p> <p>The German version is also available online at www.icudelirium.org.</p>

SLACK Incorporated, 2014. Law, M., & MacDermid, J.C. (2014). *Evidence-Based Rehabilitation: A Guide to Practice, Third Edition*. Thorafare, NJ: SLACK Incorporated.

Critically Appraised Paper #4
Quality Appraisal for Clinical Measurement
Research Reports Evaluation Form

Authors: Guenther, U., Popp, J., Koecher, L., Muder, T., Wrigger, H., Ely, E. W., & Putensen, C.
 Year: 2010
 Rater: Erica Smith, MOT/S

Evaluation Criteria	Score		
Study Question	2	1	0
1. Was the relevant background work cited to define what is currently known about the measurement properties of measures under study, and the potential contributions of the current research question to informing the knowledge base?	2		
Study Design			
2. Were appropriate inclusion/exclusion criteria defined?		1	
3. Were specific clinical measurement questions/hypotheses identified?		1	
4. Was an appropriate scope of measurement properties considered?	2		
5. Was an appropriate sample size used?		1	
6. Was appropriate retention/follow-up obtained? (For studies involving retesting; otherwise n/a)			0
Measurements			
7. Were specific descriptions provided of the measure under study and the method (s) used to administer it?	2		
8. Were standardized procedures used to administer all study measures?	2		
Analyses			
9. Were analyses conducted for each specific hypothesis or purpose?		1	
10. Were appropriate statistical tests used to obtain point estimates of the measurement of properties?		1	
11. Were appropriate ancillary analyses done to quantify in the estimates of the clinical measurement property or the confidence in the point estimate (confidence intervals, benchmark comparisons/ROC curves, alternate forms of analysis like SEM/MID, etc.)?		1	
Recommendations			
12. Were clear, specific, and accurate conclusions made about the clinical measurement properties and supported by the study objectives, analysis, and results?	2		
Subtotals (of columns 1 and 2)	10	6	
Total score (sum of subtotals divided by 24 x 100). If for a specific paper or topic an item is deemed inappropriate then you can sum of items / 2 x number of items x 100	73%		

SLACK Incorporated, 2014. Law, M., & MacDermid, J.C. (2014). *Evidence-Based Rehabilitation: A Guide to Practice, Third Edition*. Thorafare, NJ: SLACK Incorporated.

Critically Appraised Paper #5
**Data Extraction Form for Studies Evaluating the
Clinical Measurement Properties of Outcome Measures**

Authors: Mouthaan, J., Sijbrandij, M., Reitsma, J. B., Gersons, B. P. R., & Olf, M.

Year: 2014

Rater: Danielle Woodward MOT/S

DATA EXTRACTED	
Population studied	
Population	311 participants, with injuries sustained in a traumatic event.
Intervention	No intervention was present in the study.
Reliability	
Reliability (relative)	Clinical interview of the Clinician Administered PTSD scale (CAPS) – (ICC = 0.98) Startle, Physiological arousal, Anger, and Numbness (SPAN) - (ICC = 0.83) Trauma Screening Questionnaire (TSQ) – (ICC = 0.82) Impact of Event Scale-Revised (IES-R) – (ICC = 0.83)
Reliability (absolute)	No standard error of measurement was reported in the study
Minimum detectable change	The MDC was not identified in this; the one thing specified was if a patient would be diagnosed with PTSD, their score would have to have been at least 45 or higher.
Content/Structural validity	
Internal consistency	High internal consistencies were shown for the CAPS subscales (Intrusion: $\alpha = 0.91$, Avoidance: $\alpha = 0.83$, and Hyperarousal: $\alpha = 0.86$).
Content validity	All the assessments used (CAPS, SPAN, TSQ, and IES-R) measured what they were made to measure, helping predict future mental health issues with patients suffering from trauma-related injuries.
Floor-ceiling effects	Floor-ceiling effects were not measured in this study.
Factorial validity	Researchers discussed that the assessments used were excellent tools for detecting future PTSD and other mental illnesses, though further testing must be done in order to accurately diagnosis the illnesses in patients.
Item response/Rasch analyses	Items on the assessments were not ranged in difficulty.
Construct/Criterion validity	
Known groups	Known groups were not discussed in the study.
Convergent	All assessments used were able to adequately detect PTSD in patients with trauma-related injury, whether it was a short assessment or long one.
Divergent	Researchers only discussed that further testing/assessing must be done in order to accurately diagnosis mental illness in patients.
Longitudinal validity	After 6 months and being reassessed, researchers found that 5.8% of the patients studied were diagnosed with 6-month PTSD, 7.1% were

