European Journal of Pain 13 (2009) 641–648

Contents lists available at ScienceDirect

European Journal of Pain

journal homepage: www.EuropeanJournalPain.com

How many response levels do children distinguish on faces scales for pain assessment?

Céline Decruynaere *, Jean-Louis Thonnard, Léon Plaghki

Unité de Réadaptation et Médecine Physique (READ 5375), Faculté de médecine, Université catholique de Louvain, Tour Pasteur, Avenue Mounier, 53, 1200 Brussels, Belgium

ARTICLE INFO

Article history:
Received 5 September 2007
Received in revised form 18 June 2008
Accepted 13 July 2008
Available online 2 September 2008

Keywords:
Faces scale
Children
Rasch analysis
Pain intensity

ABSTRACT

Faces scales are one of the most commonly used instruments to assess pain intensity in children. Most available faces scales present five to seven faces. The present research was conducted to investigate the ability of 4- to 7-year-old children to distinguish the response categories of different faces scales. In the first study, 121 children were asked to rate painful situations on a 3- and a 6-level faces scale commencing with a smiling 'no pain' face. Children were divided into two age groups (4–5 and 6–7 years). Investigations of the category functioning were performed with a rating scale Rasch model for each age group. Results revealed the low performances of the 6-level faces scale as compared to the 3-level faces scale and also the difficulty children experienced in scoring the imaginary painfulness of items. Consequently, a second study was conducted. In this second study, 76 children were asked to rate pictures depicting painful situations on a 3-level faces scale beginning with a neutral 'no pain' face. Results of this second study confirmed an improvement in the ability to distinguish the three response categories with age. The 4–5 year-old children could only distinguish two response categories and the 6–7 year-old children were able to discern the three levels of the 3-level faces scale. In conclusion, young children do not distinguish as many faces as proposed by the majority of available faces scales. These results strongly recommend a reduction in the number of response levels of faces scales for pain assessment in children.

1. Introduction

Faces scales are self-report scales using pictures of facial expressions to assess pain intensity. They show a series of faces conveying increasing intensity of pain between 'no pain' and 'worst pain possible' (for a review, see Chambers et al., 1999). Faces scales do not require the ability to count or to use numbers in an ordinal fashion. For this reason, they are often used with children as young as 3 years old (Chambers et al., 1999). Many different face scales have been created varying, among other criteria, by the number of faces. Most faces scales have five to seven faces (Beyer, 1984; LeBaron and Zeltzer, 1984; Wong and Baker, 1988; Kuttner and LePage, 1989; Bieri et al., 1990). This categorization of pain intensity variable is intended to elicit an ordinal indication about the intensity of pain. It is supposed to give more information than expected from a simple binary 'pain'/'no pain' response. However, whether children use faces scales effectively and appropriately is still a matter of debate (Bieri et al., 1990; Yeh, 2005; Stinson et al., 2006).

The scale developer may have given more categories than a child can distinguish possibly leading to confusion in a child's mind. Some studies have already reported difficulties in differentiating between too many faces. Pesudovs and Noble (2005) analyzed adults' responses on a 7-category faces scale and showed improvement in category usage when reducing the number of categories to five or six. Hunter et al. (2000) reported that children aged from 3 to 6 years frequently confused faces near the middle of a 7-level faces scale and proposed to reduce the number of faces to six.

An innovative approach to investigate a subject's capacity to distinguish the different categories on rating scales is the use of one member of the family of Rasch models, the rating scale model (Masters, 1982). The Rasch family of models belongs to a group of probabilistic models known as Item Response Theory. Usually applied to measure human performances, attitudes and perceptions (Tesio, 2003), the rating scale analysis also provides an effective framework within which to verify and improve the functioning of rating scale categorization (Linacre, 2002).

The present studies are aimed at examining with the rating scale model how a sample of healthy children aged from 4 to 7 distinguish the different faces of faces scales when rating imaginary painful situations. In a first study, children's ability to discriminate a 3- and a 6-level faces scale was tested. A 6-level faces scale was used because it corresponds to the average number of faces on the majority of faces scales (Chambers et al., 1999) and a 3-level faces scale because the aforementioned studies emphasized the diffi-

* Corresponding author. Tel.: +32 2 764 53 74; fax: +32 2 764 53 60.
E-mail address: cdecruynaere@parnasse-deuxalice.edu (C. Decruynaere).

1090-3801/$36.00
doi:10.1016/j.ejpain.2008.07.004

culty to differentiate many response categories. As results of the first study revealed the low performance of the 6-level faces scale but also the difficulty children experienced in scoring the imaginary painful situations, a second study was conducted using only a 3-level faces scale and cartoons depicting painful situations in order to improve children's representation of situations.

2. Study 1

2.1. Methods

2.1.1. Subjects

A total of 121 healthy children (89 boys and 32 girls) participated in this experiment. As we were primarily concerned by the precision of the scale, the number of children was based on the recommendations given by Linacre (1994). They ranged in age from 4 to 7 years. They were recruited from various kindergartens, primary schools and sport centres in Belgium (French part of the country). After having received the authorization of the director of each kindergarten, primary school or sport centre, permission to assess children was obtained via written informed consent of the parents according to the rules of the Ethical Committee of the Faculty of Medicine of the Université catholique de Louvain.

The sample of children was divided into two age categories: 4 to 5 years (preschool), $n = 72$ (46 boys, 26 girls) and 6 to 7 years (primary school), $n = 49$ (43 boys, 6 girls) in order to examine age-related differences in the way children use faces scales. This stratification is consistent with school levels in Belgium and with the fact that children less than six years of age typically fail standard Piagetian tasks that their counterparts over six typically succeed in performing (Gelman and Baillargeon, 1983). The way children define pain and think about several aspects of pain has been shown to be consistent with the Piagetian developmental model (Gaffney and Dunne, 1986). No finer division into age groups was performed in order to keep sufficiently large groups for the subsequent Rasch analysis.

