A functional analysis of the Wong-Baker Faces Pain Rating Scale: linearity, discriminability and amplitude

Uma análise funcional da Wong-Baker Faces Pain Rating Scale: linearidade, discriminabilidade e amplitude

Resumen

Objetivos: Investigar el grado en que la Wong-Baker FACES Pain Rating Scale (FACES) presenta propiedades de intervalo en muestras de niños con diferentes edades (6-8; 9-11 años) y experiencias de dolor.

Metodología: El estudio se basa en la metodología de la Medida Funcional, que ofrece un criterio empírico de validación de la linealidad de la escala y la posibilidad de una medida de intervalo de los estímulos.

Resultados: La FACES presenta significativos desvíos de intervalo (igual distancia) en niños de 6-8 años, lo que ocasiona una menor amplitud de la escala. Aumenta de manera tendencialmente más linear en las muestras de niños más jóvenes (9-11 años), especialmente en el grupo con dolor crónico.

Conclusión: En niños más jóvenes, las puntuaciones de la FACES no poseen el nivel de intervalo. En niños con más de 8 años y con antecedentes de dolor crónico parece razonable atribuirles esa propiedad.

Palabras clave: dolor; escala; niño.
Introduction

Faces scales became a common pain measurement tool since the 1980s, particularly in paediatric populations. Some of these scales benefit from extensive empirical validation studies and have been recommended in systematic reviews as valid self-report measures of pain intensity (Stinson, Kavanagh, Yamada, Gill, & Stevens, 2006; Tomlinson, von Baeyer, Stinson, & Sung, 2010). The Wong Baker FACES Pain Rating Scale (FACES) (Wong & Baker, 1988), composed of six drawn faces scoring from 0 to 10, is one such scale, taken as the object of this study.

Several features contribute to the appropriateness of a self-report measure of paediatric pain, among which being easy to use, inexpensive and well-liked by children, parents and health care professionals (von Baeyer, 2006). While desirable, however, these properties do not qualify a scale as valid and reliable, which depends on another set of properties, known as psychometric. The validation of paediatric pain scales typically rests on assessing its content or construct validity (Ruskin, Amaria, Warnock, & McGrath, 2011) and the reliability of faces scales on the use of the test-retest method (Stinson et al., 2006). Scale’s responsivity or sensitivity to change, a critical property from the clinical standpoint, is also commonly targeted for assessment (Stinson et al., 2006; Tomlinson et al., 2010). This study focuses on a less often considered property: the level of measurement (ordinal, interval or ratio) afforded by the scale (Stevens, 1946; von Baeyer, 2009).

The bearing of levels of measurement on pain evaluation can be simply illustrated. Consider the goal of comparing absolute pain intensity between individuals: this would need measures at the ratio level, with an absolute zero and a common unit, hardly available if at all (von Baeyer, 2009). Conversely, comparing the pain scores of a single individual over a period of time requires no more than an ordinal level of measurement, allowing assessing whether pain has increased or decreased (von Baeyer, 2009). Lastly, establishing two reductions in pain scores (e.g., from 6 to 4 and from 4 to 2) as equal in magnitude entails measuring pain at the interval level, on a scale with a constant unit throughout its length.

The last example makes clear the benefits of interval measures to pain management. To be able to measure how much a given procedure reduces or increases pain is, from a practical standpoint, a marked advantage. Measurement at the interval level is also key for research on the interactions between pain determinants. As interactions correspond to differences in the effects of one factor (e.g., analgesic intervention in children) as a function of the levels of another factor (e.g., parents’ presence or absence), they actually require a comparison between differences (e.g., how much pain decreased with and without the parents’ presence).

Interval scales have been defined since Stevens (1946) as those allowing affine transformations of their scores (i.e., transformations of the type \( x' = ax + b \)). However, this formal definition rests silent on how to assess whether such transformations are legitimate in each case. As pointed out in Anderson (1981), the actual substantive requirement of an interval scale is that equal intervals on the psychological dimension to be measured (e.g., pain intensity) correspond to equal intervals on the observable response scale – i.e., that the unobservable response \( r \) be mapped on the external response \( R \) by a linear operation, such that \( R = b + ar \) (with \( a \) and \( b \) constants). An equal-interval scale is thus more precisely a linear response scale, and checking the equal-intervals property amounts to testing the linearity of the response scale (Anderson, 1982, 2001).

