**Article:**
Benson, P.E. orcid.org/0000-0003-0865-962X, Alshawy, E., Kettle, J.E. orcid.org/0000-0002-2776-1243 et al. (1 more author) (2022) Development of the orthodontic treatment impact questionnaire: cross-sectional validation. American Journal of Orthodontics and Dentofacial Orthopedics, 162 (4). e183-e191. ISSN 1097-6752

https://doi.org/10.1016/j.ajodo.2022.06.018

# Development of the Orthodontic Treatment Impact Questionnaire: Cross-sectional validation

**Philip E. Benson,[a] Ebrahim Alshawy,[b] Jennifer E. Kettle,[a] and Fiona Gilchrist[a]**
*Sheffield, United Kingdom, and Buraydah, Qassim, Saudi Arabia*

**Introduction:** The need to involve patients in developing and evaluating health care interventions is now well-recognized. This study assesses and refines the Orthodontic Treatment Impact Questionnaire for use as a patient-reported outcome in an interventional clinical trial to evaluate and compare any orthodontic interventions. **Methods:** The face and content validity of a previously developed questionnaire were tested in 2 focus groups involving adolescents aged 11-17 years. They were wearing a range of orthodontic appliances and at different treatment stages. A similar cross-sectional convenience sample completed the questionnaire during routine appliance adjustment appointments. A Rasch model, using item response theory, was used for item reduction, assessment of the response format, and differential item functioning. Spearman's rank correlation was used to assess construct validity, Cronbach $\alpha$ for internal consistency and reliability, and intraclass correlation coefficient for test-retest reliability. **Results:** Seven adolescents (4 females, 3 males) were involved in the initial testing; 181 (117 females, 64 males; mean age, 14.7 $\pm$ 1.5 years) completed the questionnaire once and 41 twice. The initial measure demonstrated a misfit to the Rasch model. Ten of the original 31 items had disordered thresholds and were removed. The 5-point scale was changed to a 3-point scale. None of the participants demonstrated a misfit to the model. Construct validity ($P$ = 0.480), internal consistency (Cronbach $\alpha$ = 0.827) and test-retest reliability (intraclass correlation coefficient = 0.85; 95% confidence interval, 0.73-0.92) were good. **Conclusions:** The initial Orthodontic Treatment Impact Questionnaire was tested and modified using item response theory. The modified questionnaire demonstrated good construct validity, reliability, and internal consistency. Further testing to assess generalizability and longitudinal responsiveness is required. (Am J Orthod Dentofacial Orthop 2022;162:e183-e191)

Understanding a patient's experience is essential for successfully developing, investigating, and evaluating any health care approach. Prospective observational and interventional clinical research should include the patient voice, usually in the form of patient-centered outcomes, to determine the effectiveness and efficiency of treatments designed to improve health.[1]

Some orthodontic clinical studies have used generic measures of oral health–related quality of life as a patient-reported outcome to assess the impacts of orthodontic appliances on a patient's day-to-day life.[2] However, these measures were developed and validated to assess the generic impacts of oral and dental conditions and may not identify the particular effects of treatments on the day-to-day experience of an individual. Other studies have used specific questionnaires without details about how these measures were initially developed and tested.[3]

Feldmann et al[4] used 3 focus groups involving adolescents who had recently completed orthodontic treatment and 1 involving parents to develop a questionnaire to assess motivations, expectations, and treatment experiences, but the qualitative aspects of the questionnaire development are poorly reported.