2.1.2. Faces scales

Two different faces scales were presented to children: a 3-level faces scale and a 6-level one (Fig. 1a and b). They were similar, although not identical, in face design to those previously used in pain research (e.g., Wong and Baker, 1988) and conveyed increasing levels of pain intensity, from no pain to extreme pain. Faces scales lay on cards put on the desk in front of the child. Faces were 4 cm from top to bottom.

We opted for a scale with smiling face as no pain anchor although some studies (Chambers et al., 1999; Chambers et al., 2005) reported a bias effect. This choice was justified for the following reasons: (1) the scales with smiling faces are still in use in clinical settings (e.g., Zempsky et al., 2008), (2) they allowed to increase the discriminability between adjacent categories especially for the youngest and (3) to reinvestigate the validity of the smiling face category especially for the youngest children.

a
b
c

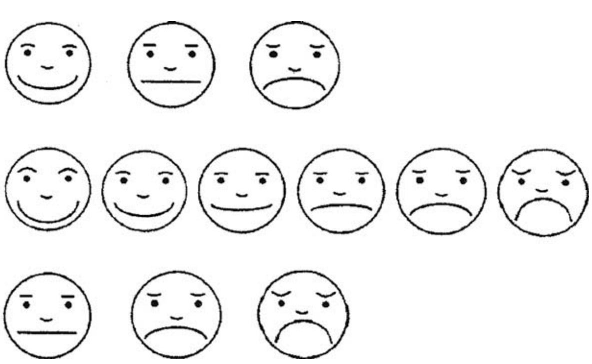

Fig. 1. The faces scales. (a) The faces of the 3-level faces scale used in Study 1 from face 0 (no pain) on the left to face 2 (extreme pain) on the right. (b) The faces of the 6-level faces scale used in Study 1 from face 0 (no pain) on the left to face 5 (extreme pain) on the right. (c) The faces of the 3-level faces scale used in Study 2 from face 0 (no pain) on the left to face 2 (extreme pain) on the right.

2.1.3. Procedure

All participants were tested individually in a quiet room. The child was sitting comfortably at a desk opposite the experimenter. Before testing, a few minutes were spent in order to meet the child and make sure that he had correct notions of pain. In this regard, the child was asked to tell about a personal and recent painful experience. Children who obviously did not understand the notion of pain (e.g. children pretending they never had experienced pain) were excluded from the study. Then, the experimenter explained the task to be presented. A series of 25 painful situations (called the items) developed for the purpose of the study (Table 1, upper list) were presented and described. Most painful situations were inspired from the Situational Pain Questionnaire (Yang et al., 1983). Others were devised to be more adapted to children. Similar painful situations have already been presented to children of the same age to investigate the validity of pain scales (McGrath et al., 1996; Hunter et al., 2000). Situations were presented in random order. The child was asked to choose the face that best reflected the amount of pain he would feel in each situation. Children rated the entire list of situations twice, once on the 3-level faces scale and once on the 6-level faces scale. The order of scale presentation was random. The entire session lasted approximately 40 min.

2.1.4. Rasch analysis

Responses were analysed with the Winsteps© Rasch analysis software (Linacre and Wright, 1998) using a rating scale model (Masters, 1982; Wright and Masters, 1982). The rating scale model is an extension of Rasch's original dichotomous model (e.g., painful/not painful) (Rasch, 1960) for polytomous response formats, i.e. for items graded on more than two response categories. For each child responding to each imaginary situation, the model formulates the probability of selecting each response category. By means of an iterative process, the rating scale model estimates a measure of pain representation for each child, the painfulness of each situation and the location of each threshold. Thresholds are defined as the location along the pain representation variable for which two adjacent response categories have an equal probability of occurrence and thus delineate the boundaries between the categories. In the rating scale model, thresholds have the same relative location from item to item. A single rating scale structure is thus modeled. All these measures are located on the same scale and expressed in logits, the natural logarithm of an odds ratio. The left panel of Fig. 2 displays the different concepts here above mentioned. The chart illustrates the category probability curves for a given situation rated on a 6-level scale. The x-axis represents the pain representation continuum and the y-axis the probability of each response category. Each curve corresponds to one response category. Subjects located on the left of the scale, with low pain representations, have category 0 as the most probable. As a subject's pain representation rises, each response category becomes in his turn the most probable. Thresholds correspond to the projection, on the x-axis, of intersections of successive category probability curves. Threshold one and four, symbolized τ_1 and τ_4 are shown as examples. Distance between two successive thresholds delimits the range for which one category is the most probable. The response scale can be represented like a ruler (bottom of Fig. 2).

Several criteria have been proposed to investigate the functioning of rating scale categories (Linacre, 2002). The criterion retained

Table 1
List of situations presented in Study 1 and 2

List of the 25 painful situations of Study 1	
1. A mosquito bite	14. Someone pulls your hair
2. Getting sunburned on your face	15. Getting your fingers caught in the car door
3. A toothache	16. Knocking the head on the corner of a piece of furniture
4. Walking barefoot on pebbles	17. Getting indigestion after a large meal
5. Cleaning a scratch	18. Catching your fingers in a zipper
6. The doctor examine your throat	19. Getting sunburned and being touched on that spot
7. Being stung by a bee	20. The dentist examines your teeth in a check-up
8. Getting a speck of dust in your eye	21. Cutting your hand with a knife
9. Burning your tongue tasting scorching hot food	22. Stubbing your toe on a chair leg
10. Walking on burning sand	23. Biting your tongue
11. Getting your fingers frozen	24. Carrying a heavy bag
12. Getting shampoo in your eye	25. Spilling some boiling water on your hands
13. Getting an injection at the doctor's	
List of the 15 cartoon situations of Study 2	
1. A mosquito bite	9. Getting indigestion after a large meal
2. Walking barefoot on pebbles	10. Cutting your hand with a knife
3. Cleaning a scratch	11. Someone stand on your foot
4. The doctor examine your throat	12. To lose a tooth
5. Getting shampoo in your eye	13. To put your hand on a hotplate
6. Getting an injection at the doctor's	14. To take a bath
7. Someone pulls your hair	15. To play football (boys)
8. Getting your fingers caught in the car door	15. To skip with a rope (girls)