Like most measurement theories, the formal theory scale-type introduced by Stevens (1946) assumes the problem solved. In contrast, the Functional Measurement theory (Anderson, 1981, 1982) rests critically on an empirical criterion for validating the linearity of response scales. This study aims at testing the linearity (equal intervals property) of the FACES in paediatric samples by means of functional measurement. As linearity consists of a functional relationship between a psychological dimension and an observable response scale, the level of measurement obtained clearly depends on both the scale and the respondent. This consideration is especially important in the realm of paediatric pain, where developmental aspects may intervene in a decisive manner (von Baeyer, 2009). In general, the same response scale may well be linear in a given age group and not another, or in a pain condition but not another (e.g., chronic pain and acute pain).
Background

Information Integration Theory (IIT) and the methodology of Functional Measurement (FM) constitute the framework of this study. IIT is an experimental-based theory that investigates how multiple pieces of information are integrated into an unitary judgment (e.g., how good something is, how risky or how painful). It embraces the notion that every psychological process is multi-determined, and rests on the use of integration tasks involving the joint manipulation of at least two information dimensions (factors), whose combinations participants evaluate on a continuous (i.e., varying in degree) response scale.

The key finding of IIT, replicated across multiple domains (Anderson, 1991; Athayde & Oliveira, 2006), is that people often rely on a limited number of algebraic models to integrate distinct pieces of information: the additive, multiplicative and averaging rules (Anderson, 1981). These models constitute the cognitive algebra, which is at the source of all potential benefits of using IIT, including the capability for functional measurement. If no integration model can be established in a given field, IIT is of no use there. If one (or more) algebraic models can alternatively be found, then IIT affords the means (1) to test the linearity (equal-intervals) of the response scale and (2) of measuring the stimulus variables on an interval scale with a common unit (Anderson, 1982).

The IIT rationale for validating the linearity of a response scale can be illustrated in the simplest way with the additive model. If an integration task with two factors gives rise to a pattern of parallel lines in a factorial plot, two conditions must have been met: (1) the integration must have obeyed an additive-type rule; (2) the internal result of the integration (r) must have been mapped linearly (without distortion) onto the external response (R). Violation of any of these conditions would have prevented parallelism to occur, which makes that the finding of parallelism supports both conditions simultaneously (Anderson, 1981). Although more complex, the other integration rules provide similar constraints, which allow testing response linearity in simultaneous with the establishment of the integration model (Anderson, 1982).

FM is based on the cognitive algebra, consisting in deriving the metric information contained in the integration models. In order to express their judgments on a continuous (varying by degree) response scale, participants must implicitly assign a value to each piece of information to be integrated. This value is dubbed functional, as it does not pre-exist in the stimulus and depends solely on how it functions in the integration. Algebraic models thus implicitly include a quantification of the stimulus variables, which FM makes explicit. For additive and multiplicative models, the marginal means of the factorial design provide proper functional values at the interval level (Anderson, 1982). In the case of averaging, recurrent estimation with the support of computer software (Vidotto & Vicentini, 2007) is needed to arrive at deriving functional values.

Three implications of the use of FM in the present study deserve mention. (1) As any metrics will be derived from the cognitive operation of integration performed by children, these metrics will be, by definition, age-suited. (2) As the locus of the integration is each individual subject, FM offers the possibility for measurement at the individual level. This feature sets it apart from other approaches to interval level measurement, such as the Thurstone’s model (Kuttner & LePage, 1989), which can only be applied at the group level (Anderson, 1981). (3) Under given conditions, FM allows to measure two types of functional parameters, with distinct psychological meanings: the stimulus magnitude (scale value) and the importance (weight) of its contribution for the judgment (Anderson, 1981, 1982).

Research questions

This study is aimed at investigating the metric properties of the FACES with functional measurement. A first goal is to verify the existence of algebraic integration models in the judgement of pain in children, which is a condition for applying FM. Should they exist, those models will imply per se the existence of a metric (interval) understanding of pain in children.

A second goal is assessing the prevalence of these models in children of different ages, particularly among younger children, aged less than eight years and with different pain experiences (no pain, chronic pain and postoperative pain).
A third goal is to provide a check on the assumption of equal intervals inherent in the scoring of FACES, while keeping to a comparative perspective regarding age and pain experience. To that end, in addition to investigating the perceptual distances among expressions, the scale’s dynamic range of variation and the profile of deviations to linearity will be compared between age groups and pain conditions.

**Methodology**

**Participants**

The study involved six samples of children distributed by two age groups (6-8 and 9-11 years old) and three types of pain experience: 1) no regular experience of pain (no pain condition); 2) postoperative pain (acute pain condition); 3) persistent pain for more than three months (chronic pain condition). Table 1 characterises the different groups regarding their experience of pain, age, gender, and sample size (n). No pain participants attended pre-schools, colleges and basic/secondary schools in the Centre region. Children with postoperative pain (acute pain condition) were hospitalised in Surgery, Orthopaedic and Neurosurgery services. Children with the chronic pain condition were followed in outpatient visits of the Oncology services of two Hospitals with paediatric services. No task was performed by children in pain. A convenience sample of participants was used.

