

Numeric Rating Scale (NRS)

Verscheidene auteurs, afgeleid van de VAS

Meetinstrument	Numeric Rating Scale
Afkorting	NRS
Auteur	Verscheidene auteurs
Thema	Symptoommanagement pijn
Doel	Pijn beoordelen bij de patiënt
Populatie	Niet gespecificeerd
Afname	Zorgverlener
Aantal items	1 item
Aanwezigheid patiënt	Aanwezigheid van de patiënt is vereist
Vindplaats meetinstrument	http://www.ndhcri.org/

Doel

Pijncontrole bij de patiënt aan de hand van zelfrapportage op een één-item schaal.

Doelgroep

De NRS werd reeds gevalideerd in een uiteenlopende, diverse populatie. Patiëntengroepen waarin de validiteit en de betrouwbaarheid van de schaal werd bestudeerd, zijn o.a. ouderen, chirurgiepatiënten, kinderen, patiënten met musculoskeletale problemen, oncologische patiënten, en andere groepen met pijnproblemen (Carey, Turpin, Smith, Whatley, & Haddox, 1997; Gagliese, Weizblit, Ellis, & Chan, 2005; Herr, Spratt, Mobily, & Richardson, 2004; Jensen, Engel, McKearnan, & Hoffman, 2003; Kaasalainen & Crook, 2003; Mawdsley, Moran, & Conniff, 2002; Paice & Cohen, 1997; Puntillo & Neighbor, 1997; Spadoni, Stratford, Solomon, & Wishart, 2004; Taylor & Herr, 2003; Ware, Epps, Herr, & Packard, 2006; Wong & Baker, 1988).

Bijgevolg kan het gebruik van de NRS naar een brede populatie van pijnpatiënten generaliseerd worden.

Beschrijving

De NRS is duidelijk afgeleid van de VAS. Het betreft eveneens een schaal waarvan de uiteinden begrensd zijn door de termen 'Geen pijn' en 'Ergst denkbare pijn'. In tegenstelling tot de VAS is de NRS genummerd in gelijke intervallen. De schaalindeling loopt van 1 tot 10. Dit levert 11 keuzemogelijkheden op voor de patiënt waarbij 0 gelijk staat aan 'geen pijn'; en 10 de 'ergst denkbare pijn' betekent.

Bedoeling is dat de patiënt de pijnintensiteit die hij/ zij beleeft op het moment van afname, aanduidt op de lijn.

Net zoals bij de VAS bestaan er tal van varianten op de NRS. Verschillen zijn er ondermeer terug te vinden in de terminologie die men hanteert om het eindpunt af te bakenen (vb. extreme pijn, meest verschrikkelijke pijn, ergst mogelijke pijn,...), het aantal punten om de schaal in te delen (vb. 6-punten NRS, 21-punten NRS), de oriëntatie van de schaal (verticaal – horizontaal), het tijdsinterval waar de pijnmeting betrekking op heeft (bv. ‘Scoor u gemiddelde pijn van de afgelopen 24h’),...

Betrouwbaarheid

De betrouwbaarheid van de NRS werd reeds in verscheidene studies onderzocht. Meerdere studies rapporteren een *intra-rater reliability* hoger dan $r = 0.70$ (Kaasalainen & Crook, 2003; Mawdsley et al., 2002; Ware, Epps, Herr, & Packard, 2006; Taylor & Herr, 2003). In het onderzoek van Spadoni et al. (2004) bedroeg deze correlatie $r = 0.63$. De waarde ervan kan men in vraag stellen aangezien de retest pas 72h later plaatvond. Een verandering in pijnbeleving bij de patiënt kan mogelijks aan de oorzaak liggen van de zwakkere correlatie.

Cronbach's alpha tussen de VAS en andere één-item meetschalen werd nagegaan in de studie van Herr et al. (2004) en in het onderzoek van Carey et al. (1997). De correlatie bedroeg respectievelijk $\alpha = 0.97$ (VAS – Numeric Rating Scale – Verbal Descriptor Scale – Verbal Numeric Scale – Faces Pain Scale) en $\alpha = 0.88$ (VAS – Faces Rating Scale – Numeric Rating Scale).

De *inter-rater reliability* genereerde een overeenstemming van 100% tussen de verschillende beoordelaars (Herr et al., 2004).