	diagnosed with Major Depressive Episode (MDE), and 7.8% with anxiety disorder (AD).
Concurrent criterion	SPAN and CAPS (Pearson $r = 0.65, p < 0.001$) TSQ and CAPS (Pearson $r = 0.72, p < 0.001$).
Predictive criterion	Researchers discussed that the assessments used were good predictors at diagnosing PTSD and other mental illnesses in patients suffering from trauma-related injuries.
Responsiveness/Clinical change	
Responsiveness	The assessments can be used to diagnose patients with mental illness and can be used as a baseline to create and plan interventions that might help with those later diagnoses.
Minimally clinical important difference	MDC was not addressed in this study.
Usefulness/practicality	
Readability	The patients were able to understand the assessments, and results were able to be understood and analyzed appropriately.
Interpretability	Researchers were able to interpret and find moderate to strong correlations between the assessments on predicting mental illnesses such as PTSD.
Time to administer	Varied, as the SPAN and TSQ were short, and the IES-R and CAPS were long.
Administration burden	The CAPS was done by those who were trained to administer it, as well as all the other assessments used in the study.
Cultural applicability	Though the study took place in the Netherlands, the assessments used are widely known and used often when it comes to trauma-related injuries in the ICU.

Rationale: This article, despite looking at mostly other patients besides burn patients, the assessments compared were great for predicting mental health complications for those who are admitted to the ICU, for whatever reason.

SLACK Incorporated, 2014. Law, M., & MacDermid, J.C. (2014). *Evidence-Based Rehabilitation: A Guide to Practice, Third Edition*. Thorafare, NJ: SLACK Incorporated.

Critically Appraised Paper #5
Quality Appraisal for Clinical Measurement
Research Reports Evaluation Form

Authors: Mouthaan, J., Sijbrandij, M., Reitsma, J. B., Gersons, B. P. R., & Olf, M.

Year: 2014

Rater: Danielle Woodward MOT/S

Evaluation Criteria	Score		
Study Question	2	1	0
1. Was the relevant background work cited to define what is currently known about the measurement properties of measures under study, and the potential contributions of the current research question to informing the knowledge base?	2		
Study Design			
2. Were appropriate inclusion/exclusion criteria defined?		1	
3. Were specific clinical measurement questions/hypotheses identified?	2		
4. Was an appropriate scope of measurement properties considered?	2		
5. Was an appropriate sample size used?	2		
6. Was appropriate retention/follow-up obtained? (For studies involving retesting; otherwise n/a)	N/A	N/A	N/A
Measurements			
7. Were specific descriptions provided of the measure under study and the method (s) used to administer it?	2		
8. Were standardized procedures used to administer all study measures?		1	
Analyses			
9. Were analyses conducted for each specific hypothesis or purpose?	2		
10. Were appropriate statistical tests used to obtain point estimates of the measurement of properties?	2		
11. Were appropriate ancillary analyses done to quantify in the estimates of the clinical measurement property or the confidence in the point estimate (confidence intervals, benchmark comparisons/ROC curves, alternate forms of analysis like SEM/MID, etc.)?	2		
Recommendations			
12. Were clear, specific, and accurate conclusions made about the clinical measurement properties and supported by the study objectives, analysis, and results?	2		
Subtotals (of columns 1 and 2)	18	2	
Total score (sum of subtotals divided by 24 x 100). If for a specific paper or topic an item is deemed inappropriate then you can sum of items / 2 x number of items x 100	91%		

SLACK Incorporated, 2014. Law, M., & MacDermid, J.C. (2014). *Evidence-Based Rehabilitation: A Guide to Practice, Third Edition*. Thorafare, NJ: SLACK Incorporated.

Critically Appraised Paper #6
**Data Extraction Form for Studies Evaluating the
Clinical Measurement Properties of Outcome Measures**

Authors: Wang, Y.P, & Gorenstein, C.