Table 1

<table>
<thead>
<tr>
<th>Exp. Pain</th>
<th>Groups</th>
<th>Ages</th>
<th>n</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>No pain</td>
<td>G1</td>
<td>6-8 (M: 6.6; SD: 0.5)</td>
<td>27</td>
<td>18 F, 9 M</td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>9-11 (M: 10.8; SD: 0.3)</td>
<td>28</td>
<td>18 F, 10 M</td>
</tr>
<tr>
<td>Acute pain</td>
<td>G3</td>
<td>6-8 (M: 7.3; SD: 0.9 )</td>
<td>11</td>
<td>5 F, 6 M</td>
</tr>
<tr>
<td></td>
<td>G4</td>
<td>9-11 (M: 10.2; SD: 0.8)</td>
<td>26</td>
<td>13 F, 13 M</td>
</tr>
<tr>
<td>Chronic pain</td>
<td>G5</td>
<td>6-8 (M: 7; SD: 0.8 )</td>
<td>15</td>
<td>7 F, 8 M</td>
</tr>
<tr>
<td></td>
<td>G6</td>
<td>9-11 (M: 10.4; SD: 0.7)</td>
<td>20</td>
<td>8 F, 12 M</td>
</tr>
</tbody>
</table>

**Stimuli**

Pairs of schematic faces, resulting from pair combinations between the six FACES expressions and the six Facial Pain Scale-Revised (FPS-R) expressions (Hicks, von Baeyer, Spafford, van Korlaar, & Goodenough, 2001) (See Figure 1).

![Figure 1. Expressions of the FACES and FPS-R scales (original obtained from the Wong-Baker FACES Foundation and the International Association for the Study of Pain, respectively). The numbers indicate the increasing intensity of the expressions.](image-url)
The construction of the stimuli involved the following steps: the cutting of faces in a graphic editor; resizing of images for them to occupy a similar area; assembling of all possible pair combinations between the FACES and the FPS-R expressions (see Figure 2).

Experimental Design and Data Analysis
Each scale was treated as one experimental factor and the levels (expressions) of both scales fully factorially combined, resulting in 36 conditions implemented by an equal number of pairs of faces. All participants assessed all pairs. The overall design was thus a full factorial $6 \times 6$ (FACES) $\times$ (FPS-R) repeated measures design. The two one-factor subdesigns, corresponding to the isolated presentation of each FACES and FPS-R expression, were added to the main design. Subdesigns are required for testing the additive versus the averaging integration rule and for independently estimating the scale and importance values in the event of averaging (Anderson, 1982). Data analysis was based on mixed and repeated measures ANOVAs. In cases of violation of sphericity, the Greenhouse-Geisser correction to the degrees of freedom was used.

Procedure
The stimuli were randomly presented, one at a time, on a computer screen. The task was carried out individually, in a reserved space of the school or hospital service. A familiarisation period, consisting of a variable number of trials, preceded the experience. The instructions included a dialogue about the concept of pain and its variation in degree, supported by illustrations. Children were asked to assess the overall amount of pain in each pair of faces.

In the group of 9- to 11-year-olds, answers were given by positioning the mouse on a horizontal graphic scale left-anchored on no pain and right-anchored on very much pain, followed by a click. In the group of 6- to 8-year-olds, children ought to press a button on a response box for a given time, measured in ms. During the instructional phase, an animated magic dog was presented, which was controlled by the answer box. Every time the dog licked, the pain was transferred into a glass shown on the screen, which became full after 13 seconds of pressing. Children pressed the button during the time deemed needed to transfer all pain expressed in a pair of faces into the glass. For the first three seconds, the glass was shown empty, allowing the children to stop pressing the button before any pain was transferred.

Results
Cognitive Algebra
The plots in Figure 3 illustrate the factorial patterns obtained in the no pain and chronic pain conditions in both age groups. The plots for the acute pain condition, omitted for reasons of space, displayed similar patterns. The right convergence of lines is consistent with a differential weighting averaging model in which the importance of the expressions of pain increases with their intensity. The dashed line corresponds to isolated presentations of the levels of the factor in the abscissa (FACES). The fact that it exhibits a steeper slope intersecting the other lines, rules out adding and supports the averaging rule (Anderson, 1981). The visual inspection of the plots was supported by repeated measures ANOVAs, which found main effects of both factors and significant interaction terms (smaller $F$ for main effects $= 38.4$, $p < .001$; smaller $F$ for the interactions $= 2.28$, $p = .018$) in all groups.
A functional analysis of the Wong-Baker Faces Pain Rating Scale: linearity, discriminability and amplitude

**Figure 3.** Factorial plots for the 6 (FACES) × 6 (FPS-R) design in the no pain and chronic pain conditions of each age group. The FACES is in the abscissa and the FPS-R is the curve parameter. The ordinates correspond to the means of the subjects’ judgments (tenths of a second in younger children; 0-40 format in older children). The dashed line stands for the isolated presentations of FACES expressions.