Validiteit

Net zoals voor de betrouwbaarheid, werd ook de validiteit van de NRS reeds uitgebreid bestudeerd. Gagliese et al. (2005), Jensen et al. (2003), Kim & Buschmann (2006), Paice & Cohen (1997), Puntillo & Neighbor (1997), Spadoni et al. (2004), Taylor & Herr (2003), Ware et al. (2006), Wong & Baker (1988) onderzochten de *concurrent validity* van de NRS, door het meetinstrument te correleren aan andere pijnschalen (o.a. VAS, Verbal Rating Scale, McGill Pain Questionnaire, Faces Pain Scale, Iowa Pain Thermometer...). De range van de bekomen correlaties in deze studies bedroeg $r = 0.48 - 0.96$. Het merendeel van de bekomen correlaties is echter overwegend sterk. Vergelijkingen tussen personen met een cognitief verminderd functioneren en cognitief intacte personen, toont aan dat deze correlaties zwakker zijn voor de eerstgenoemde groep (Gagliese et al., 2005; Kaasalainen & Crook, 2003; Taylor & Herr, 2003; Ware et al., 2006). Mogelijks leent de NRS er zich minder toe als meetinstrument bij patiënten met een cognitief verminderd functioneren (bv. dementerenden). Een alternatief kan er uit bestaan om de Faces Pain Scale of de Verbal Rating Scale te hanteren.

Om de *construct validity* van de VAS na te gaan, hebben Gagliese et al. (2005), Jensen et al. (2003), en Herr et al. (2004) een *principale componenten analyse* uitgevoerd. De 3 studies abstraherden 1 factor uit de scores op de verschillende

één-item pijnschalen. De factorlading van de NRS was telkens hoger dan 0.95, uitgezonderd in de studie van Jensen et al. (2003) waar de factorlading 0.80 bedroeg voor de NRS.

Tevens werden significante correlaties gerapporteerd tussen NRS-scores en intensiteit van de toegediende pijnprikkel (Herr et al., 2004), gedragsobservaties van de pijnpatiënt (Kaasalainen & Crook, 2003) en pijngerelateerde constructen zoals depressie en functionele beperkingen tengevolge pijn (Jensen et al., 2003).

Gebruiksvriendelijkheid

De NRS werd niet uitgetest op haar gebruiksvriendelijkheid aangezien pijnmanagement in de door ons bevraagde ziekenhuizen aan de hand van de VAS beoordeeld wordt. Doch, wij menen te kunnen stellen dat er een grote overlap bestaat tussen de gebruiksvriendelijkheid van de NRS en de VAS. Omwille van deze reden verwijzen we naar de rubriek 'Gebruiksvriendelijkheid' van de VAS. Rekening houdend dat de afstand tussen het nulpunt en de markering van de patiënt niet dient opgemeten te worden (dit in tegenstelling tot de VAS), kan men er van uitgaan dat er minder fouten zullen gemaakt worden door verpleegkundigen en dat afname van de NRS vlotter zal verlopen.

Opmerkingen

Een vergelijking tussen de verschillende meetinstrumenten toont aan dat de NRS een hoog betrouwbare en valide schaal is in het beoordelen van pijn. Niettegenstaande de VAS het frequentst gehanteerd wordt in de praktijk, verdient de NRS mogelijks de voorkeur op de VAS. Meerdere studies rapporteren namelijk moeilijkheden bij patiënten om de VAS in te vullen (Gagliese et al., 2005; Herr et al., 2004; Paice & Cohen 1997; Pautex et al., 2006). Vergelijkingen tussen de VAS en de Numeric Rating Scale (NRS) duiden op minder problemen in het hanteren van de laatstgenoemde schaal. Daarnaast geniet de NRS een vergelijkbare betrouwbaarheid en validiteit. Bovendien geven patiënten zelf aan dat de NRS en ander pijnschalen hun voorkeur verdienen op de VAS (Carey et al., 1997; Dworkin et al., 2005; Gagliese et al., 2005; Herr et al., 2004; Paice & Cohen 1997). Meerdere validatiestudies bevelen de NRS dan ook aan als het meest geschikte instrument in het beoordelen van pijn (Downie et al., 1978; Dworkin et al., 2005; Jensen et al., 2003). Dit dient echter genuanceerd te worden aangezien geen enkele pijnschaal bij alle vormen van pijn of voor alle leeftijdscategorieën kan gebruikt worden. Tevens ontbreekt, in contrast met de VAS, uitvoerig onderzoek naar de sensitiviteit van de NRS.