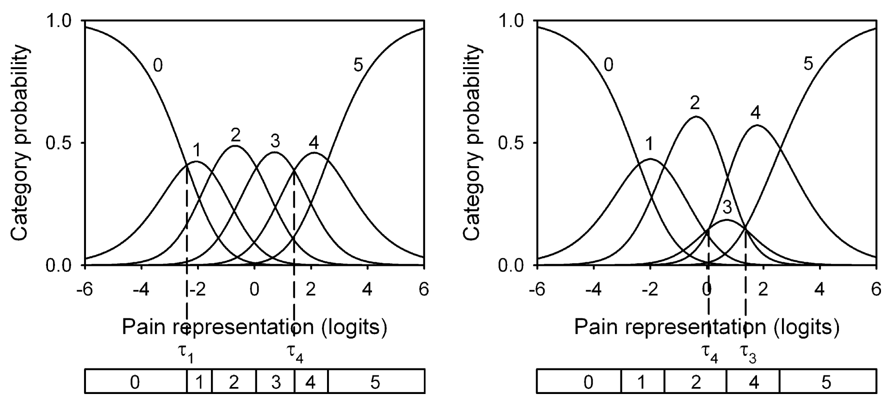

Fig. 2. Category probability curves for a given situation indicating the probability of each response category as a function of the subject's representation of pain. Thresholds (τ) are located at the pain representation level for which pairs of successive responses are equally probable. The bottom ruler is a synthetic representation of the response scale. Left panel: example of a well-functioning response scale. Thresholds 1 (τ_1) and 4 (τ_4) are displayed as example (broken lines). Right panel: example of a bad-functioning response scale. Category 3 never emerges. The probability of observing this category is lower than the probability of observing any other category whatever the subject's location along the scale. Thresholds delimiting categories 3 and 4 (τ_3) and 4 and 5 (τ_4) are reversed.

for the present study is the monotonic progression of thresholds, implying that the category probability curves form a 'range of hills' such as illustrated in the left panel of Fig. 2. An example of threshold disordering is illustrated in the right panel of Fig. 2. The threshold between category 3 and 4 (τ_3) has a higher location along the pain representation scale than the threshold between category 4 and 5 (τ_4). Probability curves show that the fourth category is never the most likely to be selected. Disordered thresholds reveal that the categories are not functioning as intended.

2.2. Results

Separate Rasch analyses were conducted for each age group and for each faces scale. Distributions of child measures and item (or situation) locations for both scales are presented in Fig. 3 (four left panels). Children reporting the highest pain scores are located on the right of the scale. Highest painful items are located on the left of the scale. Two major observations emerge. First, the range of item painfulness is extremely narrow for both scales and for both age groups reflecting the inability of children to differentiate situations according to their painfulness. For the 3-level faces scale, situation locations spread over 1.10 logit for the 4–5 year olds and over 0.97 logit for the 6–7 year olds. This effect is amplified with the 6-level faces scale: situation locations spread over 0.28 logit for the 4–5 year olds and over 0.43 logit for the 6–7 year olds. Second, children are located more on the right of the scale as compared to the item locations, reflecting children's tendency to favor high pain scores. This is confirmed by the percentage utilization of each response category, that is the percentage of occurrences of each category, presented in Table 2. Except for the first response category of the 6-level faces scale, both faces scales present percentage utilization that increases together with the response categories. This phenomenon is greater for the younger age group than for the older one. Notwithstanding these two observations, child's measures on both scales are coarsely correlated (ρ = 0.70, p < 0.001 for the 4–5 years and ρ = 0.60, p < 0.001 for the 6–7 years), supporting the idea that both scales point in the same direction and that children do not score in a random fashion.

The examination of category functioning is presented in Fig. 4. For the 6-level faces scale, the category probability curves display multiple threshold disorders regardless of the age category. On the other hand, the analysis of responses shows an improvement