Subgroups of children who complied with an equal averaging rule, signaled by a pattern of parallel lines crossed by the dashed line, were identified in all cases. These subgroups were predominant among older children in the acute pain (14 participants out of 26) and chronic pain (13 out of 20) conditions. Along with the good fit to the data provided by the averaging model (tested by ANOVAs on the model residuals: see Anderson, 1982), the emergence of parallelism in these subgroups contributed to validate both response scales as linear.

**Functional Measurement**

Based on the established averaging model, the scale values and the importance of the expressions in each faces scale were independently estimated, using the R-AVERAGE software (Vidotto & Vicentini, 2007). The scale values reflect the perceived spacing between pain expressions in both the FACES and the FPS-R, regardless of other factors potentially involved in the judgement (e.g., attentional priority or affective resonance of certain expressions), which contribute in turn to the importance (weight) assigned to each face.

**Linearity**

The plots in Figure 4 allow comparing the profile of the FACES scale values between age groups and pain conditions. If the FACES was linear (implying equal intervals between expressions), the curves should correspond to straight line segments. Therefore, the greater the deviation from a straight line the higher the departure from the ideal of equal intervals.
Panel A shows that, overall, younger children present larger deviations from linearity. This indication was confirmed by a significant interaction FACES × age group in a mixed ANOVA conducted on the estimates of the scale values, $F(5,58) = 4.2; p < .001$. Panel B shows that, overall, the pain condition has no appreciable effects: the associated statistical comparisons concluded on the absence of significant effects resulting from the pain condition: $F(2.12) = 2.6; p = .079$ for the main effect; $F < 1$ for the interactions.

Panels C, D, and, E detail the interaction FACES × age group in each pain condition. The younger children's curve is closer both to linearity and to the older children's curve in the pain versus no pain conditions. Although it does not give rise to a significant interaction FACES × age group × pain condition, a slight trend for higher linearity in the chronic pain condition is detectable for the younger children. The same happens with the older children, who provide the best fit to linearity in that same condition ($r = 0.998$). Panels F and G detail the interaction FACES × pain condition in each age group. In the group of 6- to 8-year-olds, linearity can again be seen to improve from the no pain condition to the acute pain and to the chronic pain condition. As indicated before, this trend was not statistically significant. In the group of 9-11-year-olds, the curves virtually overlap (with a slight advantage for the group with chronic pain). Deviations from linearity, especially in younger children, take the form of a negative curvature on the upper side of the scale, with the effect of reducing its dynamic range of variation.

**Amplitude and Discriminability**

Figure 5 reproduces the perceptual separations between different FACES expressions in the six groups considered. Functional values are interval-level estimates an thus give meaning to the comparison of distances between expressions. The general structure of the distribution of pain expressions was replicated in all groups: perceptual contiguity of expressions 1 and 2, isolated salience of expression 3 and, to varying degrees, proximity between expressions 4, 5 and 6.

The scale's dynamic range of variation (difference between the maximum and minimum values) is noticeably lower in children aged 6-8 years old. An ANOVA over the distributions of range values, with age group and pain condition as between-subjects factors, confirmed the existence of a significant effect of the age group, $F (1.116) =$
A functional analysis of the Wong-Baker Faces Pain Rating Scale: linearity, discriminability and amplitude

Revista de Enfermagem Referência - IV - n.° 3 - 2014

The mean ordering of expressions complied with the normative ordering in the FACES in all groups. The question of whether consecutive pain expressions are discriminable does not fully overlap with the issue of their perceptual separation, referring specifically to the consistency with which they can be discriminated. In this study, we took as a criterion of discriminability between two faces the existence of a significant difference (p set to ≤ .025) between their scale values. The mean functional values of the different faces were compared in each group by means of a repeated measures ANOVA, followed by pairwise multiple-comparison tests with Bonferroni correction. The brackets in Figure 5 indicate significant differences between expressions, thus illustrating the discriminability profile in the different groups.