Niettegenstaande de betrouwbaarheid en validiteit van de NRS ondersteund wordt bij kinderen (Jensen & Karoly in Gagliese et al., 2005), kan het aangewezen zijn te opteren voor de FPS in het controleren van pijn bij deze patiëntgroep. Dit in het bijzonder voor jonge kinderen (< 12 jaar), aangezien de FPS meer aanschouwelijk en beter begrijpbaar is voor hen. Daarenboven werden de psychometrische eigenschappen van de FPS reeds uitvoerig bestudeerd en dit bevestigt het gebruik van de FPS bij kinderen (Bieri, Reeve, Champion, Addicoat, & Ziegler, 1990; Bosenberg, Thomas, Lopez, Kokinsky & Larsson, 2003; Hicks, von Baeyer, Spafford,

van Korlaar & Goodenough, 2001; Paik & Ahn, 2002; Stinson, Kavanagh, Yamada, Gill & Stevens, 2006; Wong & Baker, 1988).

Referenties

Bieri, D., Reeve, R. A., Champion, G. D., Addicoat, L., & Ziegler, J. B. (1990). The Faces Pain Scale for the self-assessment of the severity of pain experienced by children: development, initial validation, and preliminary investigation for ratio scale properties. *Pain*, *41*, 139-150.

Bosenberg, A., Thomas, J., Lopez, T., Kokinsky, E., & Larsson, L. E. (2003). Validation of a six-graded faces scale for evaluation of postoperative pain in children. *Paediatr Anaesth.*, *13*, 708-713.

Carey, S. J., Turpin, C., Smith, J., Whatley, J., & Haddox, D. (1997). Improving pain management in an acute care setting. The Crawford Long Hospital of Emory University experience. *Orthop.Nurs.*, *16*, 29-36.

Downie, W. W., Leatham, P. A., Rhind, V. M., Wright, V., Branco, J. A., & Anderson, J. A. (1978). Studies with pain rating scales. *Ann Rheum Dis*, *37*, 378-381.

Dworkin, R. H., Turk, D. C., Farrar, J. T., Haythornthwaite, J. A., Jensen, M. P., Katz, N. P. et al. (2005). Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*, *113*, 9-19.

Gagliese, L., Weizblit, N., Ellis, W., & Chan, V. W. (2005). The measurement of postoperative pain: a comparison of intensity scales in younger and older surgical patients. *Pain*, *117*, 412-420.

Herr, K. A., Spratt, K., Mobily, P. R., & Richardson, G. (2004). Pain intensity assessment in older adults: use of experimental pain to compare psychometric properties and usability of selected pain scales with younger adults. *Clin.J.Pain*, *20*, 207-219.

Hicks, C. L., von Baeyer, C. L., Spafford, P. A., van, Korlaar, I, & Goodenough, B. (2001). The Faces Pain Scale-Revised: toward a common metric in pediatric pain measurement. *Pain*, *93*, 173-183.

Jensen, M. P., Engel, J. M., McKearnan, K. A., & Hoffman, A. J. (2003). Validity of pain intensity assessment in persons with cerebral palsy: a comparison of six scales. *J Pain*, *4*, 56-63.

Kim, E. J. & Buschmann, M. T. (2006). Reliability and validity of the Faces Pain Scale with older adults. *Int J Nurs Stud.*, *43*, 447-456.

Kaasalainen, S. & Crook, J. (2003). A comparison of pain-assessment tools for use with elderly long-term-care residents. *Can J Nurs Res*, *35*, 58-71.

Mawdsley, R. H., Moran, K. A., & Conniff, L. A. (2002). Reliability of two commonly used pain scales with elderly patients. *Journal of Geriatric Physical Therapy*, 25, 16-20.

Paice, J. A. & Cohen, F. L. (1997). Validity of a verbally administered numeric rating scale to measure cancer pain intensity. *Cancer Nurs*, 20, 88-93.

Paik, H. J. & Ahn, Y. M. (2002). Measurement of acute pain after eye surgery in children. *Korean J Ophthalmol.*, 16, 103-109.

Pautex, S., Michon, A., Guedira, M., Emond, H., Le Lous, P., Samaras, D. et al. (2006). Pain in severe dementia: self-assessment or observational scales? *J Am Geriatr Soc*, 54, 1040-1045.

Puntillo, K. A. & Neighbor, M. L. (1997). Two methods of assessing pain intensity in English-speaking and Spanish-speaking emergency department patients. *J Emerg Nurs*, 23, 597-601.

Spadoni, G. F., Stratford, P. W., Solomon, P. E., & Wishart, L. R. (2004). The evaluation of change in pain intensity: a comparison of the P4 and single-item numeric pain rating scales. *J Orthop Sports Phys Ther*, 34, 187-193.