Year: 2013

Rater: Erica Smith, MOT/S

DATA EXTRACTED	
Population studied	
Population	A variety of populations. The instrument was applied to over 60,000 respondents.
Intervention	N/A
Reliability	
Reliability (relative)	Twenty-nine of the 118 retrieved articles (25%) did not report reliability coefficients, indicating that the assumption of test score reliability generally has not prevailed in clinical practice regarding application of the BDI.
Reliability (absolute)	Retest reliability (Pearson's r) showed relative stability through re-application of the BDI-II, with good to excellent coefficients (range, 0.73 to 0.96), with a mean re-application interval of 2 weeks (range, 1 week to 6 months) for the majority of studies (82%).
Minimum detectable change	N/A
Content/Structural validity	
Internal consistency	In comparison to the internal consistency of the previous versions of the BDI (average Cronbach's alpha coefficient around 0.85), 8 most studies on BDI-II reported an average alpha coefficient around 0.9, ranging from 0.83 to 0.96.
Content validity	The English version of the BDI-II has been translated into 17 languages, and is used in Europe, the Middle East, Asia, and Latin America (Table 1). Although the English version prevailed among the studies (65%), the increasing number of language versions suggests inter-national acceptance of the instrument. The content validity of the BDI-II appears to be adequate, but narrower than that of the former version. The BDI-I reflected six of the nine criteria for DSM-based depression, while the BDI-II presented an improved performance on specificity to indicate DSM-based depression.
Floor-ceiling effects	N/A
Factorial validity	Beck reported a median item-total scale correlation of 0.59 for the BDI-II in a sample of college students (n=120). Acceptable item-total scale correlations (it is less than or equal to 0.5) were described for 17 out of 21 items. Nonetheless, this correlation can vary across studies. For the Arabic version, substantial item-total

	correlation was described for 10 items among Islamic students, whereas adequate item-total correlation of the Portuguese version in Brazilian samples was reported for 15 items. Factors such as language version, type of sample, age range, educational level, and severity of depression might affect the difficulty of item endorsement. Insight into which items should be assigned to a scale can improve its performance through item-level analysis. Factor analysis showed a robust dimension of general depression composed by two constructs: cognitive-affective and somatic-vegetative.
Item response/Rasch analyses	Most validation studies of the BDI-II were analyzed on the grounds of classic test theory (CTT), assuming a true score for each respondent and disregarding the measurement error. In other words, two individuals with the same total score may differ in terms of the relative severity and frequency of symptoms.
Construct/Criterion validity	
Known groups	The confirmatory strategy has been employed to compare the structure and model fit of previous studies in relation to the construct validity of the BDI-II.
Convergent	<p>The convergent validity between the BDI-I and the BDI-II was high, with Pearson's product-moment correlation coefficients (r) ranging from 0.82 to 0.94.</p> <p>The convergent validity between the BDI-II and scales that assess anxiety – such as the Beck Anxiety Inventory (BAI), the Hamilton Anxiety Rating Scale (HAM-A), and the State-Trait Anxiety Inventory (STAI) – was also significant, with a wide range of correlation coefficients (0.37 to 0.83; rough estimate of 0.50).</p>
Divergent	N/A
Longitudinal validity	N/A
Concurrent criterion	This study reported a comparison of the BDI-II with scales measuring depression, anxiety, and miscellaneous constructs as criterion, determined at essentially the same time to check for concurrent validity.
Predictive criterion	In general, studies reported a sensitivity of 0.70. Sensitivity should be viewed as the most important indicator to minimize the chance of false-negative diagnosis of depressive disorders.
Responsiveness/Clinical change	
Responsiveness	The mean score ranged from 5.1 to 38.4.
Minimally clinical important difference	N/A
Usefulness/practicality	
Readability	N/A
Interpretability	N/A
Time to administer	N/A

Administration burden	Scores were found to significantly decrease for the weekly administration group only, indicating that lower retest scores could be the result of a measurement effect and the frequency of administration.
Cultural applicability	<p>With the BDI-II being such a popular measure adapted for use in several countries, information on cross- cultural comparability is still remarkably scarce.</p> <p>The cross-cultural equivalence between the versions of the BDI-II stands out as a topic of fervent academic interest: the symptomatology of depression in different culture/races or languages can be compared by testing the measurement of variance of the instrument.</p>

SLACK Incorporated, 2014. Law, M., & MacDermid, J.C. (2014). *Evidence-Based Rehabilitation: A Guide to Practice, Third Edition*. Thorafare, NJ: SLACK Incorporated.

