The CALI-9: A brief measure for assessing activity limitations in children and adolescents with chronic pain

Amy Lewandowski Holley, PhD[a], **Chuan Zhou, PhD**[b,f], **Anna C. Wilson, PhD**[a], **Keri Hainsworth, PhD**[c,d], and **Tonya M. Palermo, PhD**[e,f]

[a]Department of Pediatrics, Oregon Health & Science University, Portland OR, USA

[b]Department of Pediatrics, University of Washington, Seattle WA, USA

[c]Children's Hospital of Wisconsin, Milwaukee, WI, USA

[d]Department of Anesthesiology, Medical College of Wisconsin, Milwaukee WI, USA

[e]Department of Anesthesiology and Pain Medicine, University of Washington, Seattle WA, USA

[f]Seattle Children's Research Institute, Seattle WA, USA

Abstract

Pain-related functional limitations represent an important outcome domain to assess in children and adolescents with chronic pain. The aim of this study was to extend the empirical support of the CALI-21 (21-item Child Activity Limitations Interview), a well-validated measure of activity limitations, using a large, multisite sample and to develop a brief form of the measure with more interpretable scoring. A sample of 1616 youth and 1614 parents completed the CALI-21 at an initial appointment in 1 of 3 pain specialty clinics in the Midwest or Northwest United States, or as part of a research study following this initial visit. All youth also reported on usual pain intensity. CALI-21 data from 1236 youth and parents were used in analyses. Results of the exploratory and confirmatory factor analyses supported a common 2-factor structure (Active and Routine factors) for both child and parent report versions. Using item reduction, the 9 item measure (CALI-9) was developed with both child and parent versions showing good internal consistency and high cross-informant reliability. Initial validity was shown by the ability of the CALI-9 to distinguish by level of pain intensity. Findings suggest that the CALI-9 is a promising brief tool for the evaluation of pain-related activity limitations in youth with chronic pain and for proxy-report by parents. Advantages of the shortened scale include the revised 0-100 point scale, which increases interpretability, and further validation of the subscale scoring to assess specific limitations in active and routine physical functioning domains.

1. Introduction

Assessment of pain-related function is important in the evaluation and treatment of children and adolescents with chronic pain [13]. Pain interferes with daily functioning for a sizeable

v) Corresponding Author: Amy Lewandowski Holley PhD, 3181 SW Sam Jackson – CDRC, Portland OR USA, 97239, phone: 503-494-2833, fax: 503-494-5945, holleya@ohsu.edu.

group of youth, with estimates suggesting 5-8% experience significant pain-related disability [7]. One aspect of pain-related function is the perception of difficulties in participating in daily physical, recreational, routine living, and social activities [18]. Self-report of physical and role functioning have been recommended for routine assessment in trials conducted of chronic pain interventions in youth [11]. Moreover, accurate assessment of activity limitations can be used to both inform treatment goals and document treatment improvements.

The CALI-21[1] is a well-validated measure of activity limitations for youth with chronic pain that was developed and validated using diverse, clinical samples of children and adolescents with varying chronic pain conditions. The CALI-21 is unique from other measures of pain-related function in youth in that it includes individually scorable subscales that describe limitations in active (e.g., running, gym class) versus routine (reading, eating regular meals) domains, and a prospective 8-item version to use for daily diary data collection. This is particularly important as youth with diverse chronic pain complaints can experience disability in different domains. The measure has been used widely in a variety of cross-sectional and longitudinal studies as well as to assess treatment outcomes (e.g., [9; 14; 17; 26; 28]). Previous psychometric validation has been conducted with the CALI-21, including an exploratory factor analysis [16].

In addition to the CALI-21, a number of other assessment tools have been developed and validated over recent years including the Functional Disability Inventory (FDI) [3] and the PROMIS Pain Interference Scale [24]. While valid and reliable tools, these measures have limitations, particularly length and lack of relevancy of all items (e.g., running the length of a football field) (FDI) and the inability to isolate the type of activity limitations experienced (FDI and PROMIS Pain Interference Scale). Moreover, the PROMIS Pain Interference Scale includes cognitive and affective items (e.g. pain-related anger, attention problems, impact of pain on fun), which broadens the measure's focus beyond activity limitations.

The central aim of the current study is to develop a revised version of the CALI-21, replacing the prospective and retrospective versions with a single brief instrument of activity limitations. With the new measure, we also sought to eliminate frequently skipped/rarely endorsed items and change scoring to a readily interpretable 0-100 scale, streamlining the measure for rapid assessment of activity limitations in clinical and research settings. To accomplish these aims we conducted three main sets of statistical analyses. We first examined the underlying subscale structure of the CALI-21 using exploratory and confirmatory factor analyses with data from a large clinical sample of children and adolescents with chronic pain and their parents. In the second set of analyses we reduced the number of items on the CALI-21 through variable selection techniques in regression analysis. Finally we conducted preliminary validation analysis on the shortened version of the measure. We hypothesized that the CALI-21 could be improved, with the revised version

---

[1]The 21-item assessment tool has two questionnaire versions (CALI-21, CALQ; same items, different scoring systems) and an interview-report version (CALI). For consistency, the term "CALI-21" will be used throughout the manuscript to describe all prior versions of the measure.

having minimal loss in reliability, validity, and variability. Moreover a two-factor structure would be supported, as found in the prior exploratory factor analysis of the measure.

2. Method

Participants in this study were youth ages 8-18 years who presented to a specialty pediatric pain clinic in either the Midwest (Site 1) or Northwest (Sites 2 and 3) United States for a new patient evaluation for chronic pain. Youth and their parents completed questionnaires as part of standard clinic intake forms (n=1384) or participation in a research study (4 studies, n=232). Participants who completed questionnaires as part of a research study provided consent or assent prior to participating, with all study procedures approved by the Institutional Review Board. Institutional Review Board approval was obtained at each institution permitting analysis of previously collected clinical intake data for research purposes.

Inclusion criteria were: 1) that the child was between 8-18 years and was seeking care in a specialty care clinic for evaluation of chronic pain, and 2) both children and their parents were able to read and write in English to complete questionnaires.

2.1 Questionnaire Measures

2.1.1. Demographics—Demographic characteristics (child age, sex, and race) were provided by parents or obtained from each participant's electronic medical record.