Yassir et al[5] report the evaluation of questionnaires to assess patient perceptions of fixed orthodontic

[a]Academic Unit of Oral Health, Dentistry and Society, School of Clinical Dentistry, University of Sheffield, Sheffield, United Kingdom.
[b]Department of Orthodontics and Pediatric Dentistry, College of Dentistry, Qassim University, Buraydah, Qassim, Saudi Arabia.
All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest, and none were reported.
This work was supported by the British Orthodontic Society Foundation, London, United Kingdom.
Address correspondence to: Philip E. Benson, Academic Unit of Oral Health, Dentistry and Society, School of Clinical Dentistry, University of Sheffield, Sheffield S10 2TA, United Kingdom; e-mail, p.benson@sheffield.ac.uk.
Submitted, October 2021; revised and accepted, June 2022.
0889-5406

https://doi.org/10.1016/j.ajodo.2022.06.018

appliances before, during, and after treatment. Twenty-two clinicians and 30 patients assisted the investigators in assessing the content and face validity of the questionnaires and the internal consistency, but there are no details about how the initial items were chosen or reduced. The report has no information about the patient's age, treatment stage, or specific treatment. Construct validity, criterion validity, test-retest reliability, and responsiveness were not tested, and it is unclear how the items were scored and analyzed. Similar issues arise with another recently developed questionnaire.[6]

Mandall et al[7] carried out semistructured qualitative interviewing to identify questions to include in a measure designed to assess the impact of fixed appliances. The questionnaire was only used as a primary outcome in 1 randomized clinical trial.[8] One limitation of this questionnaire is that it was developed specifically for participants wearing fixed orthodontic appliances. We believe there is a need to develop a questionnaire that can be used to assess the specific impacts of any orthodontic treatment experienced by adolescents. This would allow the measure to be applied as a patient-reported outcome in clinical trials and audits to evaluate and compare any type of orthodontic appliance; for example, removable aligners vs fixed appliances or removable functional appliances vs fixed functional appliances. Separate questionnaires for removable and fixed appliances would not allow this comparison.

This investigation aimed to refine and undertake a preliminary evaluation of the psychometric properties of a newly developed measure, the Orthodontic Treatment Impact Questionnaire (OTIQ version 1). The early development of the questionnaire, using qualitative methods, has been reported.[9]

The specific objectives were (1) to use item response theory (IRT) to reduce the items that had been identified through qualitative inquiry; and (2) to undertake a cross-section evaluation of the psychometric properties of construct validity, internal consistency, and test-retest reliability of the newly developed questionnaire.

## MATERIAL AND METHODS

The questionnaire is designed to assess the self-reported impacts of wearing any type of orthodontic appliance on the oral health of adolescents. The adolescents were aged 11-17 years, a common age for orthodontic treatment, during the early permanent dentition. Oral health is defined by the Fédération Dentaire Internationale World Dental Federation as "the ability to speak, smile, smell, taste, touch, chew, swallow, and convey a range of emotions through facial expressions with confidence and without pain, discomfort, and disease of the craniofacial complex (head, face, and oral cavity)."[10] The impacts were on oral function, pain, discomfort, and social and emotional well-being.

The questionnaire was developed using the mixed methods methodology described by Guyatt et al.[11] This was undertaken in 2 stages: (1) stage 1 involved conducting and analyzing individual interviews with adolescents wearing an orthodontic appliance to discuss their experiences and identify potential questions to include in the measure; and (2) stage 2 involved validating the measure in 2 phases. Study 1 consisted of qualitative focus groups or 1-to-1 interviews to test the face and content validity of the newly developed measure. This was to ensure that participants understood the items, instructions, and response format of the questionnaire and that no important items had been omitted. Study 2 consisted of a quantitative, cross-sectional evaluation of the instrument with the recruitment of a convenience sample.

Stage 1 has been reported previously.[9] The initial questionnaire developed in stage 1 is shown in Supplementary Material 1. It consisted of 31 specific impact questions derived from stage 1: 1 on analgesic use, 1 on global questions that allowed the participants to indicate how their brace or retainer affects their life overall, and 6 free text boxes. The lead-in question was either (1) Because of my brace or retainer I have been bothered by or (2) I have been bothered by…, or (3) Because of my brace or retainer. The recall time was since the patients' last visit to the orthodontist, and the response format for the impact questions and the global question was a 5-point scale from 0 (not a lot) to 4 (a lot). The response format for the analgesia question included yes, no, or don't know.

This is the report for stage 2.

Ethical approval was obtained from the North East Newcastle and North Tyneside Research Ethics Committee (ref no. 16/NE/0367-November 2016).