Critically Appraised Paper #6
Quality Appraisal for Clinical Measurement
Research Reports Evaluation Form

Authors: Wang, Y.P, & Gorenstein, C.

Year: 2013

Rater: Erica Smith, MOT/S

Evaluation Criteria	Score		
Study Question	2	1	0
1. Was the relevant background work cited to define what is currently known about the measurement properties of measures under study, and the potential contributions of the current research question to informing the knowledge base?	2		
Study Design			
2. Were appropriate inclusion/exclusion criteria defined?		1	
3. Were specific clinical measurement questions/hypotheses identified?	2		
4. Was an appropriate scope of measurement properties considered?	2		
5. Was an appropriate sample size used?	2		
6. Was appropriate retention/follow-up obtained? (For studies involving retesting; otherwise n/a)	2		
Measurements			
7. Were specific descriptions provided of the measure under study and the method (s) used to administer it?	2		
8. Were standardized procedures used to administer all study measures?	2		
Analyses			
9. Were analyses conducted for each specific hypothesis or purpose?	2		
10. Were appropriate statistical tests used to obtain point estimates of the measurement of properties?	2		
11. Were appropriate ancillary analyses done to quantify in the estimates of the clinical measurement property or the confidence in the point estimate (confidence intervals, benchmark comparisons/ROC curves, alternate forms of analysis like SEM/MID, etc.)?		1	
Recommendations			
12. Were clear, specific, and accurate conclusions made about the clinical measurement properties and supported by the study objectives, analysis, and results?	2		
Subtotals (of columns 1 and 2)	20	2	
Total score (sum of subtotals divided by 24 x 100). If for a specific paper or topic an item is deemed inappropriate then you can sum of items / 2 x number of items x 100	92%		

SLACK Incorporated, 2014. Law, M., & MacDermid, J.C. (2014). *Evidence-Based Rehabilitation: A Guide to Practice, Third Edition*. Thorafare, NJ: SLACK Incorporated.

Critically Appraised Paper #7
**Data Extraction Form for Studies Evaluating the
Clinical Measurement Properties of Outcome Measures**

Authors: Wei, L. A., Fearing, M. A., Sternberg, E. J., & Inouye, S. K.

Year: 2008

Rater:Carolynn Keane, MOT/S

DATA EXTRACTED	
Population studied	
Population	Patients with dementia, depression, psychosis
Intervention	The use of the Confusion Assessment Method to address recognition and management of delirium.
Reliability	
<ul style="list-style-type: none"> - Interrater reliability (.70-1.00) was moderate to high across all studies Gonzalez 2004: <ul style="list-style-type: none"> ▪ Sensitivity: .90 ▪ Specificity: 1.0 ▪ Inter-rater reliability: .89 - Laurila 2002: <ul style="list-style-type: none"> ▪ Sensitivity: .81-.86 ▪ Specificity: .63-.84 ▪ Inter-rater reliability: -- - Fabbri 2001: <ul style="list-style-type: none"> ▪ Sensitivity: .94 ▪ Specificity: .96 ▪ Inter-rater reliability: .70 - Monette 2001: <ul style="list-style-type: none"> ▪ Sensitivity: .86 ▪ Specificity: 1.0 ▪ Inter-rater reliability: .91 - Ely 2001a: <ul style="list-style-type: none"> ▪ Sensitivity: .95-1.0 ▪ Specificity: .89-.93 ▪ Inter-rater reliability: .79-.95 - Ely 2001b: <ul style="list-style-type: none"> ▪ Sensitivity: .93-1.0 ▪ Specificity: .98-1.0 ▪ Inter-rater reliability: .96 - Rolfson 1999: <ul style="list-style-type: none"> ▪ Sensitivity: .70 ▪ Specificity: 1.00 ▪ Inter-rater reliability: -- - Zou 1998: <ul style="list-style-type: none"> ▪ Sensitivity: .89 ▪ Specificity: 1.00 	