2.1.2. Activity Limitations—Activity limitations were reported by children and their parents using either the CALI-21 or CALQ. Both of these measures use the same 21-items to assess pain-related activity limitations in children and adolescents (ages 8–18). The measures have parent and child report versions, with instructions asking participants to report on pain-related limitations in 21 activities over the previous 4 weeks. On the CALI-21, participants rate the difficulty in completing each activity on a 5-point rating scale, ranging from 0 'not very difficult' to 4 'extremely difficult' and a total score is calculated by summing ratings for all 21 items (range from 0 - 84), with higher scores indicating more activity limitations due to pain. The CALQ has one additional response option "not at all difficult" which was collapsed with the "not very difficult" response option for purposes of this factor analysis. The CALI-21 and CALQ Child and Parent versions have demonstrated reliability and validity in assessing pain-related activity limitations in school aged children and adolescents [5; 16]. A previous exploratory factor analysis on the CALI-21 yielded a 2-factor structure, including a Routine and an Active factor [16].

2.1.3. Pain Characteristics—Pain was assessed using youth report of both *pain intensity* and *pain frequency*. Specifically youth were asked to report on their "usual pain intensity" using a Numerical Rating Scale (11 point NRS 0-10). The NRS is a valid and reliable tool for assessing pain intensity in children and adolescents [25]. Reports of the frequency of pain symptoms were also assessed, with youth using a 6 point ordinal scale (0-5; 0 = "less than once/month" to 5 = "daily") to describe how often pain occurs.

2.2 Statistical Analysis

2.2.1. Data Processing and Summary—Item level analysis was conducted in order to identify items with extensive missing data or poor psychometric properties. First, the data on missing responses were examined for each item on the CALI-21. Item 10 ("Working at a job") and item 11 ("After school practices") had a large amount of missing values. For child report, 69% of subjects had no response for item 10 and 56% for item 11. For parent report, 76% of subjects had no response for item 10 and 59% did not answer item 11. Given the age and health status of these children, we decided that these two items may have had limited applicability in this sample (many did not have a job or participate on sports teams) so we elected to drop items 10 and 11 from the CALI-21 in all subsequent analyses. For the remaining 19 items, we *a priori* decided that if a subject was missing 7 or more items (35%) the subject did not provide data reliable enough for our analysis and should be dropped. For those with 6 or less items missing, hot-deck imputation [1] was performed on items with missing values. We conducted imputations for child report and parent report separately.

For participants with sufficient data (<7 out of the 19 were missing) the hot-deck imputation was conducted as follows. The child report or parent report with complete items formed the "donor pool". A response with missing items was then matched to one or more "donors" in the pool based on the distance (we used L1 norm) between items without missing values. The missing values were then replaced with the item values from the donor. If multiple donors were identified, one was then randomly chosen. Note with our cutoff, a child report or parent report may fall into one of the following three categories: 1) complete, 2) missing and imputation, or 3) missing and without imputation. Only subjects with both child report and parent report in the "missing and without imputation" category were excluded from subsequent analyses.

Summary statistics were used to describe characteristics of the dataset. Means and standard deviations were used for continuous data, and categorical items were described using frequency statistics. Chi-square tests and bivariate correlations were used to explore relations among demographic variables and self-report measures.

2.2.2. Exploratory and confirmatory factor analysis—Our next step was to re-examine the underlying factor structure of the CALI-21 using exploratory and confirmatory factor analyses employing commonly used psychometric methods [12]. An exploratory factor analysis (EFA) was previously conducted [16] but this sample was limited in size (n=155) and therefore the current study provides a much larger dataset for analysis. The current study did not include any data used in the previous validation study. Before conducting the factor analyses, the dataset was divided into separate estimation and validation samples. This was achieved by randomly allocating half of the sample within each site into the estimation sample and half to the validation sample.

We then conducted an EFA on the estimation sample separately using child report and parent report data. The EFA utilized an iterative principle factor estimation method and direct oblique rotation. We used eigenvalue at one and scree test to determine the number of factors. We kept all the items regardless of factor loadings. Initially items were assigned to

factors with the highest factor loadings. The assignments were then reviewed and reconciled to generate a common factor structure for both child report and parent report. Given the CALI-21 items were scored on a Likert scale, we also conducted the factor analysis using Spearman's rank based correlation instead of Pearson's correlation matrix. The results were identical using both methods.

Next we validated the common factor structure using a multiple-factor measurement model on the validation sample (Confirmatory Factor Analysis CFA). The model was fitted using structural equation model (SEM), with latent constructs corresponding to the factors. Using a commonly accepted approach, we examined goodness of fit (GOF) indices including likelihood ratio test statistic, root mean squared error of approximation (RMSEA), AIC/BIC, comparative fit index (CFI), Tucker-Lewis index (TLI), standardized root mean squared residual (SRMR) and coefficient of determination (CD) [19].

2.2.3. Revised scoring of the CALI-21—All 21 original CALI-21 items were rated on a 5-point scale for difficulty from (0) not very difficult to (4) extremely difficult. Previously the CALI-21 total score was based on available items reported. Because the total score depended on the number of observed items, one could not easily compare impairment in specific domains (Active and Routine) without having complete CALI-21 records. To address this we re-scored the CALI-21 using the following algorithm:

1. Step 1: Transform item score. Items are linearly transformed to a 0-100 scale as follows: 0=0, 1=25, 2=50, 3=75, and 4=100.

2. Step 2: Calculate factor scores

    a) If more than 50% of the items in a factor (i.e., subscale) are missing, the factor score is not computed.

    b) For those with 50% missing in a factor, the factor score is calculated as sum of the transformed item scores divided by the number of items answered.

3. Step 3: If both factor scores are available, then a total score is calculated as sum of factor scores divided by two. No total score is calculated if at least one factor cannot be scored. With the above scoring algorithm, the total CALI-21 score ranges between 0 and 100 regardless of number of items answered; with higher scores indicating greater limitations in function.

