The Faces Pain Scale (FPS)

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.

Meetinstrument	Faces Pain Scale
Afkorting	FPS
Auteur	Bieri, D., Reeve, R. A., Champion, G. D., Addicoat, L., & Ziegler, J. B.
Thema	Symptoommanagement pijn
Doel	Pijn beoordelen bij de patiënt
Populatie	Kinderen, volwassenen en ouderen
Afname	Zorgverlener
Aantal items	1 item
Aanwezigheid patiënt vereist	Ja
Vindplaats meetinstrument	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.

Doel

Pijncontrole bij de patiënt aan de hand van zelfrapportage op een één-item schaal die bestaat uit 7 gezichtsexpressies.

Doelgroep

De Faces Pain Scale (FPS) werd initieel ontwikkeld voor 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).

Latere studies rapporteren de validiteit van de FPS als pijnschaal bij volwassenen (Benaim et al., 2007; Carey, Turpin, Smith, Whatley & Haddox, 1997; Freeman, Smyth, Dallam & Jackson, 2001; Herr, Spratt, Mobily & Richarfdson, 2004; Jensen, Engel, McKearnan & Hoffman, 2003) alsook bij ouderen (Herr, Mobily, Kohout & Wagenaar, 1998; Herr et al., 2004; Kaasalainen & Crook, 2003; Kim & Buschmann, 2006; Taylor & Herr, 2003). Meer specifiek werd de FPS reeds meermaals

gevalideerd bij ouderen met een cognitief verminderd functioneren (Kaasalainen & Crook, 2003; Pautex et al., 2005; Pautex et al., 2006; Scherder & Bouma, 2000; Taylor & Herr, 2003; Ware, Epps & Packard, 2006).

Beschrijving

Het meetinstrument bestaat uit een lijn met 7 afbeeldingen die elk een gelaatsuitdrukking weergeven. Dit varieert tussen een neutrale gevoelsexpressie en een gelaatsuitdrukking die extreme pijn weerspiegelt. Elke gelaatsafbeelding is 6 cm hoog en de score is identiek aan het schaalnummer (0 - 6): afbeelding één komt overeen met een score nul, score zes wordt toegekend aan de gelaatsafbeelding die correspondeert met extreme pijn.

Varianten
Er bestaan tal van varianten op de FPS. Hicks et al. (2001) hebben de FPS herwerkt tot een schaal met 6 gelaatsuitdrukkingen i.p.v. 7. Dit moet vergelijkingen tussen de FPS en andere meetinstrumenten op een lineaire schaal toelaten. Om gelijkaardige redenen werden eveneens schalen ontwikkeld met 11 gevoelsexpressies (Kim & Buschmann, 2006; McGrath et al. in Scherder & Bouma, 2000).

Betrouwbaarheid

De betrouwbaarheid van de FPS werd reeds uitvoerig bestudeerd en is hoog.

Tal van studies hebben de betrouwbaarheid van de FPS onderzocht aan de hand van een *test – hertest* (Benaim et al., 2007; Bieri et al., 1990; Chibnall & Tait, 2001; Herr et al., 1998; Kaasalainen & Crook, 2003; Kim & Buschmann, 2006; Pautex et al., 2005; Pautex et al., 2006; Taylor & Herr, 2003; Ware et al., 2006; Wong & Baker, 1988). Met uitzondering van de studie van Chibnall & Tait (2001) worden telkens correlaties gerapporteerd hoger dan 0.70.

De *interrater reliability* is eveneens hoog (Benaim et al., 2007; Herr et al., 2004, Pautex et al., 2005; Pautex et al., 2006,). De correlatie varieert tussen $0.70 < r \leq 1.00$. In de studie van Benaim et al. (2007) bedroeg deze correlatie slechts $r = 0.64$ en $r = 0.44$ binnen een populatie van respectievelijk linker en rechter hemisfeer CVA-patiënten.

Validiteit

In het bestuderen van de validiteit van de FPS werd de *concurrent validity* meermaals onderzocht (Benaim et al., 2007; Bosenberg et al., 2003; Freeman et al., 2001; Hicks et al., 2001; Jensen et al., 2003; Kim & Buschmann, 2006; Paik & Ahn, 2002; Pautex et al., 2005; Taylor & Herr, 2003; Ware et al., 2006, Wong & Baker, 1988). De correlatie tussen de FPS en andere pijnschalen is overwegend sterk ($r > 0.70$), hoewel enkele studies een lagere correlatie rapporteren tussen de FPS en andere pijnschalen onderling (Bosenberg et al., 2003; Herr et al., 1998; Jensen & Karoly, 1992; Ware et al., 2006). Dit kan er mogelijks op wijzen dat de FPS een ander construct meet dan pijn.

Om de *construct validity* van de FPS na te gaan, hebben Herr et al. (2004) en Jensen et al. (2003) een *principale componenten analyse* uitgevoerd. Uit de pijnscores op de verschillende één-item meetinstrumenten werd 1 factor geabstraheerd. Een *factor-analyse* op de verschillende FPS-pijnscores uit herhaalde metingen gedurende 12 dagen, resulteerde eveneens in 1 factor (Chibnall & Tait, 2001).