	<ul style="list-style-type: none"> ▪ Inter-rater reliability: 0.86 <ul style="list-style-type: none"> - Pompei 1995: <ul style="list-style-type: none"> ▪ Sensitivity: .46 ▪ Specificity: .92 ▪ Inter-rater reliability: -- - Rockwood 1994: <ul style="list-style-type: none"> ▪ Sensitivity: .68 ▪ Specificity: .97 ▪ Inter-rater reliability: -- - Inouye 1990: <ul style="list-style-type: none"> ▪ Sensitivity: .94-1.0 ▪ Specificity: .90-.95 ▪ Inter-rater reliability: .81-1.0
Reliability (relative)	ICC not stated
Reliability (absolute)	SEM not stated
Minimum detectable change	Results were combined across 7 high quality studies (n=1071), demonstrating an overall sensitivity of 94% (95% confidence interval, CI, 91-97%), and specificity of 89% (95% CI, 85-94%).
Content/Structural validity	
Internal consistency	Cronbach's alpha values not stated
Content validity	N/A
Floor-ceiling effects	N/A
Factorial validity	N/A
Item response/Rasch analyses	N/A
Construct/Criterion validity	
Known groups	N/A
Convergent	<ul style="list-style-type: none"> - CAM-ICU - CAM for ED - MDS Version 3.0 - Nursing Home CAM - Delirium Index - One day Fluctuation Scale
Divergent	In two validation studies (11-12), the CAM-ICU was compared with ratings by delirium experts using the DSM-IV, yielding a sensitivity of 95-100%, specificity of 93-98%, and interrater reliability of .79-.95. When the non-verbal ratings on the CAM-ICU were directly compared to verbal ratings (24), CAM-ICU ratings demonstrated lower sensitivity of 73% and lower interrater reliability of .64, but maintained a high specificity of 100%.
Longitudinal validity	N/A
Concurrent criterion	N/A
Predictive criterion	<ul style="list-style-type: none"> - Positive predictive accuracy of 91-94% - Negative predictive accuracy of 90-100%
Responsiveness/Clinical change	

Responsiveness	
Minimally clinical important difference	N/A
Usefulness/practicality	
Readability	<ul style="list-style-type: none"> - Comprehensive literature search for articles published between January 1, 1991 and December 31, 2006 using PubMed, EMBASE, PsychINFO, CINAHL, Ageline, and Google Scholar. Full text searches for the search term 'Confusion Assessment Method'. Applying additional search terms, such as 'delirium', 'acute confusional state'. Excluded articles based off of inclusion criteria. - In over 222 application studies, we have documented that the CAM is widely used for both clinical and research applications.
Interpretability	Results compared on sensitivity, specificity and inter-rater reliability.
Time to administer	Not stated
Administration burden	Some training is recommended for optimal use. (See Confusion Assessment Method Training Manual)
Cultural applicability	CAM has been translated into 10 languages.

SLACK Incorporated, 2014. Law, M., & MacDermid, J.C. (2014). *Evidence-Based Rehabilitation: A Guide to Practice, Third Edition*. Thorafare, NJ: SLACK Incorporated.

Critically Appraised Paper #7

Instructions for completion:

Please refer to the attached dictionary for definition of terms and instructions for completing each section. For each criteria, score by placing a check mark in the appropriate box.

First Author: Leslie A. Wei
 Year: 2008
 Journal: N/A
 Reviewer: Carolynn Keane, MOT/S

CRITERIA	YES	NO
Q1. Did the authors have a clearly focused question [population, intervention (strategy), and outcome(s)]?	X	
Q2. Were appropriate inclusion criteria used to select primary studies?	X	
Q3. Did the authors describe a search strategy that was comprehensive? <i>Circle all strategies used:</i> <ul style="list-style-type: none"> ▪ health databases ▪ psychological databases ▪ social science databases ▪ educational databases ▪ other <ul style="list-style-type: none"> ▪ handsearching ▪ key informants ▪ reference lists ▪ unpublished 	X	
Q4. Did search strategy cover an adequate number of years?	X	
Q5. Did the authors describe the level of evidence in the primary studies included in the review? <ul style="list-style-type: none"> ▪ Level I → RCTs only ▪ Level II → non-randomized, cohort, case-control ▪ Level III → uncontrolled studies 	N/A	N/A
Q6. Did the review assess the methodological quality of the primary studies, including: <i>(Minimum requirement: 4/7 of the following)</i> <ul style="list-style-type: none"> ▪ Research design ▪ Study sample ▪ Participation rates ▪ Sources of bias (confounders, respondent bias) ▪ Data collection (measurement of independent/dependent variables) ▪ Follow-up/attrition rates ▪ Data analysis 		X
Q7. Are the results of the review transparent?	X	
Q8. Was it appropriate to combine the findings of results across studies?	X	
Q9. Were appropriate methods used for combining or comparing results across studies?	X	
Q10. Do the data support the author's interpretation?	X	
TOTAL SCORE:	80%	