2.2.4. Reduction of CALI-21 items—To reduce the number of items and create a brief measure while retaining the factor structure, item reduction was conducted within each factor using commonly accepted strategies [21]. Applying the scoring algorithm to the estimation and validation samples, we calculated the factor scores based on candidate items for child report and parent report separately. Using the estimation sample, we regressed the factor score on individual items within the factor. We then applied the leaps-and-bounds algorithm [4] to identify the best combination of items with a given number of predictors in the model. For a given number of predictors, the optimal model was the one with the smallest value of AIC and BIC, the largest value of adjusted R-squared, and a value of

Mallow's Cp that was closest to the number of predictors in the model plus one or the smallest among the other Cp values. We determined a priori that the number of items left in the model should have to explain at least 90% of the variance (i.e., adjusted R-squared >=90%) in the factor scores based on full list of items. This selection process was performed for each factor and by child and parent report separately using the "vselect" add-on package in Stata [22]. We then combined the items identified from the child report and parent report for each factor to ensure a common factor structure across child and parent reports. For example, if items (4, 7, 13, 17, 19) were identified as the short list for one factor from the child report, and items (4, 7, 9, 17, 19) were identified as the short list for the same factor from the parent report, we would then consider the union of items (i.e., 4, 7, 9, 13, 17, 19) as the final common short list for the corresponding factor.

2.2.5. Preliminary Validation of the Shortened Measure—To validate the reduced items, we calculated the factor scores using both the full list of items and the reduced list of items in the validation sample. We then examined Pearson's correlation coefficients on the two sets of factor scores. We also regressed the factor scores based on the full list of items on the factor scores based on the reduced list of items, and examined the amount of variance explained (R-squared). A high value of R-squared would indicate the factor score based on the reduced list of items can be used in place of the factor score based on the full list of items. Additional regression analyses were conducted to examine relationships between the brief version of the CALI-21 and pain intensity. All data analyses were conducted using SPSS v20.0 [20] and Stata 12.1[22].

3. Results

3.1. Imputation for Missing Data and Descriptive Statistics

Combining data from the 3 sites resulted in a total sample of 1616 youth and 1614 parents. Among the 1616 youth participants, 234 (14.5%) had complete child report, 815 (50%) had 1-6 items missing, and 567 (35%) had 7 or more items missing. Of the parent reports, 207 (13%) had complete data, 831 (51%) had 1-6 items missing, and 578 (36%) had 7 or more items missing. 380 (23.5%) youth were removed from the dataset due to missing data on 7 or more items in both the child report and parent report.

We examined participants' rates of item completion for the remaining 19 items after dropping items 10 and 11. We elected to drop data from participants who had responses missing from 7 or more items (>=35%). This was done individually for parent and child reports. Hot deck imputation was conducted for participants with 6 or less missing CALI items, with on average 3 missing items (Median=3, IQR: 2-4) imputed for child report and 3 items (Median=3, IQR: 2-4) imputed for parent report. After imputation, among the final 1236 parent-child dyads, 851 had complete data in both child and parent reports, 198 had complete child report but incomplete parent report, and 187 had complete parent report but incomplete child report.

The final sample consisted of 1236 youth who were an average of 14.4 years (SD=2.4, range 8-18 years), 70.7% female, and 80% Caucasian. Average pain intensity was 6.3 (SD=2.0, 0-10 NRS) and the majority of youth (79.8%) reported pain frequency of   4 times/week.

Among the 380 (23.5% of total 1616) youth removed from analyses, 360 (94.7%) were from Site 1, 18 (4.7%) from Site 2, and two (0.5%) from Site 3. Within Site 1, youth dropped from analyses were slightly younger (13.7 vs. 14.1, p=0.02), more likely to be Caucasian (72.5% vs. 80.1%, p=0.04), had less frequent daily pain (38.6% vs. 55.7%, p<.001), and reported less severe pain (5.9 vs. 6.6, p<.001; 0-10 NRS) than the remaining sample at Site 1. The 18 participants dropped from Site 2 reported significantly higher pain (7.0 vs. 5.7, p=0.01) than the remaining sample in Site 2, but did not differ significantly in other pain or demographic characteristics.

3.2. Exploratory Factor Analysis

For the exploratory factor analysis, 616 records (49.8%) were randomly assigned to the estimation sample and 620 (50.2%) to the validation sample. For the child report, two factors were identified in exploratory factor analysis using the estimation sample (Factor 1: eigenvalue=8.25, proportion of variance explained=62.04%; Factor 2: eigenvalue=1.67, proportion of variance explained=12.57%). The factor loadings for child report after the oblimin rotation can be found in Table 2. Note because of the oblimin rotation, the loadings can only be interpreted as weights, not correlations to the factors. For the parent report, two factors were identified using the estimation sample (Factor 1: eigenvalue=8.84, % variance explained=64.67%; Factor 2: eigenvalue=1.61, % variance explained=11.78%).

As shown in Table 2, items 1, 3, 4, 8, 16, 17, 19, and 21 are clearly loaded more heavily onto Factor 2, while items 2, 5, 14, 15, 18, and 20 are clearly loaded more heavily onto Factor 1. Items 9, 12 and 13 have loadings $\leq$ 0.3 for both factors. But all three are consistently more heavily loaded unto Factor 2 therefore they are all classified into Factor 2. Based on the factor loadings, child report and parent report produced almost identical factor structures, with the exception being items 6 ("Doing a Hobby") and 7 ("Playing with Friends"). These two items showed opposite directions on factor loadings. We reviewed the content on these 2 items and made a final decision to assign both to Factor 2. The final factors were as follows: Factor 1 - items 2, 5, 14, 15, 18, 20 (6 total), and factor 2: items 1, 3, 4, 6, 7, 8, 9, 12, 13, 16, 17, 19, 21 (13 total). This factor structure is the same for child and parent reports.

Using the validation sample, the next step was to calculate the mean (SD) of each factor on the parent and child reports using the 2 factor structure presented above. The child report Routine factor score had a mean of 41.62 (20.78) and an Active factor score of 54.72 (25.34). Parent report Routine factor score had a mean of 43.14 (21.76) and an Active factor score of 56.14 (27.49). Pairwise correlations between child and parent factor scores showed they are strongly correlated. The correlation between the child report Routine score and parent report Routine score was 0.66 (p<.001). The correlation between the child report Active score and parent report Active score was 0.71 (p<.001). The Cronbach's α for Routine items was 0.89 for child report, and 0.91 for parent report. The Cronbach's α for Active items was 0.86 for child report, and 0.90 for parent report. These values support the same 2-factor structure as originally identified in the CALI-21 and demonstrate that the 2-factor structure has good reliability.