Tevens werden significante correlaties berekend tussen de FPS en pijngerelateerde constructen zoals bijvoorbeeld gedragsobservaties (Kaasalainen & Crook, 2003, Stinson et al., 2006). De samenhang tussen beide varieerde tussen r = 0.49 en r = 0.90.

Tot slot blijkt de FPS ook sensitief te zijn voor pijnbestrijding d.m.v. analgetica (Bosenberg et al., 2003; Stinson et al., 2006) en verhoogde pijnprikkels (Herr et al., 2004).

Gebruiksvriendelijkheid

De FPS 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 FPS en de VAS. Omwille van deze reden verwijzen we naar de rubriek 'Gebruiksvriendelijkheid' van de VAS. Rekening houdend dat de FPS beter begrijpbaar is voor patiënten en 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 en dat afname van de FPS vlotter zal verlopen. Dit wordt trouwens ook bevestigd in het onderzoek van Ware et al. (2006).

Opmerkingen

Een vergelijking tussen de door ons geselecteerde pijnschalen, geeft aan dat de FPS aangewezen kan zijn bij jonge kinderen en bij ouderen.

In de studie van Scherder en Bouma (2000) bleek de voltallige ouderenpopulatie het gebruik van de FPS te begrijpen. Meerdere studies genereren ook gunstige resultaten ten aanzien van het gebruik van de FPS bij ouderen met een cognitief verminderd functioneren (Freeman et al., 2001; Pautex et al, 2006, Taylor & Herr, 2003, Ware et al., 2006). Dit geldt echter eveneens voor de Verbal Rating Scale (VRS). In de studies van Kaasalainen & Crook (2003) en Pautex et al. (2005) verdient de VRS zelfs de voorkeur op de FPS bij dementerenden.

In de review van Stinson & al. (2006) wordt aangegeven dat de FPS gehanteerd kan worden bij kinderen tussen 4 en 17 jaar. De auteurs van de FPS zelf (Bieri et al., 1990) stellen dat het meetsintrument adequaat bruikbaar is bij kinderen vanaf 3 jaar oud. Het gebruik van de FPS zou voornamelijk bij jonge kinderen (leeftijd 4 tot 12 jaar) aangewezen zijn (Champion et al. In Hicks et al., 2001; Stinson et al., 2005).

Bovendien verdient de FPS de voorkeur van de patiënt op andere meetinstrumenten (Benaim et al., 2007; Carey et al., 1997; Taylor & Herr, 2003; Ware et al., 2006;

Wong & Baker, 1988). Vergelijking van de betrouwbaarheids- en validiteitsresultaten duiden niet op de superioriteit van één bepaalde pijnschaal. Vandaar dat men kan opteren voor de FPS en dit in het bijzonder voor jonge kinderen, ouderen met een verhoogde leeftijd en dementerende personen.

Onenigheid bestaat er echter omtrent de validiteit van de FPS m.b.t. het meten van het construct pijn. Aangezien meerdere studies een lagere correlatie rapporten tussen de FPS en andere pijnschalen onderling (Bosenberg et al., 2003; Herr et al., 1998; Jensen & Karoly, 1992; Ware et al., 2006), reist de vraag of de gelaatsuitdrukkingen wel degelijk peilen naar pijn en niet naar gevoelsmatige aspecten. In de studie van Herr et al. (1998) werd bijvoorbeeld vastgesteld dat patiënten de gezichtsexpressies met pijn associëren, maar eveneens met droefheid, verveling, vermoeidheid en verbitterdheid (hoewel in mindere mate). In de studie van Bieri et al. (1990) en Kim & Buschmann (2006) associeerde respectievelijk 58% en 68% van de bevraagden de gelaatsuitdrukkingen met pijn. Jensen et al. (2003) kwam tot de vaststelling dat de FPS sterker gecorreleerd was met een schaal peilend naar depressie (Center for Epidemiological Studies Depression Scale), in vergelijking met andere pijnschalen.

Referenties

Benaim, C., Froger, J., Cazottes, C., Gueben, D., Porte, M., Desnuelle, C., & Pelissier, J. Y. (2007). Use of the Faces Pain Scale by left and right hemispheric stroke patients. *Pain, 128,* 52-58.

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.

Chibnall, J. T. & Tait, R. C. (2001). Pain assessment in cognitively impaired and unimpaired older adults: a comparison of four scales. *Pain, 92,* 173-186.

Freeman, K., Smyth, C., Dallam, L., & Jackson, B. (2001). Pain measurement scales: a comparison of the visual analogue and faces rating scales in measuring pressure ulcer pain. *J.Wound.Ostomy.Continence.Nurs., 28,* 290-296.

Herr, K. A., Mobily, P. R., Kohout, F. J., & Wagenaar, D. (1998). Evaluation of the Faces Pain Scale for use with the elderly. *Clin J Pain, 14,* 29-38.

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.

Jensen, M.P., & Karoly, P. (1992). Self-report scales and procedures for assessing pain in adults. In: D. Turk & R. Melzack (Eds.), *Handbook of pain assessment* (p. 135–51). New York: Guiford Press.

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.

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.