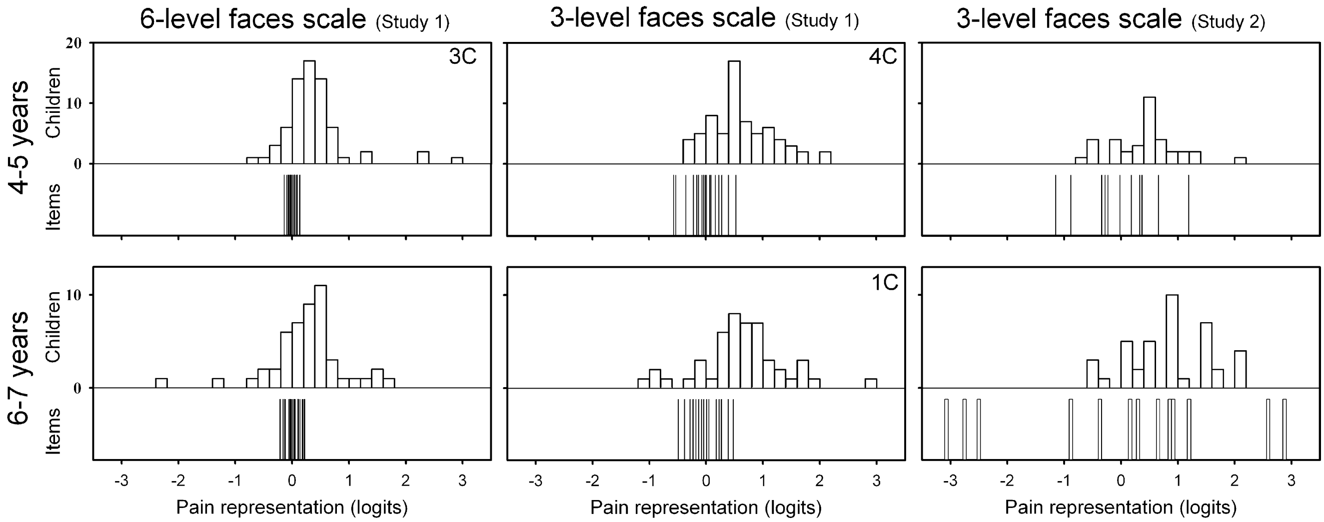

Fig. 3. Distributions of children's measures and item locations estimated from responses on the 6-level faces scale of Study 1 (left panels), on the 3-level faces scale of Study 1 (middle panels) and on the 3-level faces scale of Study 2 at test time (right panels) for the 4–5 year-old (upper panels) and for the 6–7 year-old children (bottom panels). Number of children presenting an extreme maximum score is marked on the upper right corner for each child distribution followed by a 'C' (children).

Table 2
Percentage utilization of categories

Age (years)	N	% Utilization of categories								
		6-Level faces scale						3-Level faces scale		
		0	1	2	3	4	5	0	1	2
4–5	72	11	8	12	14	17	38	19	26	55
6–7	49	12	9	15	18	18	28	18	32	50

N, number of subjects.

in the use of the 3-level faces scale by age group. The middle category is distinguished for the 6–7 year-olds (distance between τ_1 and $\tau_2 = 0.60$ logit) while it fails to emerge for 4–5 year-olds (distance between τ_1 and $\tau_2 = -0.18$ logit). These results suggest that children of 4–5 years of age only distinguish two response levels and that children of 6–7 years of age can distinguish three response levels.

The possible ability of 6 and 7 year-old children to distinguish four or five response levels on a face scale was also investigated. In order to construct four- and five-level faces scales, scores on the original 6-level faces scale were recoded by collapsing responses to adjacent categories. Recoded data were then analyzed. For the four-level response scale, the grouping procedure where face 1 was grouped with face 2 and face 3 with face 4 led to satisfying results (distance between τ_1 and $\tau_2 = 0.93$ logit and distance between τ_2 and $\tau_3 = 1.09$ logit). Alternative recoding procedures into four-level faces scales were also checked; however, no other solution provided ordered thresholds. For the five-level response scale, none of the possible groupings provided ordered thresholds.

3. Study 2

The first study revealed the low performances of the 6-level faces scale as compared to the 3-level faces scale and also the difficulty children experience in scoring the imaginary painfulness of items. Taking these facts into consideration, a second study was conducted with the principal aim of enhancing children's performance in discriminating between items by modifying the test in four ways. First, we sought to clarify the representations of the portrayed painful situations by using cartoons depicting the imaginary events. Second, we sought to make the task simpler by reducing the number of items from 25 to 15 with the intent of shortening the task, reducing the burden and keeping attention focused. Third, we included two control non-painful situations in the list of 15 items in order to check if children are really able to imagine and rate the painfulness of hypothetical situations or systematically choose high scores whatever the situation. Fourth, we replaced the 'smiling face' with a 'neutral face' in the 3-level faces scale. Indeed, the low percentage utilization (see Table 2) of the lowest scores may result from the use of a smiling face as the 'no pain' anchor. Indeed, several studies have shown that children have significantly higher pain ratings when using scales that commence with smiling 'no pain' faces in comparison to faces scales with neutral 'no pain' faces (Chambers and Craig, 1998; Chambers et al., 1999; Chambers et al., 2005). Children may be reluctant to point to a smiling 'no pain' face even if the situation is not painful probably because there is nothing happy about the circumstances even if pain is absent. This may have skewed the ratings towards faces at the upper end of the scale. With these four modifications, we commenced the second study.

3.1. Methods

3.1.1. Subjects

Seventy-six other children (41 boys and 35 girls) also aged from 4 to 7 years participated in this second study. The recruitment procedure was otherwise identical to that of Study 1. The sample of children was similarly divided into two age categories: 4–5 years, $n = 36$ (22 boys, 14 girls) and 6–7 years, $n = 40$ (19 boys, 21 girls).

3.1.2. Faces scale

A 3-level faces scale with a neutral face as 'no pain' anchor was presented to the children (Fig. 1c). This 3-level faces scale was similar in face design to those used in Study 1.