Stinson, J. N., Kavanagh, T., Yamada, J., Gill, N., & Stevens, B. (2006). Systematic review of the psychometric properties, interpretability and feasibility of self-report pain intensity measures for use in clinical trials in children and adolescents. *Pain*, 125, 143-157.

Taylor, L. J. & Herr, K. (2003). Pain intensity assessment: a comparison of selected pain intensity scales for use in cognitively intact and cognitively impaired African American older adults. *Pain Manag.Nurs*, 4, 87-95.

Ware, L. J., Epps, C. D., Herr, K., & Packard, A. (2006). Evaluation of the Revised Faces Pain Scale, Verbal Descriptor Scale, Numeric Rating Scale, and Iowa Pain Thermometer in older minority adults. *Pain Manag.Nurs*, 7, 117-125.

Wong, D. L. & Baker, C. M. (1988). Pain in children: comparison of assessment scales. *Pediatr Nurs*, 14, 9-17.

Vindplaats meetinstrument

<http://www.ndhcri.org/hcqip/>

NUMERIC RATING SCALE (NRS)

VERSCHEIDENE AUTEURS SINDS1974

Author (year)	Setting	Sample (n)	Design	Reliability	Validity
Mawdsley, R. H., Moran, K. A., & Conniff, L. A. (2002)	An outpatient setting.	Elderly with musculoskeletal injury or disorders with pain aged 61 to 84. (n = 32)	Repeated measures design (VAS and NRS measurements 15 minutes apart).	S	
Gagliese, L., Weizblit, N., Ellis, W., & Chan, V. W. (2005)	Toronto General Hospital, University Health Network and Mount Sinai Hospital, Toronto, Ontario.	Patients scheduled to receive patient-controlled analgesia following general surgery. (n = 504)	Comparative study: the VAS was compared with the Numeric Rating Scale (NRS), the Verbal Descriptor Scale (VDS) and the pain intensity measures of the McGill Pain Questionnaire (MPQ).		FV CrV CsV

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Results reliability	Results validity	Commentary
<p>(S) <i>Intra-rater reliability NRS:</i> r = 0.74</p> <p><i>Intra-rater reliability VAS:</i> r = 0.92</p>		
	<p>(FV) <i>Face validity:</i> Patients rated the VAS as the least accurate and least preferred instrument for future assessments. The VAS had the highest proportion of patients making errors ($p \leq 0.005$).</p> <p>(CrV) <i>Concurrent validity:</i> - Correlation between VAS and VDS and NRS: 1. Patients aged > 60 years: r = 0.60-0.93 ($p < 0.001$) 2. Patients aged ≤ 60 years: r = 0.72-0.91 ($p < 0.001$) - Correlation between VAS and MPQ was lower.</p> <p>(CsV) <i>Principal components analysis:</i> A one-factor model was extracted from the different measurement scales and accounted for 82% of the total variance in the young group and 85% of the variance in the older group. Factor loading was 0.96.</p>	

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Author (year)	Setting	Sample (n)	Design	Reliability	Validity
Spadoni, G. F., Stratford, P. W., Solomon, P. E., & Wishart, L. R. (2004)	20 outpatient physical therapy clinics. The clinics were located in the Canadian Provinces of Alberta, Ontario, Quebec and the State of Georgia, USA.	220 patients with musculoskeletal problems receiving physical therapy. 213 patients (97%) were available for retest; and 183 patients (83%) provided data at follow-up. (n = 220)	Comparative study: to determine whether the P4 was more adept at assessing pain change than 2 versions of a single-item NRS. One version inquired about pain intensity over the past 24 hours and the second version asked about pain intensity over the past 2 days. NRSs and the P4 were administered on 3 occasions - initial visit, within 72 hours of baseline, and 12 days following baseline assessment.	S	CrV
Paice, J. A. & Cohen, F. L. (1997)	A large tertiary care hospital.	A convenience sample of 50 adult oncological patients with pain. (n = 50)	Comparative study: the NRS was compared with VAS and Simple Descriptor Scale (SDS).		CrV
Carey, S. J., Turpin, C., Smith, J., Whatley, J., & Haddox, D. (1997)	Inpatient units of the Crawford Long Hospital, Atlanta.	The admitted diagnosis for 39.5% of the sample was acute pain, 40.3 with chronic pain and 20.2 with no pain. (n = 267)	Comparative study: the VAS was compared with FRS and NRS.	IC	