Quality Assessment
 Rating:

Strong
 (total score 8 – 10)

Moderate
 (total score 5 – 7)

Weak
 (total score 4 or less)

Critically Appraised Paper #8
**Data Extraction Form for Studies Evaluating the
Clinical Measurement Properties of Outcome Measures**

Authors: Chia-Yi Wu, Ming-Been Lee, Yi-Yin Lin, Shih-Cheng Liao

Year: 2019

Rater:Carolynn Keane, MOT/S

DATA EXTRACTED	
Population studied	
Population	<ul style="list-style-type: none"> - 3 groups of subjects <ul style="list-style-type: none"> ▪ 931 psychiatric outpatients ▪ 931 non-psychiatric medical outpatients ▪ 2120 community residents - Recruited from convenience samples of the outpatient clinics of the psychiatric department and the non-psychiatric departments of family medicine, oncology, and emergency medicine in a medical center in Northern Taiwan.
Intervention	Participants were invited to answer a battery of structured questions in written form (patient subjects) or by telephone form (community subjects) by personnel with specific training for the study. The structured questionnaire including CMHC items and additional questions for demographic characteristics and suicidality was administered via a standardized procedure in different settings.
Reliability	
Reliability (relative)	No ICC given
Reliability (absolute)	No standard error of measurement given
Minimum detectable change	CMHC-9 with AUCs (recent SI 92.9%: 95% CI = 91.5-94.3; lifetime SI 75.9%: 95% CI= 73.9-77.9)
Content/Structural validity	
Internal consistency	The internal consistency of the CMHC-9 was satisfactory based on Cronbach's alpha values for the total sample (alpha= 0.79) and the three sub-samples (i.e., alpha= 0.76 for psychiatric subjects; 0.67 for community subjects; and 0.69 for medical outpatient subjects).
Content validity	To reduce the risk of "lack of assessment" that limits opportunities for mental health care, universal screening is recommended
Floor-ceiling effects	<ul style="list-style-type: none"> - It can be part of clinical assessment in detecting patients' recent psychopathology and suicidality, potentially improving patient-centered care and subjective and objective intervention outcomes. - The lowest percentage of scoring level of 0 was found in the psychiatric outpatient group, while over 30-45% of the other two samples also scored 0. This demonstrates the floor effect in the psychiatric group as well as distinct discrimination between lower scoring groups from the higher scoring group.

Factorial validity	<ul style="list-style-type: none"> - The factorial structure of CMHC-9 was invariant across groups and the summed scores could differentiate a high-risk group in the psychiatric sub-sample from other medical or community sub-samples. - The results indicated that three common first-order factors were extracted with an eigenvalue greater than one and parallel analysis. In total, 54% of the total variance was explained by these three factors: <ul style="list-style-type: none"> ▪ the highest % of variance was explained by Factor 1 (31%), including five items of psychopathology (with factor loadings of 0.61-0.91); ▪ Factor 2 (31%), containing four items of suicidality from the CSPA such as lifetime suicide attempt, future suicidal intent, substance abuse and lack of support (loadings = 0.95, 0.54, 0.42, and 0.34, respectively) ▪ Factor 3 (8%), consisting of the two items of age and marital status (loadings= 0.68 and 0.65, respectively).
Item response/Rasch analyses	Post hoc analysis using the Turkey HSD test revealed significant differences in paired comparisons among community subjects, medical patients, and psychiatric patients (Means = 1.2, 1.9, and 4, SDs = 1.6, 1.8, and 2.3, respectively, $p < 0.001$).
Construct/Criterion validity	
Known groups	<ul style="list-style-type: none"> - The demographics differed significantly among the three sub-samples, with a higher rate of females and those divorced/separated/widowed in psychiatric subjects and younger age in community subjects. - The descriptive statistics of the suicide risk factor items/scales differed significantly among the three sub-samples, with the highest percentages of all risk factor items and overall scale scores of the psychopathology and CMHC-9 in psychiatric subjects, and the lowest percentages in community subjects. - The psychiatric subjects had more severe suicide risk and psychological distress than the other two groups.
Convergent	<ul style="list-style-type: none"> - The CMHC-9 was originated from the CMHC, which contained 16 items for suicide risk assessment with a three-factor structure of demographics, psychopathology and suicidality. - CMHC consisted of 5 items of psychopathology from the BSRS-5R, nine items from the CSPA and 2 additional measures for recent and lifetime suicide ideation
Divergent	There is a lack of correlation among the tests.
Longitudinal validity	N/A