3.3. Confirmatory Factor Analysis

We fitted the two-factor measurement models to both the child report and parent report data in the validation sample using structural equation models. Similar to the EFA, the two-factor model demonstrated good fit for both parent and child report. Common thresholds for goodness of fit indices are: 0.07 on RMSEA, 0.90 on GFI, 0.05-0.08 on SRMR, and 0.95 on CFI. For the child report, fit indices from the CFA yielded a 2 factor solution (RMSEA=.10, coefficient of determination =.98, comparative fit index =.84, Tucker-Lewis index =.82). For the parent report, fit indices from the CFA yielded a 2 factor solution (RMSEA=.10, coefficient of determination =.99, comparative fit index =.88, Tucker-Lewis index =.86). See Figure 1.

3.4. Item Reduction

As reported above, items 6 and 7 showed inconsistent loading patterns across child and parent report, suggesting these items may be interpreted differently by children and parents. Due to this discrepancy across reporters, these items were excluded from possible inclusion in the brief measure. This left 11 items in Routine factor and 6 items in Active factor for the item reduction process. In the following item reduction and validation, Routine factor score based on 11 item and Active factor score based on 6 items were used exclusively. For item reduction we used linear regression and variable selection techniques to identify a smaller set of items that captured the majority of variance in the factor scores. The leaps-and-bounds algorithm [4] identified the best combination of items as predictors for each of the four factor scores (child report Routine score, child report Active score, parent report Routine score, and parent report Active score), conditional on the number of predictors. This algorithm was applied separately to child report and parent report. Unions of the set of items identified from child and parent reports are considered the final results allowing for a common structure for both child and parent report.

By applying the re-scoring algorithm to both the child report and parent report in the *estimation sample,* we were able to generate the factor scores for the Routine and Active factors, respectively. The results are summarized below. The resulting brief measure ("CALI-9") has 9 items, 6 Routine and 3 Active. The final items on the Routine factor are: 4, 12, 16, 17, 19 and 21. This set of 6 items explains 94% variation in the factor score calculated from all 11 items on the Routine for child report and 95% variation for parent report. The final items on the Active factor are: 5, 14, and 18. This set of 3 items explains 93% of variation in the factor score calculated from all 6 items on the Active factor for child report, and 93% variation for parent report. The CALI-9 parent and child versions as well as detailed scoring instructions are located in Appendices A & B.

3.5. Preliminary validation of CALI-9

Using the *validation sample*, we calculated the Routine scores using all 11 items from the complete Routine factor and the reduced list of 6 items, for child report (Routine 11c vs. Routine 6c) and parent report (Routine 11p vs. Routine 6p), respectively. Similarly, we calculated the Active scores using all 6 items from the complete Active factor and the reduced list of 3 items for child report (Active 6c vs. Active 3c) and parent report (Active 6p vs. Active 3p), respectively. See Table 3.

On the child report, the correlation between the 11 item and 6 item Routine factors was 0.95 (p<.001) and the correlation between the 6 and 3 item Active factors was 0.95 (p < .001), demonstrating that both shortened child factors performed comparably to the longer item sets. To further validate the factor structure of the CALI-9 child report, a linear regression model analysis was performed. Under linear regression without intercept, Routine 6c captured 98.06% variation in Routine 11c (R^2 =0.98); Active 3c captured 98.16% of variation in Active 6c (R^2 =0.98). Similar patterns of associations emerged using the parent data (see Table 3). These high values of R^2 confirmed that there was little information loss from factor scores calculated with the reduced list of items.

For additional validation of the CALI-9, we evaluated associations among CALI-9 subscale scores and individual factors (child age, sex and BMI) using the full sample. T-tests examining group differences in younger (8-12 years) versus older (13-18 years) youth revealed that younger children reported significantly *fewer* limitations on the Routine subscale than older children (40.12 vs. 43.81, p=0.02). There were no other age differences in CALI-9 total or subscales scores for either parent or child report. Sex differences among CALI-9 scores also emerged. Specifically, girls reported *greater* limitations on the Active subscale than boys (60.10 vs. 54.47, t(1126)= 3.13, p=.002), and *greater* limitations on the CALI-9 total score than boys (51.89 vs. 47.97, t(1104)=2.70, p=0.007). Parent-report of child activity limitations was not different by child sex. Associations among BMI and CALI-9 subscales were significant, with higher BMI correlated with greater activity limitations on Active and Routine subscales using parent and child report data (Pearson's correlation coefficient r's = .10 to .13, p's <.001).

The patterns of associations between pain intensity and parent and child CALI-9 subscale scores were examined using the full sample. Specifically, child report of usual pain intensity (0-10 NRS) was grouped into three categories: 0-3 NRS, 4-7 NRS, and 8-10 NRS. These cut-offs were selected based on recent data suggesting cut-offs of 4 for moderate and 8 for severe pain in youth [6]. The distribution of youth who reported pain intensity in each category is as follows: missing (4.21%), 0-3 NRS (9.14%), 4-7 NRS (58.66%) and 8-10 NRS (27.99%). Boxplots indicate higher pain intensity is associated with higher CALI-9 scores on both the Active (Figure 2) and Routine (Figure 3) subscales.

4. Discussion

Results of this study show that the CALI-9 is a promising brief assessment tool that can be used to rate pain-related activity limitations in children and adolescents with chronic pain. Using a large, multi-site sample of children and adolescents with chronic pain, the length of the CALI-21 was shortened from 21 to 9 items, with minimal loss in reliability, validity and variability. A common factor structure was identified for both child report and parent report forms, which further improves usability. The CALI-9 demonstrates good internal consistency for both child and parent forms and high cross-informant reliability. Initial validity is demonstrated by associations among CALI-9 score and pain intensity.