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

Pautex, S., Herrmann, F., Le Lous, P., Fabjan, M., Michel, J. P., & Gold, G. (2005). Feasibility and reliability of four pain self-assessment scales and correlation with an observational rating scale in hospitalized elderly demented patients. *J.Gerontol.A Biol.Sci.Med.Sci., 60,* 524-529.

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

Scherder, E. J. & Bouma, A. (2000). Visual analogue scales for pain assessment in Alzheimer's disease. *Gerontology, 46,* 47-53.

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

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.

THE FACES PAIN SCALE (FPS)

BIERI, D., REEVE, R. A., CHAMPION, G. D., ADDICOAT, L., & ZIEGLER, J. B. (1990)

Australia (English)

Author (year)	Setting	Sample (n)	Design	Reliability	Validity
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
Freeman, K., Smyth, C., Dallam, L., & Jackson, B. (2001)	Hospital.	Adults who had one or more stage 1 to 4 pressure ulcer, and had some ability to explain their pain experience. (n = 44)	Comparative study: the VAS was compared with the FPS.		CrV

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
(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: - SDS: 72.73% - NRS: 75.44% - FPS: 74.24% - Glasses: 75.38% - Chips: 77.27% - Colors: 68.18%	**(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: - SDS: 62.81% - NRS: 60.00% - FPS: 60.43% - Glasses: 63.70% - Chips: 69.06% - Colors: 58.39%	The FPS was the most preferred scale by all age groups. 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. No one scale demonstrates superiority in validity or reliability. No significant differences exist among the scales for any age group.
	(CrV) *Concurrent validity:* A mathematical transformation of the FPS in a numeric value was highly correlated with the VAS ($r = 0.92$). There was a significant increase in variability in VAS with increasing values of the FPS ($p < 0.05$).	

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
Pautex, S., Michon, A., Guedira, M., Emond, H., Le Lous, P., Samaras, D. et al. (2006)	Geriatrics hospital and a geriatric psychiatry service.	129 patients aged 65 and older hospitalized during a 15 month-period who met DSM criteria for dementia, with a Mini-Mental State Examination score less than 11 and a Clinical Dementia Rating of 3 or greater. (n = 129)	Repeated measures design. Patients were asked to indicate their current level of pain on different scales (VAS, Faces Pain Scale and Verbal Rating Scale). This was repeated 30 minutes later either by the same investigator (50% of the cases) or by a different examiner who was blinded to the first assessment.	S E	
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 FPS and NRS.	IC	
Hicks, C. L., von Baeyer, C. L., Spafford, P. A., van Korlaar, I, & Goodenough, B. (2001)	A children's hospital.	Children aged 4 to 12 and who were hospitalized for surgical treatment in 68 cases (75%) including abdominal (n = 18), ear/ nose/ throat (n = 12), orthopedic (n = 12), urological (n = 7), and other (n = 19). The remaining 22 cases (25%) were hospitalized for non-surgical painful conditions: abdominal (n = 5), respiratory (n = 5), orthopedic/ rheumatological (n = 4), and other (n = 8). (n = 90)	Validation study. To validate a revised version of the FPS (FPS-R) with 6 faces instead of 7.		CrV

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
(S) *Intra-rater reliability:* R = 0.93 (p < 0.001) **(E)** *Interrater reliability:* R = 0.94 (p < 0.001)	**(CrV)** *Concurrent validity:* - Correlation between FPS and different scales at first assessment: 1. Correlation FPS – Verbal Rating Scale: r = 0.80 (p<0.001) 2. Correlation FPS – VAS: r = 0.45 (p<0.001) 3. Correlation FPS – Doloplus 2 (observational pain scale): r = 0.36 (p<0.001) - Correlation between FPS and different scales at second assessment: 4. Correlation FPS – Verbal Rating Scale: r = 0.79 (p<0.001) 5. Correlation FPS – VAS: r = 0.66 (p<0.001) 6. Correlation FPS – Doloplus 2 (observational pain scale): r = 0.48 (p<0.001)	A significant better comprehension of the Verbal Rating Scale (VRS) and the Faces Pain Scale (FPS) scale was found. This might suggest that the VAS is not the most appropriate pain assessment scale for patients with severe dementia.
(IC) *Cronbach's alpha:* Intercorrelations between 3 scales was α = 0.88.		Patients selected the FPS as 'easiest to use' (48.6%), followed by the NRS (35.3%) and the VAS (16.1%).
	(CrV) *Concurrent validity:* The child was asked to estimate his or her current pain on the FPS-R, followed by either the VAS or the colored analogue scale (CAS). Each child was randomly assigned to use either the VAS or the CAS. Correlations between the FPS-R and the CAS and between the FPS-R and the VAS were respectively r = 0.84 and r = 0.92.	

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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)	Acute care facilities.	68 subjects aged 60 years and older whith 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

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
(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).	**(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).	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. 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. 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.