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Results reliability	Results validity	Commentary
<p>(S) Test-retest reliability: The second assessment was obtained within 72 hours of the baseline visit.</p> <ul style="list-style-type: none"> - P4: $r = 0.78$ (95% CI [0.72, 0.83]) - NRS 24-hour: $r = 0.63$ (95% CI [0.54, 0.70]) - NRS 2 day : $r = 0.70$ (95% CI [0.62, 0.77]) 	<p>(CrV) Concurrent validity: The criterion validity analysis produced a correlation coefficient of 0.57 between the P4 and 24-hour NRS. This difference is statistically significant ($z = 1.73$, $p = .043$). The longitudinal validity analysis yielded correlation coefficients of 0.61 and 0.56 between the retrospective rating of change and the P4 and 2-day NRS, respectively. This difference is statistically significant ($z = 2.53$, $p = .006$).</p>	<p>Ratings are retrospective. This might bias the validity of the NRS-scores. However, the aim of the NRS is to rate the present pain of the patient and not the pain of the past few days. We cast doubt on such longitudinal measurements; an adequate pain policy requires pain assessment on a regularly base and not every two days.</p>
	<p>(CrV) Concurrent validity: -The strong positive correlation between the NRS and VAS was statistically significant: $r = 0.85$, $p < 0.001$. -NRS was also correlated to SDS: $r = 0.83$, $p < 0.001$</p>	<p>A majority of subjects (50%) preferred the use of the NRS when comparing the three scales used to measure pain intensity. Fewer patients preferred the SDS (38%), with the VAS chosen least often (12%).</p> <p>Eleven subjects (20%) in this study were unable to complete the VAS or did so with great difficulty. All subjects were able to complete the NRS and SDS without apparent difficulty.</p>
<p>(IC) Cronbach's alpha: Intercorrelations between 3 scales was $\alpha = 0.88$.</p>		<p>Patients selected the FPS as 'easiest to use' (48.6%), followed by the NRS (35.3%) and the VAS (16.1%).</p>

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Author (year)	Setting	Sample (n)	Design	Reliability	Validity
Ware, L. J., Epps, C. D., Herr, K., & Packard, A. (2006)	Not specified.	68 subjects aged 60 years and older with cognitive impairments. The mean score for the Mini Mental State Examination (MMSE) was 23 (standard deviation = 5.4) with a range from 10 to 30. Fifty-nine percent (n = 40) of the sample scored 24 or greater indicating no cognitive impairment (CI). Forty-one percent (n = 28) scored less than 24 indicating some degree of CI. (n = 68)	Comparative study: Subjects were instructed to recall a vividly remembered pain and rate this remembered pain using the Iowa Pain Thermometer (IPT), the Verbal Descriptor Scale (VDS), a 0 to 10 Numeric Rating Scale (NRS), and the Faces Pain Scale Revised (FPS-R).	S	CrV

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Results reliability	Results validity	Commentary
<p>(S) Test-retest reliability: The researcher returned 2 weeks later and reminded subjects of the “vividly remembered pain” identified at the first assessment and asked them to rate that pain again using all four scales. In the cognitively intact group, Spearman rank correlation coefficients between the two-week vividly remembered pain ratings were 0.87 (NRS), followed by the VDS (0.86), IPT (0.81), and FPS-R (0.76). Correlations ranged from 0.77 to 0.89 in the CI group. The FPS-R had the strongest test–retest reliability coefficient (0.89) in this group followed by the IPT (0.82), VDS (0.79), and NRS (0.77).</p>	<p>(CrV) Concurrent validity: Spearman correlations between pain ratings on the selected scales ranged from 0.64 to 0.90 in the cognitively intact group and from 0.56 to 0.83 in the CI group. The lowest correlations in the CI group and intact group (0.64–0.84) were found between the FPS-R and the other scales (0.56–0.66).</p>	<p>In terms of the concurrent validity, moderate to high inter-tool correlations for the CI and cognitively intact groups were found with the exception of low correlations associated with the FPS-R, suggesting that the FPS-R may measure overall affect as opposed to pain.</p> <p>Four participants with moderate CI were unable to follow directions and complete the VDS and IPT. The NRS had the highest failure rate with six participants with moderate CI and one mildly impaired participant unable to use the scale. No failures occurred when using the FPS-R to evaluate pain.</p> <p>The NRS (n = 12/36, 33%) was the preferred scale in the cognitively intact group, and the FPS-R (n = 13/24, 54%) was the preferred scale in the CI group.</p>