The accurate assessment of pain-related disability is important to both clinicians and researchers in pediatric chronic pain. This revised version of the CALI-21 presents the first

brief tool for assessing pediatric pain-related activity limitations that provides scoring for specific domains of activity impairment. Similar to the full measure, the exploratory and confirmatory factor analyses demonstrated a 2-factor fit to the data that provides assessment of Active and Routine domains. Assessment of disability in both Active and Routine domains enables clinicians and researchers to more specifically identify the type of pain-related limitations that are experienced by the child. This is particularly relevant as pain experiences vary, and prior research has demonstrated that individual children are differentially impacted in their functional activities [28]. For example, in the original validation sample, [16] children with headache pain reported fewer limitations in the Active domain and greater limitations in the Routine domain than youth with other pain diagnoses. Additionally, children with comorbid disease and their parents reported greater limitations in the Active domain and parents reported higher total disability scores in their children, than parents of those without comorbid disease. The ability to assess the impact of pain in these two domains on the CALI-9 may allow for tailored treatment approaches, based on diagnosis and individual need, and increased sensitivity in detecting changes following treatment.

The CALI-9 has two significant improvements that will increase ease and interpretability in future use. First, the CALI-9 eliminated rarely endorsed or frequently skipped items on the CALI-21, resulting in a brief and highly relevant measure for assessment of pain-related impairment. Additionally, with further validation the CALI-9 may be used as a daily measure to assess short-term trajectories of pain-related disability or changes in patient functioning within the context of treatment. The two existing studies examining in-treatment activity limitations data are limited by use of an uncontrolled study [10] and lack of a standard item set in which participants rated disability [15]. Future research can use the CALI-9 to track change over time and determine whether disability can be modified by treatment or fluctuates with changes in symptoms, such as pain intensity or mood. The CALI-9 is a more precise measure of pain-related activity limitations than the PROMIS Pain Interference Scale, which contains an affective component (e.g., "I felt angry when I had pain", "It was hard to have fun when I had pain", and "It was hard for me to pay attention when I had pain"). Further, the CALI-9 maintains two subscales to assess separate domains of function which may also be useful for hypotheses testing the impact of various treatment approaches on the domains of active and routine function over time. Sex differences found on CALI-9 Active subscale also supports separate scoring of active versus routine domains. While a previous large-scale validation study of the FDI [8] did not find sex differences in pain-related disability in boys versus girls, it is possible that sex differences in FDI scores did not emerge because the FDI total score does not allow for examination of individual domains. Subscale scoring may also support targeted interventions focusing on areas of greatest impairment.

A second significant improvement involves the new scoring algorithm for the CALI-9, which takes into account missing items and number of items answered. The improved scoring system has a more straightforward interpretation of a child's pain-related disability on a 0-100 scale for both a total score and on the Active and Routine factors with higher scores indicating greater activity limitations. Using a 0-100 point scale, the CALI-9 is similar to measures of health-related quality of life. This new scoring system permits ease of

interpretation in tracking changes in pain-related disability over time, giving clinicians and researchers the ability to readily see improvement or decline in children's pain-related disability over time.

Important future directions include examining test-retest reliability and discriminant validity of the CALI-9 as well as administering the revised version to new samples of youth testing both the retrospective format (e.g., activity limitations in last 4 weeks) and daily diary report version (e.g., activity limitations in last 24 hours). Test-retest reliability will identify the stability in pain-related disability over a short period (e.g., 1 week, 2 weeks) and how sensitive the measure is in predicting clinical outcomes. For example, it is possible that the measure can be used as a screening tool to identify patients at greatest risk for poor outcomes in certain domains (e.g., significant limitations in the Active domain) or be used to track progress in a domain over time. Tests of discriminant validity would provide data on the ability of the CALI-9 to separate patients and participants based on pain diagnosis, pain location, or individual characteristics, such as BMI. Another important direction is to validate the CALI-9 using measures of objective assessment methods of activity such as an accelerometer. Previous research has demonstrated significant associations among peak activity level and sedentary time (assessed via accelerometer) and CALI-21 scores [27].

Findings of this study should be considered in light of several limitations. While the study included a large sample size with youth and parents from different geographical areas in the U.S, the multi-site sample lacked racial diversity. Data were compiled across three institutions and two geographic regions, however 80% of the sample was Caucasian and 70% of youth were female. While these numbers reflect typical compositions of patients seen in specialty pediatric pain clinics, findings may not be generalizable to all youth with chronic pain. An additional limitation is that due to reporting differences across sites, data on child pain location and parent/family level variables were not available for inclusion in analyses. Data on pain location were collected using a variety of formats (e.g., multiple choice items, a body diagram) and questions (e.g., primary pain location, identifying all pain locations on body). Given this variability in assessment of pain location, we could not summarize pain location across the whole sample. Future research can examine how CALI-9 (total and individual factor) scores differ by children's pain location or pain etiology (disease versus non-disease related). It will also be important to examine associations among CALI-9 scores and measures of psychological distress/emotional function (e.g., pain catastrophizing, depressive symptoms, fear of pain) not available with the current data set given established associations among these factors and pain-related disability in youth (e.g., [2; 23]). Finally, all participants who completed the CALI-21 reflect clinical samples of youth with chronic pain. Responses from clinical samples may not generalize to general populations of youth with chronic pain, making validation of the CALI-9 with non-clinical samples an important next step.

It is also important to recognize that missing data could have influenced our results. Approximately 50% of youth and parents were missing data on 1-6 items of the CALI-21 data included in the EFA, CFA, and item reduction analyses. While the robust sample size and data analytic methods were able to accommodate this missing data, it is important to note that youth who completed the questionnaires as part of clinical intake had more missing

data than youth whose data was collected as part of a research study. This suggests that study methods in research (e.g., electronic data collection) may have more effectively prevented item-level missingness, and also points to how measures may be impacted by the demands of the clinical environment. The shortened version (including elimination of less applicable/irrelevant items) is expected to improve the degree to which patients, participants and parents complete the measure in the context of routine clinical care.

