	Scales		
	Pain	**Delirium**	**Sedation**
Scale Development: Item Selection and Content Validation			
Was the process of item selection described? 2: Scale was developed for a specific population, using a theoretical or conceptual framework, or a qualitative approach was used (e.g. consultation with clinicians or patients) 1: Scale was developed based on the literature review only 0: No information is provided about item selection	√	√	√
Was content evaluated by experts? (content validation) 2: Content was evaluated by experts in the field, a Delphi technique may have been used, and Content Validity Index (CVI) were calculated for each item included in the scale 1: Content was evaluated by experts, but no CVI is reported 0: No information is provided about content validation	√	√	√
Are limitations of some items presented or discussed? 1: No limitations or if any limitations, they are presented and item modifications have been made or precautions have been stated 0: No information is provided	√	√	√
Subtotal – Scale Development	5	5	5
*Subtotal weighted score – Scale Development	2	2	2
Scale testing: Reliability			
Was internal consistency of the scale calculated? 2: 0.70<α<0.90 1: 0.60<α<0.70 or α>0.90 0: α<0.60 or no information provided	√		
Was interrater reliability calculated? 2: kappa>0.60 or ICC>0.80 1: 0.60<kappa>0.40 or 0.60<ICC<0.80 0: kappa<0.40, ICC<0.60 or no information provided	√	√	√
Was interrater reliability tested with other raters besides research team? 1: Other raters then research staff members were involved 0: Only research staff members were involved	√	√	√
Optional – To be examined if ICC>0.80 not consistent in all studies Was intrarater reliability tested? Specify test-retest interval: _____ 2: kappa>0.60 or ICC>0.80 1: 0.60<kappa>0.40 or 0.60<ICC<0.80 0: kappa<0.40, ICC<0.60 or no information provided	√	√	√

Subtotal – Reliability	5 or 7	3 or 5	3 or 5
*Subtotal weighted score – Reliability	6	6	6
Scale Testing: Construct Validity			
What is the total of participants for the purpose of testing the scale? 2: N>50 1: 20<N<50 0: N<20	√	√	√
Criterion validation: Was the scale correlated with the "gold standard" measure in the field of interest (e.g. the patient's self-report of pain)? 2: r>0.60 with the "gold standard" measure 1: 0.40<r<0.60 0: r<0.40 or no information provided	√		
Criterion validation: Was the sensitivity of the scale calculated? 2: Sensitivity≥80% 1: 60%≤Sensitivity<80% 0: Sensitivity<60% or no information provided	√	√	
Criterion validation: Was the specificity of the scale calculated? 2: Specificity≥80% 1: 60%≤Specificity<80% 0: Specificity<60% or no information provided	√	√	
Predictive validation: Is the scale score able to predict some outcome(s) that will be available later on during the patient's ICU stay, e.g. delirious patients with higher ICU mortality rate? 2: A clinically important difference between groups (presence versus absence of delirium) and the outcome was found 1: A difference was found but was not considered clinically important 0: No difference was found or no information is provided		√	
Convergent validation: Was the scale correlated with another tool, ideally using a different method (e.g. BIS, EEG if analyzing a subjective sedation scale), measuring the same construct or related construct? 2: r>0.60 with another type of measure of same construct or related construct 1: 0.40<r<0.60 0: r<0.40 or no information provided			√

Discriminant validation: Was the scale able to discriminate between different situations, e.g. between pain and no pain (e.g. at rest and during a nociceptive procedure, before and after the administration of an analgesic)? 2: A clinically important difference was found 1: A difference was found but was not considered clinically important 0: No difference was found or no information is provided	√		√
Subtotal – Validity	10	8	6
*Subtotal weighted score – Validity	8	8	8
Scale Feasibility			
Was the feasibility (i.e. ease of usage with which clinicians can apply the instrument in the clinical setting) of the scale examined? 1: Scale is considered to be feasible to use by more than 80% of the clinicians 0: Scale is considered to be complex to use by more than 20% of the clinicians or no information is provided	√	√	√
Are directives of use of the scale clearly described? 1: Yes, directives of use including the scoring method are described 0: No information about directives of use is provided	√	√	√
Subtotal – Feasibility	2	2	2
*Subtotal weighted score – Feasibility	2	2	2
Scale Relevance or Impact of Implementation in ICU patient outcomes			
Was the relevance of the scale or impact of its implementation in ICU patient outcomes examined? 1: Scale is considered to be relevant to practice by more than 80% of the clinicians; use of the scale yielded a significant change into practice (e.g. better use of medication, increase in patients' assessments) 0: Scale is not considered to be relevant to practice by more than 20% of the clinicians; use of the scale did not yield to a significant change into practice or no information provided	√	√	√
Subtotal – Impact of implementation at bedside	1	1	1
*Subtotal weighted score – Impact of implementation at bedside	2	2	2
Total Score	23 or 25	19 or 21	17 or 19
Total Weighted Score	20	20	20

*The subtotal weighted score represents a different range than the subtotal score, but keeps the same proportions. It is calculated using the rule of three.