Concurrent criterion	Time-specific assessment of suicide risks including recent psychopathological symptoms (within seven days), past suicide attempt (over the lifetime) and future suicide attempt. <ul style="list-style-type: none"> ▪ BSRS-5R- Satisfactory psychometric properties as a measure to detect psychiatric morbidity and SI in medical settings or in the community. ▪ CPS
Predictive criterion	The ROC curves of the CMHC-9 scores revealed that the optimal cut-off to predict recent SI was $\frac{3}{4}$ for all subjects with 92.1% sensitivity, 82.0% specificity, 99.2% negative predictive value and 30.6% positive predictive value.
Responsiveness/Clinical change	
Responsiveness	Yes, it can assess suicide risks including recent psychopathological symptoms (within seven days), past suicide attempt (over the lifetime) and future suicide intent.
Minimally clinical important difference	The optimal cut-offs to predict recent SI in all subjects and in different demographic subgroups were examined. The significance of all tests was set at $p < 0.05$.
Usefulness/practicality	
Readability	The structured questionnaire including CMHC items and additional questions for demographic characteristics and suicidality (e.g., serious consideration of suicide over the past week or over the lifetime) was administered via a standardized procedure in different settings.
Interpretability	The CMHC-9 scores among the three samples were compared with ANOVA and found significant between-group differences ($F(2, 3979) = 821.5, p < 0.001$).
Time to administer	Not given, only 9-items so fairly short
Administration burden	Can be self-administered or interview format
Cultural applicability	Checklist does not change amongst cultures

SLACK Incorporated, 2014. Law, M., & MacDermid, J.C. (2014). *Evidence-Based Rehabilitation: A Guide to Practice, Third Edition*. Thorafare, NJ: SLACK Incorporated.

Critically Appraised Paper #8
Quality Appraisal for Clinical Measurement
Research Reports Evaluation Form

Authors: Chia-Yi Wu, Ming-Been Lee, Yi-Yin Lin, Shih-Cheng Liao

Year: 2019

Rater:Carolynn Keane, MOT/S

Evaluation Criteria	Score		
Study Question	2	1	0
1. Was the relevant background work cited to define what is currently known about the measurement properties of measures under study, and the potential contributions of the current research question to informing the knowledge base?	2		
Study Design			
2. Were appropriate inclusion/exclusion criteria defined?	2		
3. Were specific clinical measurement questions/hypotheses identified?			0
4. Was an appropriate scope of measurement properties considered?		1	
5. Was an appropriate sample size used?			0
6. Was appropriate retention/follow-up obtained? (For studies involving retesting; otherwise n/a)	n/a	n/a	n/a
Measurements			
7. Were specific descriptions provided of the measure under study and the method (s) used to administer it?	2		
8. Were standardized procedures used to administer all study measures?	2		
Analyses			
9. Were analyses conducted for each specific hypothesis or purpose?	2		
10. Were appropriate statistical tests used to obtain point estimates of the measurement of properties?	2		
11. Were appropriate ancillary analyses done to quantify in the estimates of the clinical measurement property or the confidence in the point estimate (confidence intervals, benchmark comparisons/ROC curves, alternate forms of analysis like SEM/MID, etc.)?			0
Recommendations			
12. Were clear, specific, and accurate conclusions made about the clinical measurement properties and supported by the study objectives, analysis, and results?		1	
Subtotals (of columns 1 and 2)	13	2	
Total score (sum of subtotals divided by 24 x 100). If for a specific paper or topic an item is deemed inappropriate then you can sum of items / 2 x number of items x 100	68%		

SLACK Incorporated, 2014. Law, M., & MacDermid, J.C. (2014). *Evidence-Based Rehabilitation: A Guide to Practice, Third Edition*. Thorafare, NJ: SLACK Incorporated.