In summary, the CALI-9 is a promising brief assessment tool that can be used to rate pain-related limitations and track outcomes over time among youth with chronic pain. The measure includes both child and parent reports with a common factor structure and scoring that is easily interpretable. The factor analyses demonstrated that the CALI-9 can be used to assess limitations in active and routine domains, which may allow for better understanding of ways in which pain impacts the day-to-day functioning of children and adolescents with chronic pain.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Andridge RR, Little RJ. A Review of Hot Deck Imputation for Survey Non-response. Int Stat Rev. 2010; 78(1):40–64. [PubMed: 21743766]

2. Caes L, Fisher E, Clinch J, Tobias JH, Eccleston C. The role of pain-related anxiety in adolescents' disability and social impairment: ALSPAC data. Eur J Pain. 2015; 19(6):842–851. [PubMed: 25380528]

3. Claar RL, Walker LS. Functional assessment of pediatric pain patients: psychometric properties of the functional disability inventory. Pain. 2006; 121(1-2):77–84. [PubMed: 16480823]

4. Furnival GM, Wilson RW. Regressions by leaps and bounds. Technometrics. 1974; 16(4):499–511.

5. Hainsworth KR, Davies WH, Khan KA, Weisman SJ. Development and preliminary validation of the Child Activity Limitations Questionnaire: flexible and efficient assessment of pain-related functional disability. J Pain. 2007; 8(9):746–752. [PubMed: 17631055]

6. Hirschfeld G, Zernikow B. Variability of "optimal" cut points for mild, moderate, and severe pain: neglected problems when comparing groups. Pain. 2013; 154(1):154–159. [PubMed: 23182623]

7. Huguet A, Miro J. The severity of chronic pediatric pain: an epidemiological study. J Pain. 2008; 9(3):226–236. [PubMed: 18088558]

8. Kashikar-Zuck S, Flowers SR, Claar RL, Guite JW, Logan DE, Lynch-Jordan AM, Palermo TM, Wilson AC. Clinical utility and validity of the Functional Disability Inventory among a multicenter sample of youth with chronic pain. Pain. 2011; 152(7):1600–1607. [PubMed: 21458162]

9. Lewandowski AS, Palermo TM, Kirchner HL, Drotar D. Comparing diary and retrospective reports of pain and activity restriction in children and adolescents with chronic pain conditions. Clin J Pain. 2009; 25(4):299–306. [PubMed: 19590478]

10. Lynch-Jordan AM, Sil S, Peugh J, Cunningham N, Kashikar-Zuck S, Goldschneider KR. Differential changes in functional disability and pain intensity over the course of psychological treatment for children with chronic pain. Pain. 2014; 155(10):1955–1961. [PubMed: 24954165]

11. McGrath PJ, Walco GA, Turk DC, Dworkin RH, Brown MT, Davidson K, Eccleston C, Finley GA, Goldschneider K, Haverkos L, Hertz SH, Ljungman G, Palermo T, Rappaport BA, Rhodes T, Schechter N, Scott J, Sethna N, Svensson OK, Stinson J, von Baeyer CL, Walker L, Weisman S, White RE, Zajicek A, Zeltzer L. Core outcome domains and measures for pediatric acute and chronic/recurrent pain clinical trials: PedIMMPACT recommendations. J Pain. 2008; 9(9):771–783. [PubMed: 18562251]

12. Mulaik, SA. Foundations of Factor Analysis. Boca Raton, FL: CRC Press; 2009.

13. Palermo TM. Assessment of chronic pain in children: current status and emerging topics. Pain Res Manag. 2009; 14(1):21–26. [PubMed: 19262912]

14. Palermo TM, Law EF, Fales J, Bromberg MH, Jessen-Fiddick T, Tai G. Internet-delivered cognitive-behavioral treatment for adolescents with chronic pain and their parents: a randomized controlled multicenter trial. Pain. 2016; 157(1):174–185. [PubMed: 26335910]

15. Palermo TM, Law EF, Zhou C, Holley AL, Logan D, Tai G. Trajectories of change during a randomized controlled trial of internet-delivered psychological treatment for adolescent chronic pain: how does change in pain and function relate? Pain. 2015; 156(4):626–634. [PubMed: 25775441]

16. Palermo TM, Lewandowski AS, Long AC, Burant CJ. Validation of a self-report questionnaire version of the Child Activity Limitations Interview (CALI): The CALI-21. Pain. 2008; 139:644–652. [PubMed: 18692316]

17. Palermo TM, Wilson AC, Peters M, Lewandowski A, Somhegyi H. Randomized controlled trial of an Internet-delivered family cognitive-behavioral therapy intervention for children and adolescents with chronic pain. Pain. 2009; 146(1-2):205–213. [PubMed: 19695776]

18. Palermo TM, Witherspoon D, Valenzuela D, Drotar DD. Development and validation of the Child Activity Limitations Interview: a measure of pain-related functional impairment in school-age children and adolescents. Pain. 2004; 109(3):461–470. [PubMed: 15157707]

19. Rencher, AC., Christensen, WF. Methods of Multivariate Anaysis Hoboken. New Jersey; John Wiley & Sons; 2012.

20. SPSS. SPSS for Windows, v 20 0. Chicago, IL: SPSS Inc.; 2011.

21. Stanton JM, Sinar FE, Balzer WK, Smith PC. Issues and strategies for reducing the length of self-report scales. Person Psychol. 2002; 55:167–194.

22. StataCorp. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP; 2011.

23. Tran ST, Jastrowski Mano KE, Hainsworth KR, Medrano GR, Anderson Khan K, Weisman SJ, Davies WH. Distinct influences of anxiety and pain catastrophizing on functional outcomes in children and adolescents with chronic pain. J Pediatr Psychol. 2015; 40(8):744–755. [PubMed: 25840447]

24. Varni JW, Stucky BD, Thissen D, Dewitt EM, Irwin DE, Lai JS, Yeatts K, Dewalt DA. PROMIS Pediatric Pain Interference Scale: an item response theory analysis of the pediatric pain item bank. J Pain. 2010; 11(11):1109–1119. [PubMed: 20627819]

25. von Baeyer CL, Spagrud LJ, McCormick JC, Choo E, Neville K, Connelly MA. Three new datasets supporting use of the Numerical Rating Scale (NRS-11) for children's self-reports of pain intensity. Pain. 2009; 143(3):223–227. [PubMed: 19359097]

26. Wilson AC, Moss A, Palermo TM, Fales JL. Parent pain and catastrophizing are associated with pain, somatic symptoms, and pain-related disability among early adolescents. J Pediatr Psychol. 2014; 39(4):418–426. [PubMed: 24369365]

27. Wilson AC, Palermo TM. Physical activity and function in adolescents with chronic pain: a controlled study using actigraphy. J Pain. 2012; 13(2):121–130. [PubMed: 22099608]