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
Pautex, S., Herrmann, F., Le Lous, P., Fabjan, M., Michel, J. P., & Gold, G. (2005)	The inpatient dementia consultation of the Geneva Geriatric Hospital.	Elderly who met DSM criteria for dementia. (n = 160)	Repeated measures design. Patients were asked to indicate their current level of pain on different scales (horizontal/ vertical VAS, Faces Pain Scale and Verbal Rating Scale). This was repeated 30 minutes later either by the same investigator (50% of the cases) or by a different examiner who was blinded to the first assessment.	S E	CrV

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
(S) *Intra-rater reliability FPS:* r = 0.97 (p < 0.001) **(E)** *Interrater reliability:* r = 0.71 (p < 0.001)	**(CrV)** *Concurrent validity:* - Correlation between FPS and different scales at first assessment: 1. Correlation FPS – Verbal Rating Scale: r = 0.89 (p<0.001) 2. Correlation FPS – horizontal VAS: r = 0.88 (p<0.001) 3. Correlation FPS – vertical VAS: r = 0.89 (p<0.001) 4. Correlation FPS – Doloplus 2 (observational pain scale): r = 0.34 (p<0.001) - Correlation between FPS and different scales at second assessment: 1. Correlation FPS – Verbal Rating Scale: r = 0.89 (p<0.001) 2. Correlation FPS – horizontal VAS: r = 0.87 (p<0.001) 3. Correlation FPS – vertical VAS: r = 0.90 (p<0.001) 4. Correlation FPS – Doloplus 2 (observational pain scale): r = 0.35 (p<0.001)	

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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) and 89 older adults (age 65-94). (n = 175)	Repeated measures design. Noxious heat stimili were delivered to the ventral forearm by an electronically controlled contact thermode. The heat stimili 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

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
(IC) *Internal consistency:* Intercorrelations between the scales were all statistically significant at every temperature ($p < 0.001$). *Cronbach's alpha within each scale across the 7 temperature values:* VAS: 0.88 NRS: 0.88 VDS: 0.86 VNS: 0.88 FPS: 0.88 *Cronbach's alpha within temperature across the 5 scales:* 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 **(E)** *Interrater reliability:* VAS: 93.5% agreement NRS: 100% agreement VDS: 100% agreement VNS: 100% agreement FPS: 100% agreement	**(CsV)** *Principal components analysis:* A one-factor model was extracted from the different measurement scales. The correlation of each scale to the isolated factor was as follows: VAS: 0.94 NRS: 0.96 VDS: 0.95 VNS: 0.95 FPS: 0.86 **(Sen)** Each tool demonstrated significant increases in score associated with increase in temperature ($p < 0.001$).	The VAS had a significant higher failure rate (6.7%) in comparison with the NRS, VDS, VNS, and FPS. The VNS demonstrated significantly higher levels of pain report than the other 4 scales. The psychometric scale evaluation was conducted using an experimental pain stimulus instead of using clinical pain stimili. The scales most preferred in order by the total sample: NRS (35.3%); VDS (25.3%); VNS (15.9%); FPS (12.9%); VAS (10.6%).

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
(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.	**(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 ($r = 0.30$, $p < 0.12$).	Test-retest for the 3 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 these tools. 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. For those with moderate cognitive impairment, the PPI seems to be a more appropriate and reliable tool than the FPS or the NRS.

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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)	Not specified.	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

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
(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).	**(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 ($r = 0.74$).	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. The FPS showed strong test-retest stability in the cognitively impaired minority older adults. 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.

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
Stinson, J. N., Kavanagh, T., Yamada, J., Gill, N., & Stevens, B. (2006)	Not appropriate (review).	Children aged 3 – 18 years.	Review on self report measures of single-item ratings of pain intensity for use in clinical trials in children and adolescents.	S	CrV CsV Sen

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
(S) *Test – retest reliability*: Both the Faces Pain Scale and the Faces Pain Scale-Revised have evidence of test–retest reliability. The Faces Pain Scale demonstrated adequate stability at a two-week interval (r = 0.79) in healthy children (Bieri et al. in Stinson et al., 2006), and at one and two days post-surgery in hospitalized children (Perrott et al. in Stinson et al., 2006). The Faces Pain Scale-Revised indicated adequate stability at one month following a surgical or non-surgical painful condition (r = 0.63) (Miro and Huguet in Stinson et al., 2006).	**(CrV)** *Concurrent validity:* Strong positive correlations (r = 0.59–0.90) have been found with the Faces Pain Scale and other well established self-report pain intensity measures (e.g. Pieces of Hurt tool, Oucher, Wong–Baker FACES Pain Scale) (Goodenough et al. in Stinson et al., 2006; Jacobson et al. in Stinson et al., 2006; Chambers et al. in Stinson et al., 2006). Similarly, the Faces Pain Scale-Revised has demonstrated strong positive correlations (r = 0.84 –0.92) with visual analogue scales (Hicks et al. in Stinson et al., 2006; Migdal et al. in Stinson et al., 2006). **(CsV)** *Convergent validity:* Moderate to strong positive correlations (r = 0.49–0.90) have been shown between the Faces Pain Scale and behavioural scales, such as the Children's Hospital of Eastern Ontario Pain Scale (Jacobson et al. in Stinson et al., 2006; Cassidy et al. in Stinson et al., 2006) and the Child Facial Coding System (Cassidy et al. in Stinson et al., 2006). **(Sen)** The Faces Pain Scale has demonstrated responsivity following procedural pain (Goodenough et al. in Stinson et al., 2006; Wolf et al. in Stinson et al., 2006) and the Faces Pain Scale-Revised has demonstrated responsivity following administration of lidocaine during venipuncture (Migdal et al. in Stinson et al., 2006; Taddio et al. in Stinson et al., 2006).	Disadvantages of the Faces Pain Scale and the Faces Pain Scale-Revised include the limited evidence regarding interpretability and mixed evidence regarding the acceptability of the scale with children. However, the Faces Pain Scale has been reported as being well accepted by children aged 4–17 years (Jacobson et al. in Stinson et al., 2006; Goodenough et al. in Stinson et al., 2006). Children as young as 3 years old have used the scale with adequate comprehension (Bieri et al. in Stinson et al., 2006).