3.1.3. Procedure

All participants were tested individually in the same conditions as Study 1. Again, before testing, a few minutes were spent in order to meet the child, make sure that he had correct notions of pain and to explain the task. Then, children were presented a series of 15 cartoon pictures portraying painful situations. For a list of the pictures, see Table 1, bottom list. Ten situations were selected from the items used in Study 1 on the basis of the analysis of responses

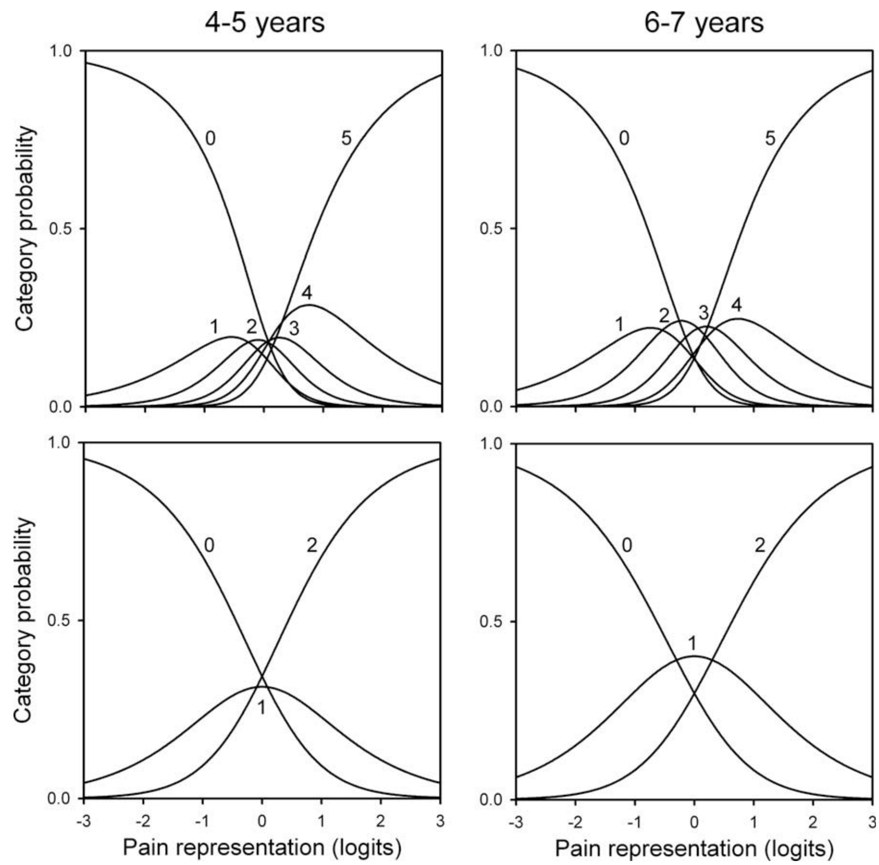

Fig. 4. Upper panels: category probability curves for the 6-level faces scale for the 4–5 year-old (left panel) and for the 6–7 year-old children (right panel). Bottom panels: category probability curves for the 3-level faces scale for the 4–5 year-old (left panel) and for the 6–7 year-old children (right panel).

and of children's understanding of the situations (items 1–10). Three painful situations were added to the list to extend the range of painfulness (items 11–13) and two non-painful situations were added as a control to check that children do not unthinkingly select high scores (items 14 and 15). Item 15 differs according to the sex of the child: to play soccer for boys and to skip a rope for girls. Cartoon depictions were presented successively and in random order. Cartoons presented to girls showed a child with long hair and a dress. Cartoons presented to boys showed a child with short hair and shorts. All other elements of the cartoon were perfectly similar. The facial expression of the depicted child was neutral. The child was asked to pretend that he was the child in the picture and to choose on the 3-level faces scale, the face that best reflects the amount of pain he would feel in each situation. The session lasted approximately 15 min. Test–retest stability of responses was investigated by resubmitting the test after an interim of ten days. Retest also served to further validate results found the first time. From the 76 children tested at the time of the first assessment, four could not be reassessed because of their absence from school at the time of the second visit.

3.1.4. Rasch analysis

The analysis of responses was performed exactly as for Study 1. Additionally, test–retest stability for item hierarchy and for children's pain representation was determined by an intraclass correlation coefficient (ICC).

3.2. Results

Separate Rasch analyses were conducted for each age group and for each assessment (test and retest). Items 14 and 15 elicited responses in the first response category from almost all children (74 out of 76 children at test time and 71 out of 72 children at retest time for item 14 and 62 out of 76 children at test time and 70 out of 72 children at retest time for item 15). These two control items were removed from the subsequent analyses for several reasons. Basically, these items had been added to check that children did not unthinkingly select high scores. In addition, item 14 presented an extreme minimal total score of 0 at test time for the 4–5 years and at retest time for the 6–7 years. Items with an extreme total score (zero or perfect total score) imply extreme and indefinitely located measures and are difficult to use in further investigations of the functioning of response scales (Wright, 1998). Finally, these two items presented a very poor fit to the model indicating that they did not satisfy the model requirement of unidimensionality.

Distributions of child measures and item locations at test time are presented in Fig. 3 (two right panels). Analyses led to item distributions that spread over 2.34 logits at test and over 2.59 logits at retest for the 4–5 years. For the group with 6–7 years of age, item distributions spread over 5.95 logits at test and over 7.31 logits at retest. Compared with Study 1, the item locations are spread over a range about twice as wide for the 4–5 years old and six times as wide for the 6–7 years old. In other words, children's ability to differentiate the different situations according to their painfulness is considerably improved in this second study. The comparison of the two age groups highlights the wider range of item locations for the older children compared with the younger or in other words the higher capacity of the older ones for distinguishing the different situations according to their painfulness. On the other hand, children still favor high pain scores as revealed by the percentage utilization of each response category (Table 3). As for Study 1, this

Table 3
Percentage utilization of categories

Age(years)	N	% Utilization of categories					
		Test			Retest		
		0	1	2	0	1	2
4–5	36	25	28	47	26	28	46
6–7	40	20	39	41	19	36	45

N, number of subjects.

percentage increases together with the response categories and is very similar at test and at retest time. The test–retest stability of situation calibrations is presented in Fig. 5 (upper panels). Situation locations are highly correlated between both assessments (ICC = 0.89, $p < 0.001$ for the 4–5 years and ICC = 0.96, $p < 0.001$ for the 6–7 years). On the other hand, children measures estimated by the model at the first and the second assessment are poorly but nevertheless significantly correlated (ICC = 0.34, $p = 0.0305$ for the 4–5 years and ICC = 0.44, $p = 0.0024$ for the 6–7 years) (Fig. 5, bottom panels).