28. Wilson AC, Samuelson B, Palermo TM. Obesity in children and adolescents with chronic pain: associations with pain and activity limitations. Clin J Pain. 2010; 26(8):705–711. [PubMed: 20664337]
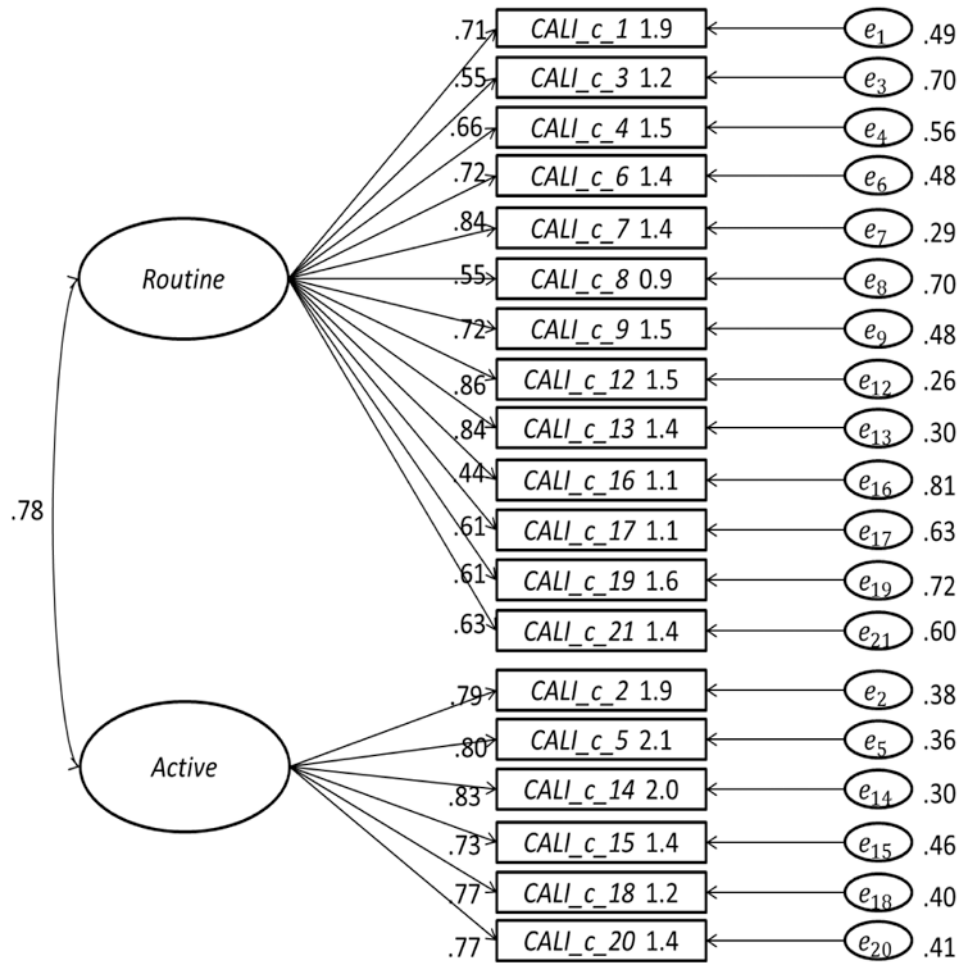
Figure 1.
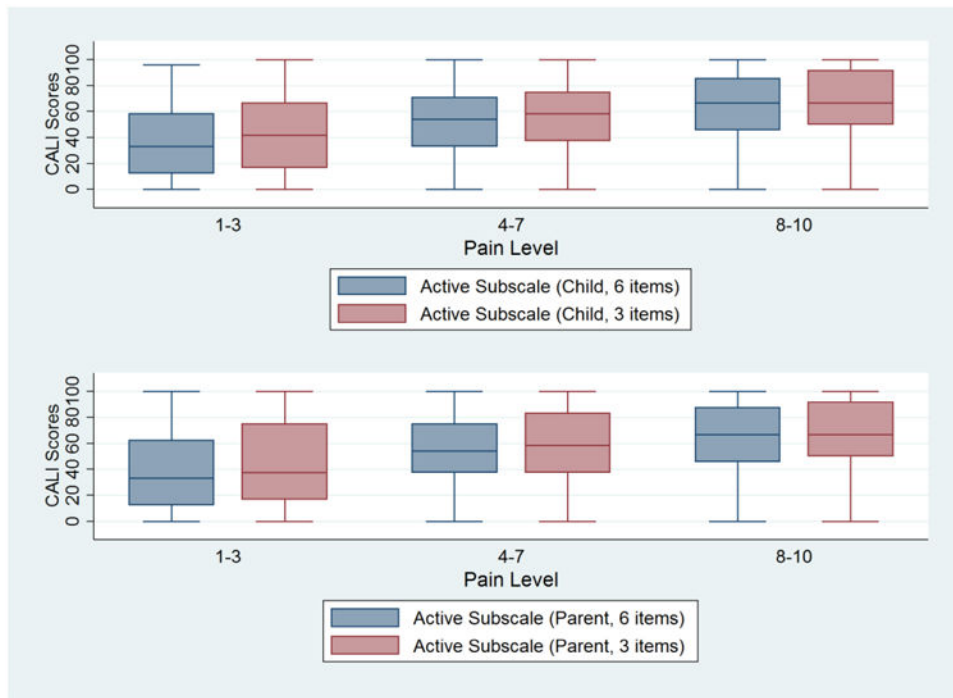CFA with two factor latent structure models on child report.