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
Benaim, C., Froger, J., Cazottes, C., Gueben, D., Porte, M., Desnuelle, C., & Pelissier, J. Y. (2007)	2 rehabilitation units.	Patients who suffered a first unilateral middle cerebral artery stroke. A distinction was made between left and right hemispheric stroke patients (LHSP –RHSP). (n = 127)	5 year period prospective study. FPS was compared with vertical VAS and Verbal Ratings Scale (VRS).	S E	CrV
Paik, H. J. & Ahn, Y. M. (2002)	Not specified.	A convenient sample of 64 children who had undergone strabismus surgery (76.3%) or received lid surgery (23.4%). (n = 64)	Repeated measures design. Children were asked to express how much pain they experienced at 2, 4, 6, 8 and 24h after eye surgery using the FPS and a Numeric/ Word Graphic Rating Scale (NWGRS).		CrV Sen

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
(S) *Intra-rater reliability (n = 33):* Kappa coefficients for FPS were 0.74 (0.13) and 0.53 (0.10) in LHSP and RHSP, respectively. Kappa coefficients for VRS were 0.39 (0.14) and 0.57 (0.15) in LHSP and in RHSP, respectively. ICC for VAS were 0.78 (0.46–0.92) and 0.90 (0.74–0.96) in LHSP and in RHSP, respectively. **(E)** *Interrater reliability (n = 43):* Kappa coefficients for FPS were 0.64 (standard error = 0.11) and 0.44 (0.09) in LHSP and RHSP, respectively. Kappa coefficients for VRS were 0.46 (0.12) and 0.52 (0.12) in LHSP and in RHSP, respectively. ICC for VAS were 0.72 (95% CI = 0.44–0.88) and 0.86 (0.68–0.94) in LHSP and in RHSP, respectively.	**(CrV)** *Concurrent validity (n = 51):* LHSP scores on the FPS were highly correlated with VAS ($r = 0.82$, $p < 0.001$) and with VRS scores ($r = 0.65$, $p < 0.01$). In the RHSP group, correlations were also high ($r = 0.72, 0.72$; $p < 0.001$ respectively).	Among 71 patients, most LHSP preferred FPS (16/32) to VAS (6/32) and VRS (10/32), most RHSP preferred VAS (18/39) to FPS (11/39) and VRS (10/39). The difference was statistically significant ($p < 0.05$).
	(CrV) *Concurrent validity:* There was a high correlation in between pain measurements with the FPS and NWGRS over the 5 time points ($0.887 > r > 0.735$). **(Sen)** Two third of the children experienced pain equal to, or greater, than moderate to severe, and about one fifth of the subjects expressed the most severe pain at 2h after surgery . At 4h after surgery, 82.8% of the patients experienced ‘a little bit’ or ‘a little more’ pain or discomforts . At 1 day after surgery, only 34.4% of children were free of pain . The decrease in pain scores across the 5 time points was statistically significant ($F = 35.12$, $p < 0.001$)	

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
Bosenberg, A., Thomas, J., Lopez, T., Kokinsky, E., & Larsson, L. E. (2003)	The study was performed in two South African hospitals, one with a mainly rural population (King Edward VIII Hospital in Durban) and the other with an urban population (Red Cross Children's War Memorial Hospital in Cape Town).	A total of 110 children aged 4–12 years, scheduled for inguinal surgery in the two South African hospitals, were included in the study. (n = 110)	Repeated measures design. Postoperative pain assessments were made every hour for 8 h after the caudal block was performed. A designated nurse assessed pain by using a four-graded descriptive scale (no, mild, moderate or severe pain) and thereafter the child reported pain by using the six-graded faces pain scale.		CrV Sen
Hicks, C. L., von Baeyer, C. L., Spafford, P. A., van, K., I, & Goodenough, B. (2001)	Two urban jewelry stores.	Children aged 5 to 12. (n = 76)	Validation study. To validate a revised version of the FPS (FPS-R) with 6 faces instead of 7.		CrV

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
	(CrV) *Concurrent validity:* Comparison of pain ratings by the observer's assessment and the faces pain scale showed a high correlation ($r = 0.76$, $p < 0.0001$). The correlation between the nurses' observations and the faces scale was significant in both hospital populations ($r = 0.81$, $p < 0.0001$ in Cape Town, and $r = 0.53$, $p < 0.0001$ in Durban). **(Sen)** At the first analgesic administration, the median face score was 5 (2–6). After analgesic administration the median score was 1 (1–6) ($p < 0.0001$). The proportion of patients with pain scores above 2 was 66 of 77 (86%) before treatment. This proportion was significantly different compared with before and after treatment ($p < 0.001$). Also in the youngest children aged 4–5 years, the proportion of patients with pain scores >2 decreased from 83% to 33% after administration of analgesics ($p < 0.001$).	
	(CrV) *Concurrent validity:* A strong positive correlation ($r = 0.93$, $p < 0.001$) was found between the ratings of pain intensity on the VAS and FPS-R.	The psychometric scale evaluation was conducted using an experimental pain stimulus instead of using clinical pain stimili.