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Author (year)	Setting	Sample (n)	Design	Reliability	Validity
Herr, K. A., Spratt, K., Mobily, P. R., & Richardson, G. (2004)	Subjects were recruited through college bulletin board displays, community faith centers, senior-citizen centers, senior-housing, and long-term facilities.	86 younger adults (age 25-55). 89 older adults (age 65-94). (n = 175)	Repeated measures design (Noxious heat stimuli were delivered to the ventral forearm by an electronically controlled contact thermode. The heat stimuli were programmed to last 5 seconds and to present randomly 43°C, 45°C, 46°C, 47°C, 48°C, 49°C or 51°C. A 2 minute trial interval followed each stimulus presentation during which the subject rated the stimulus on 5 scales).	IC E	CsV Sen

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Results reliability	Results validity	Commentary
<p>(IC) Internal consistency: Intercorrelations between the scales were all statistically significant at every temperature ($p < 0.001$). <i>Cronbach's alpha within each scale across the 7 temperature values:</i> VAS: 0.88 NRS: 0.88 VDS: 0.86 VNS: 0.88 FPS: 0.88</p> <p><i>Cronbach's alpha within temperature across the 5 scales:</i> 43°C: 0.96 45°C: 0.96 46°C: 0.96 47°C: 0.97 48°C: 0.97 49°C: 0.97 51°C: 0.97</p> <p>(E) Inter-rater reliability: VAS: 93.5% NRS: 100% VDS: 100% VNS: 100% FPS: 100%</p>	<p>(CsV) Principal components analysis: A one-factor model was extracted from the different measurement scales. The correlation of each scale to the isolated factor are as follows: VAS: 0.94 NRS: 0.96 VDS: 0.95 VNS: 0.95 FPS: 0.86</p> <p>(Sen) Each tool demonstrated significant increases in score associated with increase in temperature ($p < 0.001$).</p>	<p>The VAS had a significant higher failure rate (6.7%) in comparison with the NRS, VDS, VNS, and FPS.</p> <p>The VNS demonstrated significantly higher levels of pain report than the other 4 scales.</p> <p>The psychometric scale evaluation was conducted using an experimental pain stimulus instead of using clinical pain stimuli.</p> <p>The scales most preferred in order by the total sample is as follows: NRS (35.3%); VDS (25.3%); VNS (15.9%); FPS (12.9%); VAS (10.6%).</p>

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Author (year)	Setting	Sample (n)	Design	Reliability	Validity
Kaasalainen, S. & Crook, J. (2003)	A 240-bed long-term-care facility in urban southwestern Ontario, Canada.	4 groups of 130 elderly long-term-care residents: (1) cognitively intact, (2) mildly cognitively impaired, (3) moderately cognitively impaired, and (4) extremely cognitively impaired. (n = 130)	Repeated measures design: FPS, NRS, Present Pain Intensity Scale (PPI) were conducted twice 48 hours apart.	S	CsV

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Results reliability	Results validity	Commentary
<p>(S) Test-retest reliability: Test-retest reliability for the three verbal-report scales was moderate to strong for the cognitively intact group (FPS: ICC = 0.84; PPI: ICC = 0.55; NRS: ICC = 0.87) but decreased for the other groups. In addition, the error variances were low for the cognitively intact group (FPS: s2 error = 0.53; PPI: s2 error = 0.71; NRS: s2 error = 1.45) but increased with increasing cognitive impairment.</p>	<p>(CsV) Convergent validity: The Pearson r correlations of the Pain Assessment in the Communicatively Impaired (PACI) tool, a behavioural-observation measure, with the three verbal-report scales (FPS, PPI, NRS) were low to moderate. For the cognitively intact group, all of these correlations were moderate and significant (FPS: $r = 0.66, p < 0.001$; PPI: $r = 0.62, p < 0.01$; NRS: $r = 0.65, p < 0.01$). For the mildly impaired group, none were significant at the $p < 0.05$ level. For the moderately impaired group, the PACI correlated moderately and significantly with the FPS ($r = 0.63, p < 0.001$) and PPI ($r = 0.64, p < 0.001$). However, the correlation between the PACI and NRS for those with moderate impairment was low and non significant ($r = 0.30, p < 0.12$).</p>	<p>Test-retest for the 3 verbal-report scales was moderate to strong for elderly persons with no cognitive impairment but decreased for the other groups. Similarly, error variances were low for those with no cognitive impairment but increased with increasing cognitive impairment. These findings indicate that the level of cognitive impairment decreases the reliability of verbal reports of pain.</p> <p>Test-retest reliability for both the NRS and the FPS was strong for residents without cognitive impairment but declined considerably for those with mild and moderate impairment, suggesting that these tools may not be good choices for use with these two groups.</p> <p>For those with moderate cognitive impairment, the PPI seems to be a more appropriate and reliable tool than the FPS or the NRS.</p>