Figure 2.
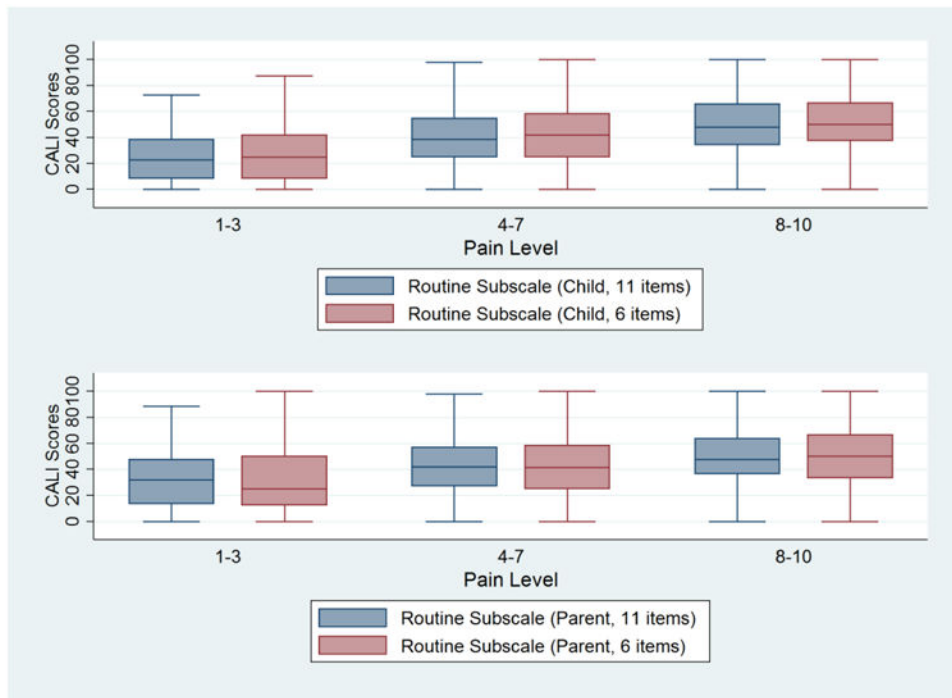Associations among usual pain intensity and Active subscale scores comparing 6 item and 3 item versions.

Figure 3.
Associations among usual pain intensity and Routine subscale scores comparing 11 item and 6 item versions.

Table 1

Descriptives of study sample.

Variable	Total (N=1236)	Estimation sample (N=616)	Validation sample (N=620)	p-values
Age in years, Mean (SD)	14.44 (2.37)	14.47 (2.35)	14.40 (2.38)	0.59
Sex, N(%)				0.94
Female	874 (70.71)	435 (70.62)	439 (70.81)	
Male	362 (29.29)	181 (29.38)	181 (29.19)	
Race/ethnicity, N(%)				0.54
Caucasian/White	995 (80.50)	499(81.01)	496 (80.00)	
African American/Black	79 (6.39)	43 (6.98)	36 (5.81)	
Asian	11 (0.89)	5 (0.81)	6 (0.97)	
Native American	13 (1.05)	8 (1.30)	5 (0.81)	
Other or multi-racial	123 (9.95)	56 (9.09)	67 (10.81)	
Missing	15 (1.21)	5 (0.81)	10 (1.61)	
BMI, Mean (SD)	23.64 (6.32)	23.80 (6.55)	23.47 (6.09)	0.42
Usual Pain Intensity (0-10 NRS), Mean (SD)	6.29 (2.01)	6.38 (1.99)	6.21 (2.02)	0.16
Usual Pain Frequency				0.67
Less than once a month	12 (0.97)	8 (1.30)	4 (0.65)	
1-3 times per month	19 (1.54)	11 (1.79)	8 (1.29)	
Once a week	19 (1.54)	8 (1.30)	11 (1.77)	
2-3 times a week	148 (11.97)	70 (11.36)	78 (12.58)	
4-6 times a week	244 (19.74)	116 (18.83)	128 (20.65)	
Daily	742 (60.03)	374 (60.71)	368 (59.35)	
Missing	52 (4.21)	29 (4.71)	23 (3.71)	
Sample size by study sites, N(%)				0.99
Site 1	742 (60.03)	371 (60.23)	371 (59.84)	
Site 2	263 (21.28)	131 (21.27)	132 (21.29)	
Site 3	231 (18.69)	114 (18.51)	117 (18.87)	

Table 2

Rotated factor loadings from EFA on estimation sample

CALI item	Content	Child report[1]		Parent report	
		Factor #1	Factor #2	Factor #1	Factor #2
2	Gym	**0.76**	-0.07	**0.65**	0.17
5	Sports	**0.82**	-0.03	**0.74**	0.09
14	Running	**0.88**	-0.14	**0.88**	-0.09
15	Walking up stairs	**0.67**	0.07	**0.79**	-0.02
18	Walking one or two blocks	**0.67**	0.11	**0.72**	0.10
20	Riding a bike or scooter	**0.70**	0.15	**0.83**	-0.02
1	Going to school	0.21	**0.55**	0.14	**0.67**
3	Reading	-0.17	**0.79**	-0.18	**0.83**
4	Schoolwork	-0.07	**0.79**	-0.14	**0.87**
6	Doing a hobby	**0.49**	0.31	0.29	*0.50*
7	Playing with friends	**0.49**	0.39	0.35	*0.55*
8	Watching TV	-0.01	**0.69**	0.01	**0.64**
9	Housework or chores	0.41	**0.43**	0.37	**0.47**
12	Doing things with friends	0.44	**0.46**	0.33	**0.59**
13	Going to clubs/church activities	0.41	**0.48**	0.33	**0.55**
16	Eating regular meals	0.15	**0.47**	-0.004	**0.52**
17	Riding in the school bus or car	0.13	**0.52**	0.23	**0.41**
19	Sleep	0.20	**0.38**	0.10	**0.40**
21	Being up all day (without a nap or rest)	0.20	**0.42**	0.20	**0.43**
Eigenvalue[2]		8.25	1.67	8.84	1.61
% variance[2]		62.04	12.57	64.67	11.78

[1] Factor loadings are based on oblique rotation.

[2] Unrotated.

Table 3

Routine and active sub-scale scores based on child report and parent report on the validation sample

Variable	N	Mean (SD)	Variable	N	Mean (SD)	r	R^2
Child			**Child**				
11 item Routine	574	41.31 (20.91)	6 item Routine	580	42.97 (22.20)	0.95	0.98
6 item Active	570	54.75 (25.34)	3 item Active	572	58.54 (26.85)	0.95	0.98
Parent			**Parent**				
11 item Routine	559	43.26 (21.59)	6 item Routine	573	43.71 (22.80)	0.96	0.98
6 item Active	569	56.14 (27.49)	3 item Active	568	59.23 (28.85)	0.96	0.99