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
Chibnall, J. T. & Tait, R. C. (2001)	A proprietary subacute care facility.	Cognitively impaired and unimpaired older adults over 55 years. (n = 90)	Repeated measures design. Patients made ratings of current pain three times/day for 7 days. They also made retrospective daily, weekly, and bi-weekly ratings of usual, worst, and least pain levels over a 14-day period. Ratings were made on four different scales: a five-point verbal rating scale, a seven-point faces pain scale, a horizontal 21-point (0±100) box scale, and a vertical 21-point (0±20) box scale (measuring pain intensity).	S	CsV
Herr, K. A., Mobily, P. R., Kohout, F. J., & Wagenaar, D. (1998)	Not specified.	Cognitively intact non-institutionalized elderly aged 65 or older. (n = 168)	Validation study.	S	FV

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
(S) *Intra-class correlation:* FPS' reliability coefficients for cognitively unimpaired and impaired patients was respectively 0.45 < r < 0.70 and 0.15 < r < 0.50.	**(CsV)** *Factor analyse;* For each day, 12 day-retrospective ratings (usual, worst, least pain for each of four scales) were subjected to a principal-axis factor analysis. A single 'pain intensity' factor emerged and factor loadings for FPS ranged from 0.60 to 0.80.	The horizontal 21-point box scale emerged as the best scale with respect to both psychometrics and validity, regardless of mental status.
(S) *Test-retest reliability:* 41 subjects were instructed to remember a vividly remembered painful experience. Two weeks later, the subjects had to recall the same pain experience and rate it again with the FPS. The correlation coefficient was 0.94 (p = 0.01)	**(FV)** *Face validity:* 33 subjects were asked whether the faces represented 6 different constructs, including pain, sourness, sadness, anger, boredom, sleepiness on a 5 point Likert scale. Subjects agreed that the faces represented pain, but there was also some agreement that the faces could represent these other constructs (with the exception of anger), although at a lesser level.	Subjects were asked to place the 7 faces in order, from the most painful to the least painful expression. This resulted in a near perfect agreement.

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
Bieri, D., Reeve, R. A., Champion, G. D., Addicoat, L., & Ziegler, J. B. (1990)	26 schools: 1 Jewish school and 25 selected randomly from a listing of Catholic parish schools in a metropolitan area.	Children from grade 1 and grade 3. (n = 553)	Validation study.	S	FV

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
(S) *Test-retest reliability:* Subjects were instructed to remember a vividly remembered painful experience. Two weeks later, the subjects had to recall the same pain experience and rate it again with the FPS. The correlation coefficient was 0.79 for 6 year old children. Even when the rater varied, a high rank correlation coefficient was obtained (r = 0.82). Children were also asked to reconstruct a presented order of the set of faces either immediately or after a delay. The percentages of reconstructions were 50 (immediate) and 77.5 (delayed). When the faces were presented in random order, the correct recall of the presented order was achieved in 15.8% (immediate) and 0% (delayed) of the cases.	**(FV)** *Face validity:* Children were asked to the meaning of the faces. Clear statements of pain, hurt, ache, being sick, and of emotional pain such as from teasing were made by 92 children (57.9%). Forty-one (25.8%) gave other interpretations such as sadness, anger, boredom, crying for no reason. Sixteen children (10.1%) did not know what the faces showed and 10 (6.3%) could not be asked because of time limitations.	Children rank-ordered the faces. All 7 faces were correctly ranked by 64% of grade 1 children and by 86% of grade 3 children. When the faces were presented inn all possible paired combinations, 62% of the younger and 75% of the older subjects placed all 7 faces correctly.

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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 ≥ 6 months duration and for which the cause of the pain could not be removed) were recruited. (n = 85)	Repeated measures design.	S	FV CrV