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Author (year)	Setting	Sample (n)	Design	Reliability	Validity
Taylor, L. J. & Herr, K. (2003)	Recruitment took place through Elder Care Services, a private non-profit organization.	A convenience sample of 57 volunteers age 58 and older. Seventy-seven percent (n = 44) of the sample scored 24 or less on the mental status exam, indicating some degree of cognitive impairment. The remaining 23% (n = 13) were cognitively intact. (n = 57)	Comparative study: Subjects were instructed to recall a vividly remembered pain and rate this remembered pain using the FPS, the VDS, the NRS and the Iowa Pain Thermometer (IPT).	S	CrV

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Results reliability	Results validity	Commentary
<p>(S) Test-retest reliability: The researcher returned 2 weeks later and reminded subjects of the “vividly remembered pain” identified at the first assessment and asked them to rate that pain again using all four scales. Spearman rank correlation coefficients between the 2-week vividly remembered pain ratings ranged from 0.52 to 0.83 in both groups. In the cognitively impaired group, the FPS had the strongest reliability coefficient (0.79), followed by the VDS (0.63), NRS (0.57) and IPT (0.52). In the intact group, the strongest correlation was noted with the IPT (0.83), followed by the FPS (0.81), NRS (0.74), and VDS (0.73).</p>	<p>(CrV) Concurrent validity: Spearman correlations between present pain ratings on the selected scales were statistically significant ($p = 0.01$) and ranged from 0.81 to 0.96 in the intact group and from 0.74 to 0.83 in the impaired group. The lowest correlation was found between the FPS and VDS ($r = 0.74$).</p>	<p>All of the participants were able to use each of the pain intensity scales to rate their present pain in a manner that allowed interpretation of a single pain score (e.g., not selecting more than one response, selection of options outside the scale range). No failures were noted.</p> <p>The FPS showed strong test-retest stability in the cognitively impaired minority older adults.</p> <p>Of the 35 older adults who identified a scale preference, the FPS was the preferred scale in both the intact group ($n = 5$, 62.5%) and the impaired group ($n = 15$, 56%). However, it should be noted that almost 40% of both groups had no preference for a specific pain intensity scale.</p>

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Author (year)	Setting	Sample (n)	Design	Reliability	Validity
Puntillo, K. A. & Neighbor, M. L. (1997)	A level 1 trauma center emergency department in San Francisco.	Ninety-five English-speaking patients (67 male and 28 female) and 21 Spanish-speaking patients (16 male and 5 female) participated in the study. Study patients were treated in the emergency department for sprains or strains (35%), fractures (19%), contusions (14%), cellulitis or abscesses (12%), or other miscellaneous conditions such as headaches or abdominal/flank pain (20%). (n = 59)	Repeated measures design: patients were asked to use the NRS and VRS (English or Spanish version according the language of the patient) seven times over a 2-hour period, once immediately before and six times after receiving an analgesic.		CrV
Wong, D. L. & Baker, C. M. (1988)	Pediatric units of two general hospitals in the South-Central United States.	Hospitalized children in three age groups: 3 to 7 (n = 52), 8 to 12 (n = 52), and 13 to 18 years (n = 46). (n = 150)	Comparative study. The following six scales were compared on reliability and validity: Simple Descriptive Scale (SDS), NRS, FPS, the Glasses Scale, the Chips Scale, and the Color Scale.	S	CrV