Participants were adolescents aged 11-17 years, attending the Orthodontic Department of the Charles Clifford Dental Hospital, Sheffield Teaching Hospitals National Health Service Foundation Trust, and wearing any orthodontic appliance, including removable, functional, and fixed appliances. Patients at different stages of treatment, including retention, were recruited. There were no restrictions for ethnicity or gender, but patients with a cleft of the lip and or palate, a severe skeletal discrepancy undergoing a combined orthodontic-orthognathic surgical treatment, or who had a complicated medical history or difficulties understanding English were excluded. The socioeconomic status of the participants in stage 2 study 2 was recorded using the index of multiple deprivation (IMD) score on the

basis of home postcode.[12] The ethnicity of the participants was not recorded.

Potential participants were identified and approached on the telephone before or when they attended their routine appliance adjustment appointment. The purpose of the study was explained to them and their parent, and they were provided with written information about the project. Participants in stage 2 of study 1 were given 2-3 weeks to think about taking part and then contacted by the research assistant to confirm if they wanted to take part or not. Arrangements were made for participants to attend focus groups, which were undertaken in a quiet room in the School of Clinical Dentistry, away from the clinic, usually at their routine appliance adjustment appointment. Written consent was provided by the participants aged ≥16 years or their parents if aged <16 years.

Participants in stage 2 study 2 were asked to complete the questionnaire on the same day. Formal written consent was not obtained from participants or their parents, but consent was implied by completing the questionnaire, which was undertaken in a quiet part of the clinic. The questionnaire was formatted as a Google Form on an electronic device (Kindle Fire 7; Amazon, Seattle, Wash). The clinician entered some basic demographic (study number, gender, and age) and clinical (type of appliance and date of fitting) data. The participant completed the remainder of the questionnaire independently. The last question asked if they would agree to complete the questionnaire again to assess the repeatability of the questionnaire. A printed copy of the questionnaire was posted to the home addresses of those who agreed to repeat the questionnaire, after at least 2 weeks, along with a prepaid return envelope. The repeat questionnaire was identical to the electronic questionnaire, except for the first question, which asked participants if anything had occurred after completing the first questionnaire that might affect how they answered the repeated questions. An example of having teeth taken out was given. Those who thought something had changed were asked to return the questionnaire in the prepaid envelope without answering the remaining questions. Only data from participants who ticked the box indicating that nothing had changed were used to assess test-retest reliability.

Recruitment started in June 2019 and finished in March 2020, when the first wave of the coronavirus pandemic stopped all clinical activity in the department for 6 weeks.

The responses submitted through the Google Form were saved as an Excel spreadsheet (Microsoft Corp, Redmond, Wash). Each item was scored on a scale of 0-4. The scores of the 4 positively worded items were reversed to maintain the consistency that higher scores reflect more negative impacts. The scores for each item were added together to produce an overall total score. Responses with >25% missing items were excluded from the analysis. When less than 25% of items were missing, each missing item was replaced with the mean for the participant (individual mean technique).[13]

Sample sizes of 150-200 participants are considered adequate to undertake a Rasch analysis.[14] The aim was to include over 200 participants, but unfortunately, the first wave of the coronavirus pandemic in the United Kingdom stopped all clinical activity, and a decision was made to stop recruitment at this point.

## Statistical analysis

Descriptive analyses were undertaken using Excel. Structural validity was tested using an IRT Rasch model using specialized software (RUMM2030; RUMM Laboratory Pty Ltd, Duncraig, Wash). A reflective model was used because it was assumed those questionnaire items represented the observable effects of an orthodontic appliance on the individual's function, discomfort, and social and emotional well-being that can be assessed and are a manifestation of the underlying construct, namely, the effect on the oral health of the respondent.

The followings were used to validate the response format and items and assess differential item functioning test as recommended by Tennant and Conaghan:[15]

1. Category discrimination: the borders between 2 adjacent response categories were assessed to determine if reducing the number of response categories was appropriate.
2. