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.

Instrument de mesure	Faces Pain Scale
Abréviation	FPS
Auteur	Bieri, D., Reeve, R. A., Champion, G. D., Addicoat, L., & Ziegler, J. B.
Thème	Management des symptômes de la douleur
Objectif	Evaluer la douleur chez le patient
Population	Enfants, adultes et personnes âgées
Relevé	Dispensateur de soins
Nombre d'items	1 item
Présence du patient requise	Oui
Localisation de l'instrument de mesure	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.

Objectif

Contrôle de la douleur chez le patient à l'aide d'un autorapport sur une échelle à un seul item se composant de 7 expressions faciales.

Groupe cible

La Faces Pain Scale (FPS) a été initialement conçue pour des enfants (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).

Des études ultérieures font rapport de la validité de la FPS comme échelle de douleur chez l'adulte (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), ainsi que chez les personnes

âgées (Herr, Mobily, Kohout & Wagenaar, 1998; Herr et al., 2004; Kaasalainen & Crook, 2003; Kim & Buschmann, 2006; Taylor & Herr, 2003). Plus spécifiquement, la FPS a déjà été validée à plusieurs reprises chez des personnes âgées avec un fonctionnement cognitif amoindri (Kaasalainen & Crook, 2003; Pautex et al., 2005; Pautex et al., 2006; Scherder & Bouma, 2000; Taylor & Herr, 2003; Ware, Epps & Packard, 2006).

Description

L'instrument de mesure se compose d'une ligne avec 7 illustrations représentant chacune une expression faciale. Cela varie entre une expression neutre et une expression faciale traduisant une douleur extrême. Chaque illustration fait 6 cm de hauteur et le score est identique au numéro d'échelle (0 - 6) : l'illustration un correspond à un score nul, le score six est attribué à l'illustration qui correspond à une douleur extrême.

Variantes

Il existe de nombreuses variantes à la FPS. Hicks et al. (2001) ont retravaillé la FPS en une échelle comptant 6 expressions faciales au lieu de 7. Cela doit permettre des comparaisons entre la FPS et d'autres instruments de mesure sur une échelle linéaire. Pour des raisons similaires, des échelles ont également été conçues avec 11 expressions faciales (Kim & Buschmann, 2006; McGrath et al. dans Scherder & Bouma, 2000).

Fiabilité

La fiabilité de la FPS a déjà été amplement étudiée et est élevée.

De nombreuses études se sont penchées sur la fiabilité de la FPS à l'aide d'un *test – retest* (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). A l'exception de l'étude de Chibnall & Tait (2001), des corrélations supérieures à 0.70 sont chaque fois rapportées.

L'*interrater reliability* est également élevée (Benaim et al., 2007; Herr et al., 2004, Pautex et al., 2005; Pautex et al., 2006,). La corrélation varie entre $0.70 < r \leq 1.00$. Dans l'étude de Benaim et al. (2007), cette corrélation n'atteignait que $r = 0.64$ et $r = 0.44$ au sein d'une population de respectivement patients AVC hémisphère gauche et droit.

Validité

Dans l'étude de la validité de la FPS, la *concurrent validity* a été étudiée à plusieurs reprises (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). La corrélation entre la FPS et d'autres échelles de douleur est principalement élevée ($r > 0.70$), bien que quelques études fassent état d'une corrélation inférieure entre la

FPS et d'autres échelles de douleur entre elles (Bosenberg et al., 2003; Herr et al., 1998; Jensen & Karoly, 1992; Ware et al., 2006). Cela indique peut-être que la FPS mesure un autre élément que la douleur.

Afin de vérifier la *construct validity* de la FPS, Herr et al. (2004) et Jensen et al. (2003) ont réalisé une *analyse des principaux composants.* Au départ des scores de douleur sur les différents instruments de mesure à un seul item, 1 facteur a été abstrait. Une *analyse des facteurs* sur les différents scores de douleur FPS issus de mesures répétées pendant 12 jours a également débouché sur 1 facteur (Chibnall & Tait, 2001).

Des corrélations significatives ont également été calculées entre la FPS et des composants relatifs à la douleur, comme par exemple des observations comportementales (Kaasalainen & Crook, 2003, Stinson et al., 2006). La cohésion entre les deux varie entre r = 0.49 et r = 0.90.

Pour conclure, il s'avère que la FPS est également sensitive au traitement antidouleur au moyen d'analgésiques (Bosenberg et al., 2003; Stinson et al., 2006) et aux sensations douloureuses accrues (Herr et al., 2004).

Convivialité

La FPS n'a pas été testée sur le plan de sa convivialité, étant donné que le management de la douleur dans les hôpitaux que nous avons interrogés est étudié à l'aide de la VAS. Nous pensons toutefois pouvoir dire qu'il existe un chevauchement important entre la convivialité de la FPS et de la VAS. Pour ces raisons, nous faisons référence à la rubrique « Convivialité » de la VAS. En tenant compte du fait que la FPS est plus facilement compréhensible pour les patients et que l'écart entre le point zéro et le marquage du patient ne doit pas être mesuré (ce contrairement à la VAS), nous pouvons partir du principe que moins d'erreurs seront commises et que le relevé de la FPS se fera plus rapidement. Cela est d'ailleurs également confirmé dans l'étude de Ware et al. (2006).