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
(S) *Test-retest reliability:* In order to evaluate the test–retest reliability of the FPS, data collected at 2-week intervals were analyzed using Cohen's kappa and the Spearman's rank order correlation. Cohen's kappa in cognitively intact subjects was r = 0.61 (p < 0.001), indicating that this proportion of subjects consistently rated the same face on both the initial and the second ratings of pain intensity. The Spearman's rank order correlation coefficient for the test–retest was r = 0.60 (p = 0.004) in all subjects and r = 0.74 (p = 0.003) in the cognitively intact subjects.	**(FV)** *Face validity:* In order to determine construct validity of the FPS, the subjects were asked to rate their degree of agreement between the faces and a given feeling/emotion. Most subjects responded 'agreed' on the construct of pain (n = 21, 67.7%). The mean rating of pain is significantly different from the mean of sourness ($F(1,30) = 27.25$, $p < 0.001$), the mean of sleepiness ($F(1, 30) = 31$, $p < 0.001$), sadness ($F(1, 30) = 7.83$, $p = 0.009$), and boredom ($F(1, 30) = 13.87$, $p = 0.001$) and has marginally significant difference from anger ($F(1, 30) = 3.21$, $p = 0.083$). This means that the subjects perceived the FPS as representing pain rather than other constructs. **(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 FPS had moderately strong correlation with the NRS ($r = 0.73$, $p < 0.001$) and the VAS ($r = 0.73$, $p < 0.001$). 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 ($r = 0.70$, $p< 0.001$).	Subjects placed the faces accurate in rank of increasing pain. Face #9 was placed with the highest accuracy of 93.5% and face #4 was placed with 90.3% accuracy. Faces #6 and #7 were placed with the lowest accuracy of 61.3%. Kendall's W was 0.93 ($p < 0.001$), indicating that agreement on a rank order among subjects is near perfect, and the rank order the subjects produced would not simply have occurred by chance.

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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
	(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. **(CsV)** *Factorananalyse:* 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. *Convergent validity:* 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).	It is possible that the FPS-7 reflects other dimensions such as affect because of the strong correlation with depression.

Reliability: Stability (S), Internal consistency (IC), Equivalence (E)
Validity: 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)

Faces Pain Scale - Revised (FPS-R)

Bron: 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.

Faces Pain Scale – Revised (FPS-R)

From *Pediatric Pain Sourcebook*, **www.painsourcebook.ca**
Version: 7 Aug 2007 CL von Baeyer

In the following instructions, say "hurt" or "pain," whichever seems right for a particular child.

"These faces show how much something can hurt. This face *[point to left-most face]* **shows no pain. The faces show more and more pain** *[point to each from left to right]* **up to this one** *[point to right-most face]* **– it shows very much pain. Point to the face that shows how much you hurt** [right now]**."**

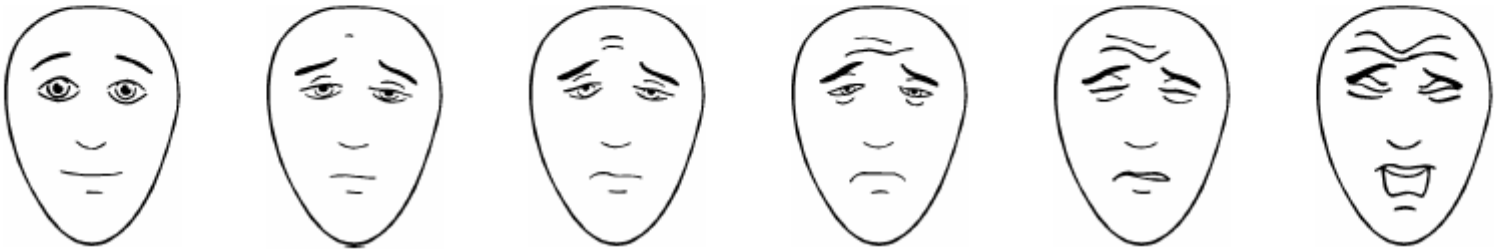
Score the chosen face 0, 2, 4, 6, 8, or 10, counting left to right, so '0' = 'no pain' and '10' = 'very much pain.' Do not use words like 'happy' and 'sad'. This scale is intended to measure how children feel inside, not how their face looks.

Permission for use. Copyright in the FPS-R is held by the International Association for the Study of Pain (IASP) © 2001. This material may be photocopied for non-commercial clinical and research use. To request permission from IASP to reproduce the FPS-R in a publication, or for any commercial use, please e-mail **iaspdesk@iasp-pain.org** For all other information regarding the FPS-R contact **Tiina.Jaaniste@sesiahs.health.nsw.gov.au** (Pain Medicine Unit, Sydney Children's Hospital, Randwick NSW 2031, Australia).

Sources. Hicks CL, von Baeyer CL, Spafford P, van Korlaar I, Goodenough B. The Faces Pain Scale – Revised: Toward a common metric in pediatric pain measurement. *Pain* 2001;93:173-183. Bieri D, Reeve R, Champion GD, Addicoat L, Ziegler J. 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* 1990;41:139-150.

Fold here

0 2 4 6 8 10



Reprinted with permission from the International Association for the Study of Pain (IASP).