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Results reliability	Results validity	Commentary
	<p>(CrV) Concurrent validity: Spearman correlations between scores on the two scales were computed for each of the seven data collection times. Correlations between the two scales were moderate to very high ($r = 0.48$ to 0.96) and statistically significant ($p < 0.05$ to $p < 0.001$) at each of the seven time periods. In fact, there was an increase in the relationship between the two pain scores (NRS and VRS) from the beginning to the end of the study.</p>	<p>Slightly more English speaking patients preferred the NRS (50) over the VRS (35), and slightly more Spanish-speaking patients preferred the VRS (11) over the NRS (9). However, these findings were not statistically significant.</p>
<p>(S) Test-retest reliability: Retest occurred the day after the initial test except in a few instances when it was done 2 to 4 days later. Percentage agreement for the 6 scales was:</p> <ul style="list-style-type: none"> - SDS: 72.73% - NRS: 75.44% - FPS: 74.24% - Glasses: 75.38% - Chips: 77.27% - Colors: 68.18% 	<p>(CrV) Concurrent validity: Children were asked to list painful events that they had experienced since being hospitalized and rank these from most to least painful. Each scale was also used to rate these painful events. A percentage consistency between the ranking of the events and the responses for each pain scale was calculated. Percentage agreement for the 6 scales was:</p> <ul style="list-style-type: none"> - SDS: 62.81% - NRS: 60.00% - FPS: 60.43% - Glasses: 63.70% - Chips: 69.06% - Colors: 58.39% 	<p>The FPS was the most preferred scale by all age groups.</p> <p>The finding of an increase in validity and reliability with age is consistent with children's advancing cognitive ability. However, reliability increased only from the 3 to 7 year age group. Reliability decreased in the 13 to 18 year age group for all the scales except for the color scale and the SDS.</p> <p>No one scale demonstrates superiority in validity or reliability. No significant differences exist among the scales for any age group.</p>

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Author (year)	Setting	Sample (n)	Design	Reliability	Validity
Jensen, M. P., Engel, J. M., McKearnan, K. A., & Hoffman, A. J. (2003)	Not specified.	Persons with cerebral palsy (CP) who have reported the presence of a chronic pain problem. Pain intensity assessment data were available for 24 participants from the survey study and 45 participants from the longitudinal study (n = 69)	Comparative study . Study participants were recruited from 2 other ongoing studies (a single-assessment survey and a 2-year longitudinal study). A subgroup of those who reported ongoing problems with pain were recruited to participate in a longitudinal study and completed measures of pain and pain impact at 5 time points (11- and 21 point NRS, 5- and 16 point VRS, 6- and 7 point FPS). All of the current study participants came from the same population.		CrV CsV

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Results reliability	Results validity	Commentary
	<p>(CrV) Concurrent validity: There was a strong association among all measures, with the 21-point NRS showing the most consistently strong (all r's > 0.80) association with the other measures. However, even the weakest association ($r = 0.59$), found between the NRS-11 and the FPS-7, indicated a great deal of overlap between these 2 measures.</p> <p>(CsV) Factoranalysis: A single factor emerged. The loadings, all 0.90 or greater (except NRS-11 0.80), support the validity of each of the scales as measures of pain intensity.</p> <p><i>Convergent validity:</i> Correlation coefficients between each of the 6 measures and measures of pain interference (a modified version of the Pain Interference Scale of the Brief Pain Inventory) and depression (Center for Epidemiological Studies–Depression scale (CES-D), were all in the expected direction, although some variability in the coefficients can be seen. The 7-point Faces scale appeared to be most strongly associated with these 2 measures, the NRS-11 and VRS-5 showed the weakest associations with pain interference, and the VRS-5 showed the weakest association with depression. The difference between the association between the NRS-11 and pain interference ($r = 0.25$) and the 7-point Faces scale ($r = 0.50$) was statistically significant ($t(42) = 2.46$, $p < 0.05$, for the difference between coefficients).</p>	<p>It is possible that the FPS-7 reflects other dimensions such as affect because of the strong correlation with depression.</p>

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Author (year)	Setting	Sample (n)	Design	Reliability	Validity
Kim, E. J. & Buschmann, M. T. (2006)	A general hospital and an oriental medical hospital in Korea.	85 older adults with chronic pain (i.e., a state of pain \geq 6 months duration and for which the cause of the pain could not be removed) were recruited. (n = 85)	Comparative study.		CrV

Betrouwbaarheid: Stability (S), Internal consistency (IC), Equivalence (E)

Validiteit: Face validity (FV), Content validity (CtV), Criterion validity (CrV), Construct validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR)

Results reliability	Results validity	Commentary
	<p>(CrV) Concurrent validity: Results indicate that pain intensities by the FPS, the NRS and the VAS were not significantly different ($F(1.00, 68.00) = 2.93, p = 0.09$); likewise the tools were not different in reporting the intensity of patient's pain. The NRS had moderately strong correlation with the FPS ($r = 0.73, p < 0.001$) and the VAS ($r = 0.91, p < 0.001$). Correlation between the VAS and the NRS of the cognitively impaired group ($r = 0.88, p < 0.001$) was weaker than the correlation of the cognitively intact group ($r = 0.92, p < 0.001$), correlation between the FPS and the NRS of the cognitively impaired group ($r = 0.75, p < 0.001$) was slightly stronger than correlation of the cognitively intact group ($r = 0.70, p < 0.001$).</p>	