Remarques

Une comparaison entre les échelles de douleur que nous avons sélectionnées montre que la FPS peut être indiquée chez les jeunes enfants et chez les personnes âgées.

Dans l'étude de Scherder et Bouma (2000), toute la population de personnes âgées s'est avérée comprendre l'utilisation de la FPS. Plusieurs études génèrent aussi des résultats positifs concernant l'utilisation de la FPS chez les personnes âgées avec un fonctionnement cognitif amoindri (Freeman et al., 2001; Pautex et al, 2006, Taylor & Herr, 2003, Ware et al., 2006). Cela vaut cependant aussi pour la Verbal Rating Scale (VRS). Dans les études de Kaasalainen & Crook (2003) et Pautex et al. (2005), la VRS mérite même la préférence sur la FPS chez les déments.

Le review de Stinson & al. (2006) indique que la FPS peut être utilisée chez les enfants âgés de 4 à 17 ans. Les auteurs de la FPS eux-mêmes (Bieri et al., 1990) déclarent que l'instrument de mesure peut être adéquatement utilisé chez les enfants

à partir de l'âge de 3 ans. L'utilisation de la FPS serait surtout indiquée chez les jeunes enfants (âgés de 4 à 12 ans) (Champion et al. In Hicks et al., 2001; Stinson et al., 2005).

De plus, la FPS mérite la préférence du patient sur d'autres instruments de mesure (Benaim et al., 2007; Carey et al., 1997; Taylor & Herr, 2003; Ware et al., 2006; Wong & Baker, 1988). Une comparaison des résultats de fiabilité et de validité ne montre aucune supériorité d'une échelle de douleur donnée. C'est pourquoi on peut opter pour la FPS, et ce en particulier chez les jeunes enfants, les personnes d'un âge avancé et les déments.

Il existe un désaccord quant à la validité de la FPS en rapport avec la mesure de l'élément douleur. Etant donné que plusieurs études font état d'une moindre corrélation entre la FPS et d'autres échelles de douleur entre elles (Bosenberg et al., 2003; Herr et al., 1998; Jensen & Karoly, 1992; Ware et al., 2006), on se demande si les expressions faciales jugent bien la douleur, et non des aspects émotionnels. Dans l'étude de Herr et al. (1998), il a par exemple été constaté que des patients associent bien les expressions faciales à la douleur, mais aussi à la tristesse, à l'ennui et à l'amertume (bien que dans une moindre mesure). Dans l'étude de Bieri et al. (1990) et Kim & Buschmann (2006), respectivement 58% et 68% des personnes interrogées ont associé les expressions faciales à la douleur. Jensen et al. (2003) en sont arrivés à la constatation que la FPS était plus nettement corrélée à une échelle évaluant la dépression (Center for Epidemiological Studies Depression Scale), en comparaison avec d'autres échelles de douleur.

Références

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.

Localisation de l'instrument de mesure

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)

Qu'est-ce que BEST ?

BEST pour Belgian Screening Tools est le nom d'une étude réalisée par l'Université de Gand, service des Sciences Infirmières, à la demande du Service Public Fédéral de la Santé Publique, Sécurité Alimentaire et Environnement.

Objectif de BEST ?

Le but de ce projet est de construire une base de données contenant des instruments de mesures validés scientifiquement. Dans le but d'objectiver les diagnostics et résultats des interventions infirmières, des instruments de mesures fiables et valides doivent être disponibles pour démontrer l'efficience des soins infirmiers.

Notre attention se porte sur les instruments de mesure utilisables pour scorer les interventions infirmières du nouveau Résumé Infirmier Minimum ou DI-RHM.

Que pouvez-vous trouver dans ce rapport ?

Le rapport décrit les différents instruments de mesure. En plus, si nous en avons reçu l'autorisation des auteurs, l'instrument est mis à votre disposition. Les instruments de mesure présentant une fiabilité et une validité élevées ont également fait l'objet d'une traduction vers le néerlandais et le français.

Les chefs de projet UGent
Prof. dr. T. Defloor
Prof. dr. M. Grypdonck

Les collaborateurs du projet UGent
M. Daem
Dr. K. Vanderwee

Le chef de projet UCL
Dr. M. Gobert

Le collaborateur du projet UCL
C. Piron

Le chef de projet FOD
B. Folens

Le collaborateur du projet FOD
M. Lardennois

Daem, M., Piron, C., Lardennois, M., Gobert, M., Folens, B., Spittaels, H., Vanderwee, K., Grypdonck, M., & Defloor T. (2007). Mettre à disposition une base de données d'instruments de mesure validés: le projet BEST. Bruxelles: Service Publique Fédéral Santé Publique, Sécurité de la Chaîne alimentaire et Environnement.