Figure 5. Perceptual separation between FACES expressions in the different groups. The functional estimates in the ordinate constitute an interval metric, expressing from 0 to 1 after normalisation to the amplitude of the response scales. The brackets stand for significant differences with an associated p-value ≤ 0.025 (after Bonferroni correction).

The largest number of discriminable intervals in the group of 6-8-year-olds was 3 (involving 4 levels of expression, one of which aggregating faces 4 and 5). In the group of 9-11-year-olds, the number of intervals was 3 in the no pain condition and 4 in the remaining conditions, which thus involved the discrimination of 5 levels of expression. As noted in the previous section, the best approximation to the ideal of equal intervals was found among older children, in the chronic pain condition.

Discussion

The results illustrate the practicability of the IIT methodology in the field of pain assessment in young children (e.g., 6 years old). Whether this methodology may be extended to children under the age of 6 years to assess pain remains an open empirical issue. The fact that children in both age groups provide algebraic integration patterns means that even the youngest children were capable of an interval, not just ordinal, understanding of pain.

In both age groups and in all pain conditions, the model of integration found was the averaging model, with higher prevalence of equal-weighting averaging (equal importance assigned to all expressions within a scale) in the group of 9-11-year-olds. The general structure of the perceptual separation between expressions, obtained via functional measurement, was replicated in all groups: indistinctness of the two first faces, perceptual salience of face 3 on the middle, and relative compression of the intervals between the last three levels of the scale. The replication of this
structure across groups speaks for its consistency and
generality and against a possible objection based on
the relatively small number of participants in each
group (for the role of replication in supporting the
generality of results, see Anderson, 2001).
Participants in the chronic pain condition were closer
to the ideal of equal intervals in both age groups.
This may be due to a more consolidated experience
of the quantitative variation of pain, or to a more
frequent contact with pain assessment instruments.
This result confirms linearity as a property dependent
on both the scale and the respondent and, therefore,
susceptible of improvement through learning.
The pattern of discriminability between expressions
showed that younger children could only
distinguish between four levels of expression of
pain (two of them being aggregated: faces 1-2 and
4-5). As the first level stood for no pain there were
actually 3 discriminable pain levels in the FACES. This
indication is consistent with the literature, which
points out to the distinction of two or three levels of
pain in children aged between three and seven years
(Belter, McIntosh, Finch Jr., & Sylor, 1988; Decruynaere,
Thonnard, & Plaghki, 2009). In older children in the
acute and chronic pain conditions, this number rose
to 4 (5, if no pain is included).
Measuring the FACES expressions at the interval level
opens way to the use of standard criteria for clinically
significant differences. Changes ranging from 10 to 20%
in the pain scores of the Visual Analogue Scale (VAS)
have been accepted as clinically significant (Powell,
Kelly, & Williams, 2001). On a scale like the FACES,
assuming equal intervals, this would correspond to
a change of one or two faces. However, uncertainty
as to the property of equal intervals did not allow
ensuring that such change would keep a constant
meaning all through the scale (Bulloch & Tenenbein,
2002). In contrast, once the FACES expressions have
been received an interval metric (see Figure 5), the standard
percentage of 10-20% can be legitimately applied to
identify those differences between expressions that
meet the clinical significance criterion.

Conclusion

The FACES presented sizable deviations from linearity.
These deviations typically reduced the dynamic range
of the scale by compressing together its last three
levels. The values of the first two levels (expressions
with smiles) were not discriminable in any of the
groups. Given they are redundant, eliminating one
of them seems like a reasonable suggestion (which
might help in simplifying the scale). In general,
scoing the FACES from 0 to 10 at equal steps of 2
doesn’t seem to justify. Deviations from linearity were
more pronounced for the younger children, and less
marked for children in the chronic pain condition.
The FACES scores should not be considered interval
measures in children under eight years old, but they
may be taken as a close approximation to that in older
children with a history of chronic pain.

Project PTDC/PSI-PCO/107910/2008, funded by the
Portuguese Foundation for Science and Technology
within the framework of the COMPETE/QREN
programme.

References

de sinais na avaliação da severidade de flebite associada à
Belter, R. W., McIntosh, J. A., Finch Jr., A. J., & Sylor, C. F.
(1988). Preschoolers’ ability to differentiate levels of pain: Relative
significant changes in acute pain in children. Academic Emergency Medicine, 9(5), 199–202. doi: 10.1197/aemj.9.5.199
response levels do children distinguish on faces scales for
Hicks, C. L., Von Baeyer, C. L., Spafford, P. A., Van Korlaar, I., &
Pain, 93(2), 173–183. doi:10.1016/S0304-3959(01)00314-1
A functional analysis of the Wong-Baker Faces Pain Rating Scale: linearity, discriminability and amplitude