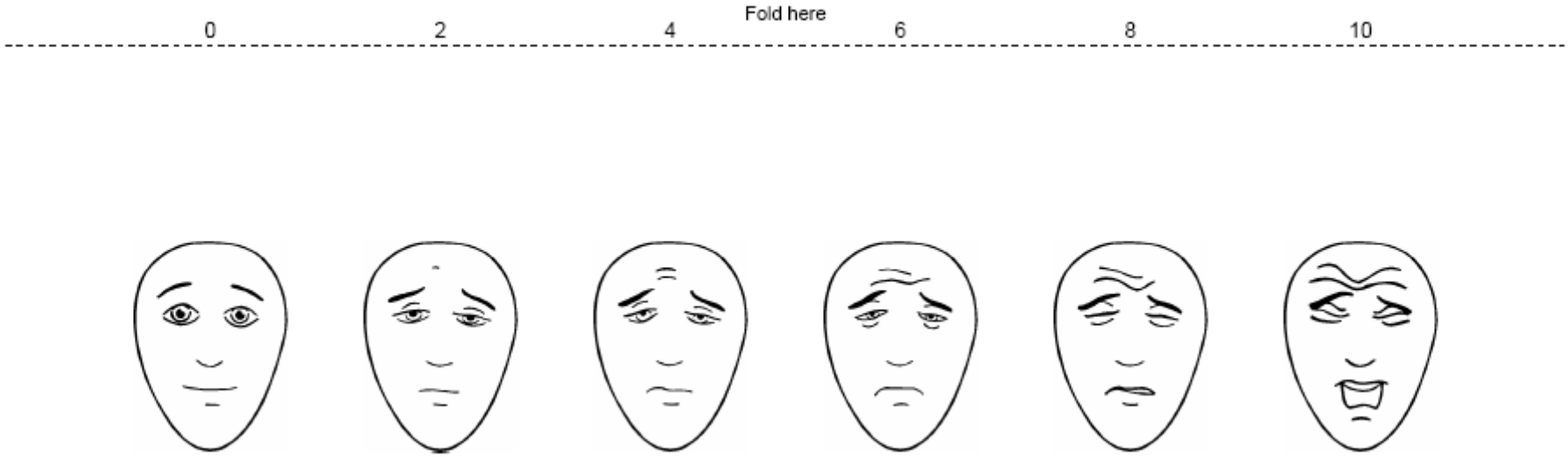
Vertaling Faces Pain Scale - Revised (FPS-R)

Dutch

"Gebruik bij de onderstaande instructie 'au' of 'pijn' al naar gelang het kind waar het om gaat. "Aan deze gezichtjes kun je zien hoe veel pijn je kan hebben. Aan dit gezichtje [wijs het meest linkse gezichtje aan] zie je geen pijn. Je ziet steeds meer pijn aan de gezichtjes [wijs ze aan van links naar rechts] tot aan deze [wijs het meest rechtse gezichtje] - aan deze zie je heel veel pijn. Kun je het gezichtje aanwijzen dat laat zien hoe veel pijn je voelt [op dit moment]?"Scoor het gekozen gezichtje 0, 2, 4, 6, 8, of 10, van links naar rechts rekenend, dus '0' = 'geen pijn'en '10'= 'heel veel pijn' Gebruik geen woorden zoals 'blij'en 'bedroefd'. Deze schaal is bedoeld om te meten hoe kinderen zich van binnen voelen, niet hun gezichtsuitdrukking.

Translation credit: Ko Hagoort and Monique van Dijk, Sophia Children's Hospital Rotterdam, The Netherlands

---



Reprinted with permission from the International Association for the Study of Pain (IASP).

Wat is BEST?

BEST staat voor BElgian Screening Tools en is een studie uitgevoerd door de Universiteit Gent, afdeling Verplegingswetenshap in opdracht van de Federale Overheidsdienst Volksgezondheid, Veiligheid van de voedselketen en Leefmilieu, en in samenwerking met de Université Catholique de Louvain.

Doel van BeST

Bedoeling van het project is het oprichten van een databank met wetenschappelijk gevalideerde meetinstrumenten. Met het oog op het objectiveren van de diagnostiek en van de resultaten van verpleegkundige interventies, zijn valide en betrouwbare meetinstrumenten een basisvoorwaarde om effectieve verpleegkundige zorg te kunnen bieden. Onze aandacht gaat uit naar meetinstrumenten voor de verpleegkundige interventies die bij de Minimale Verpleegkundige Gegevens gescoord worden.

Wat kan u vinden in dit rapport?

In dit rapport wordt de inhoud van het project alsook de gehanteerde methodologie beschreven. Vervolgens worden de verschillende meetinstrumenten per thema besproken. Bovendien wordt het instrument ter beschikking gesteld indien we hiertoe toestemming verkregen. Meetinstrumenten met een hoge betrouwbaarheid en validiteit werden tevens naar het Nederlands en het Frans vertaald.

Projectleiders UGent:
Prof. dr. T. Defloor
Prof. dr. M. Grypdonck

Projectmedewerkers UGent:
M. Daem
Dr. K. Vanderwee

Projectleider UCL:
Dr. M. Gobert

Projectmedewerkers UCL:
C. Piron

Projectleider FOD:
B. Folens

Projectmedewerkers FOD:
M. Lardennois

Gelieve bij elk gebruik van dit rapport als volgt te refereren:

Daem, M., Piron, C., Lardennois, M., Gobert, M., Folens, B., Vanderwee, K., Grypdonck, M., & Defloor T. (2007). Opzetten van een databank met gevalideerde meetinstrumenten: BEST-project. Brussel, Federale Overheidsdienst Volksgezondheid, Veiligheid van de voedselketen en Leefmilieu.