The examination of category functioning is presented in Fig. 6 for test time. The middle category emerges for both age groups. Nevertheless, the category probability curves show an increase of the distance between both thresholds with age. The extremely narrow range for which category 1 is the most probable in the 4–5 year-olds signifies that this category is not really distinguished. These results confirm that children of 4–5 years old only distinguish two response levels. On the other hand, the 6–7 year olds clearly distinguish all three levels of the 3-level faces scale even better than in Study 1.

4. Discussion

The number of response categories on a pain intensity scale is increased with the intent of obtaining a more sensitive indication of pain severity. However, beyond a certain number of categories, confusion appears, as the different response categories are not distinguished anymore, compromising the validity of the ratings. The aim of the present studies was to determine the number of faces constituting the best trade off between precision and distinction capacity in children aged between four and seven years. In the first study, children were asked to rate 25 painful situations on a 3- and a 6-level faces scale with a smiling face as the 'no pain' anchor. This first study revealed the low performance of the 6-level faces scale as compared to the 3-level faces scale and also the difficulty children experienced in distinguishing the different situations according to their painfulness. Consequently, the second study was conducted with the objective of improving children's performance in discriminating between situations. To improve the representation of the situations, pictures accompanied them. To keep attention focused, the number of situations was reduced to 15. To

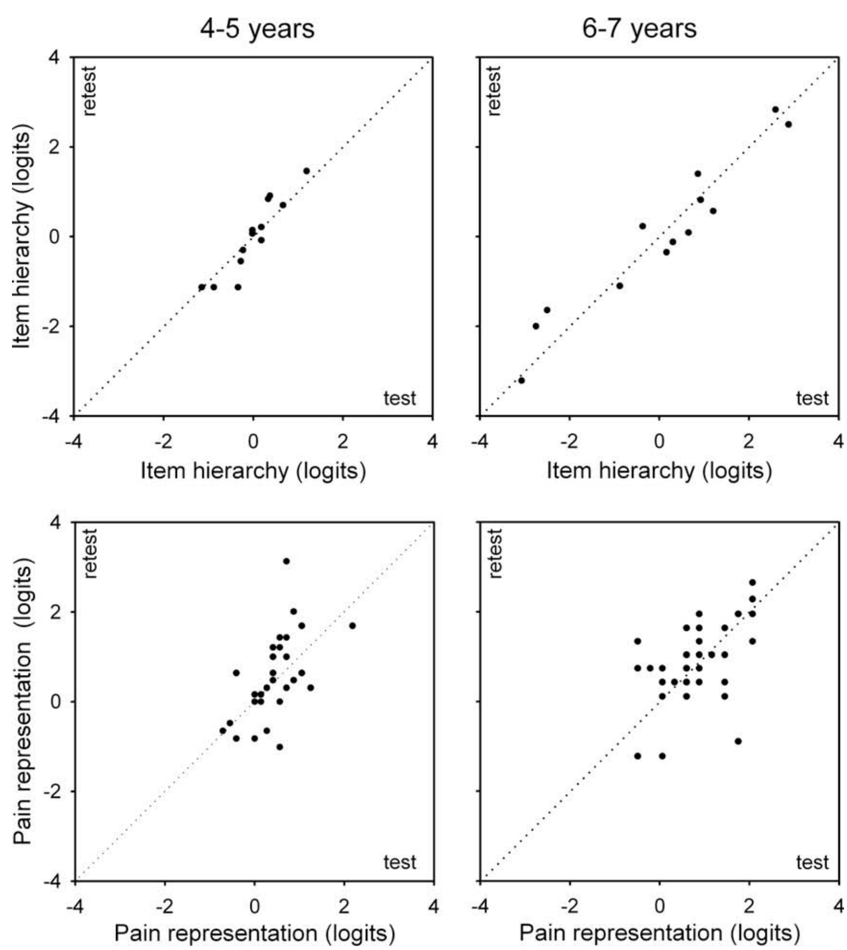

Fig. 5. Upper panels: relationship between item painfulness across time for the 4–5 year-old (left panel) and for the 6–7 year-old children (right panel). Most painful items are plotted in the bottom/left part of the panels. Bottom panels: relationship between the pain representation measures of children at the first and second assessment (delay: 10 days) for the 4–5 year-old (left panel) and for the 6–7 year-old children (right panel). Children with a high representation of pain intensity are plotted in the top/right part of the panels.

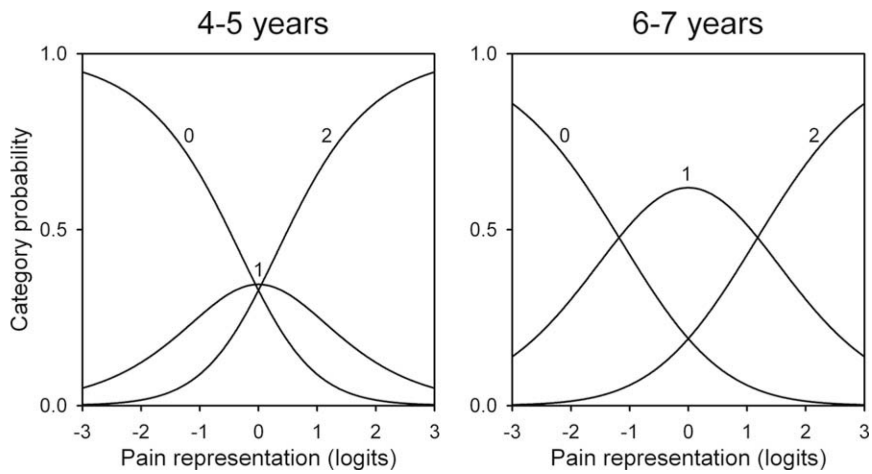

Fig. 6. Category probability curves for the 3-level faces scale at test time for both age groups.

enhance the use of the first category of the 3-level faces scale, a neutral face replaced a smiling face as the 'no pain' anchor. In addition, participants were tested twice to examine the stability of responses.

Study 2 led to a great improvement in children's capacity to differentiate the different situations according to their painfulness. This improvement was greater for the older age group indicating that the ability to differentiate painful situations not only depend on the way they are presented but mostly on the age of the child. The use of a faces scale with a neutral face as the 'no pain' anchor in Study 2 only slightly raised the response rate in the lower intensity categories. The percentage utilization still increased with the response categories. Other factors must be invoked to explain the lack of low ratings. Probably, the highest contributing factor is the small number of barely painful situations. Indeed, the two least painful situations (items 4 and 12) produced a pattern of score frequencies that decreases with the response categories indicating that the addition of such situations would have increased the percentage utilization of the lowest scores. Another factor explaining the lack of low scores is the previously reported age-related difference in children's pain report. Several studies have shown that young children (about the same age as those enrolled in the present research) report more pain than older children (Arts et al., 1994; Goodenough et al., 1997; Chambers and Craig, 1998). However, the reasons for these higher scores are not clear. The cause may be a rating bias due to cognitively immature judgment of pain intensity and/or of usage of rating scales. The bias may also be due to true age difference in children's perceptions of painful events due to the activation of more nociceptors (spatial and temporal summation) in smaller bodies as discussed in Arts and coworkers (1994). This rating bias of young children towards the upper end of the response scale could partly explain the reported parental underestimation of children's pain (Chambers et al., 1999; Chambers et al., 2005).

Ratings provided by children at the first and at the second assessment resulted in item locations that are highly correlated for each age group. This signifies that, when considering the responses of all children, items are consistently rated across time. The most painful items at test time are still estimated as the most painful at retest time and *vice versa*. On the other hand, the same ratings resulted in measures of child pain representations that are significantly but weakly correlated for the 4–5 year-olds as well as for the 6–7 year-olds. This signifies that each individual child rated the different situations moderately consistently from one time to another. Children probably have difficulty imagining such hypothetical events despite the help of pictures. More than adults, their pain representation is also likely to be influenced by external factors such as the emotional status and the level of concentration at the time of the assessment.

In Study 2, modifications in the protocol resulted in an improvement of the discrimination of the response categories of the 3-level faces scale when assessing the pain intensity of hypothetical painful events. Nevertheless, in the 4–5 year-old group, the distance between thresholds is extremely small (0.10 logits at test time and 0.48 logits at retest time). On the other hand, for the 6–7 year-old children, the distance between thresholds is much wider (2.37 logits at test time and 2.42 logits at retest time) indicating that the middle category is well distinguished. Linacre (2002) recommended a minimal distance of 1.4 logit in order for a 3-category scale to be equivalent to two dichotomies. On the whole, these results show that 4–5 year-old children distinguish only two response categories when rating imaginary painful situations on a faces scale and that 6–7 year-old children are able to use the 3-level faces scale for the same task. The difficulty of young children to use faces scales is consistent with findings of recent research (Stanford et al., 2006). This study examined young children's ability to use the 6-level Faces Pain Scale-Revised (Hicks et al., 2001) in response to cartoons depictions and investigated the role of developmental factors (numerical reasoning, language and overall cognitive development) in predicting their ability to use the scale. Authors calculated an error score for each child based on unexpected responses of pain severity depicted in the cartoons. They found that many young children aged 3–6 years experienced difficulties using the Faces Pain Scale-Revised and that age was the only significant predictor of children's ability to use the scale. They explained this latter result by the fact that age encompasses several cognitive domains of child development. The present study suggests that this reported difficulty to use a faces scale is probably due to difficulty in discriminating the six different levels of responses and that reduction in the number of faces could improve the validity of children's responses.

Finally, several important directions for future research can be recommended. First, an important next step is refining the developmental progress in using faces scales by using a larger sample of participants, which would allow finer division of the sample according to age. A second point to investigate is the possible ability of 6 and 7 year-old children to distinguish four faces. An exploratory analysis (grouping of some categories) performed on responses of the 6-level faces scale used in Study 1 suggested that the 6–7 year-old children could distinguish four response categories. However, these *a posteriori* combinations are somewhat artificial in the sense that the investigator cannot be sure that children would use the new 4-level faces scale as postulated by the combining of categories. Even if some studies have shown that the characteristics of the optimal categorization identified by the Rasch post-hoc analysis in a previous study (Zhu et al., 2001) could be maintained when the revised scale was applied to the same populations (Zhu, 2002), each new combination must be validated empirically. Third, assessing older children would inform about the age at which the ability to distinguish the different response levels reaches its maximum. Finally, these two studies involved a particular task: the assessment of imaginary painful situations. Of course, the use of 'imagined' pain, while a reasonable surrogate for 'real' pain, is still a method with limitations and should be extrapolated to clinical situations with caution. Imaginary painful situations were chosen as a first approach. Indeed, it prevents to expose a large sample of subjects, particularly children, to an experimental nociceptive stimulus without an empirical basis. Future studies should investigate the optimal number of response levels in situations of 'real' pain. According to the results of the present study, we may recommend two response options for 4–5 years old children and three or four for the 6–7 years old.

To conclude, appropriate management of pain in children depends on valid, reliable and age-appropriate pain measures. In the past two decades, there has been an increased interest and significant improvements in pediatric pain measurement with, among others, the development of faces scales. However, the central issue of the distinction of the response categories on these scales has not been thoroughly studied. The present studies were conducted to further examine this question. From their results, it appears that children from four to seven do not distinguish as many faces as proposed by the majority of available faces scales when rating hypothetical situations. If these results are confirmed in a clinical pain context, they will have important implications. For instance, since young children (before six) perceive pain intensity in a dichotomous way, the more valid self-assessment of pain is maybe simply to ask them if they are in pain. A possible substitute for young children's assessment of pain would be to use behavioral measures in conjunction with self-report measures (Crellin et al., 2007).

References

Arts SE, Abu-Saad HH, Champion GD, Crawford MR, Fisher RJ, Juniper KH, et al. Age-related response to lidocaine–prilocaine (EMLA) emulsion and effect of music distraction on the pain of intravenous cannulation. Pediatrics 1994;93:797–801.

Beyer JE. The Oucher: a user's manual and technical report. Evanston, IL: The Hospital Play Equipment Company; 1984.

Bieri D, Reeve RA, Champion GD, Addicoat L, Ziegler JB. 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;41:139–50.

Chambers CT, Craig KD. An intrusive impact of anchors in children's faces pain scales. Pain 1998;78:27–37.

Chambers CT, Giesbrecht K, Craig KD, Bennett SM, Huntsman E. A comparison of faces scales for the measurement of pediatric pain: children's and parents' ratings. Pain 1999;83:25–35.

Chambers CT, Hardial J, Craig KD, Court C, Montgomery C. Faces scales for the measurement of postoperative pain intensity in children following minor surgery. Clin J Pain 2005;21:277–85.

Crellin D, Sullivan TP, Babl FE, O'Sullivan R, Hutchinson A. Analysis of the validation of existing behavioral pain and distress scales for use in the procedural setting. Paediatr Anaesth 2007;17:720–33.

Gaffney A, Dunne EA. Developmental aspects of children's definitions of pain. Pain 1986;26:105–17.

Gelman R, Baillargeon R. A review of some Piagetian concepts. In: Flavell JH, Markman EM, editors. Handbook of child psychology. Cognitive development 1983;vol. 3. New York: Wiley Press; 1983. p. 167–230.

Goodenough B, Kampel L, Champion GD, Laubreaux L, Nicholas MK, Ziegler JB, et al. An investigation of the placebo effect and age-related factors in the report of needle pain from venipuncture in children. Pain 1997;72:383–91.

Hicks CL, von Baeyer CL, Spafford PA, van Korlaar I, Goodenough B. The faces pain scale-revised: toward a common metric in pediatric pain measurement. Pain 2001;93:173–83.

Hunter M, McDowell L, Hennessy R, Cassey J. An evaluation of the faces pain scale with young children. J Pain Symptom Manage 2000;20:122–9.

Kuttner L, LePage T. Faces scales for the assessment of pediatric pain: a critical review. Can J Behav Sci 1989;21:198–203.

LeBaron S, Zeltzer L. Assessment of acute pain and anxiety in children and adolescents by self-reports, observer reports, and a behaviour checklist. J Consult Clin Psychol 1984;52:729–38.

Linacre JM. Sample size and item calibration stability. Rasch Meas Trans 1994;7:328.

Linacre JM. Optimizing rating scale category effectiveness. J Appl Meas 2002;3:85–106.

Linacre JM, Wright BD. A user's guide to Winsteps: Rasch-model computer program. Chicago: Mesa Press; 1998.

Masters GN. A Rasch model for partial credit scoring. Psychometrika 1982;47:149–74.

McGrath PA, Seifert CE, Speechley KN, Booth JC, Stitt L, Gibson MC. A new analogue scale for assessing children's pain: an initial validation study. Pain 1996;64:435–43.

Pesudovs KP, Noble BA. Improving subjective scaling of pain using Rasch analysis. J Pain 2005;6:630–6.

Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research; 1960 [Expanded edition, Chicago: Mesa Press; 1980].

Stanford EA, Chambers CT, Craig KD. The role of developmental factors in predicting young children's use of a self-report scale for pain. Pain 2006;120:16–23.

Stinson JN, Kavanagh T, Yamada J, Gill N, Stevens B. 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 2006;125:143–57.

Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. J Rehabil Med 2003;35:105–15.

Wong DL, Baker CM. Pain in children: a comparison of assessment scales. Pediatr Nurs 1988;14:9–17.

Wright BD. Estimating measures for extreme scores. Rasch Meas Trans 1998;12:632–3.

Wright BD, Masters GN. Rating scale analysis. Chicago: Mesa Press; 1982.

Yang JC, Wagner JM, Clark WC. Psychological distress and mood in chronic pain and surgical patients: a sensory decision theory analysis. In: Bonica J, Lindblom U, Igg A, editors. Advances in pain research and therapy. New York: Raven Press; 1983. p. 901–6.

Yeh C-H. Development and validation of the Asian version of the Oucher: a pain intensity scale for children. J Pain 2005;6:526–34.

Zempsky WT, Robbins B, McKay K. Reduction of topical anesthetic onset time using ultrasound: a randomized controlled trial prior to venipuncture in young children. Pain Med 2008(Mar):11 [Epub ahead of print].

Zhu W. A confirmatory study of Rasch-based optimal categorization of a rating scale. J Appl Meas 2002;3:1–15.

Zhu W, Timm G, Ainsworth B. Rasch calibration and optimal categorization of an instrument measuring women's exercise perseverance and barriers. Res Q Exerc Sport 2001;72:104–16.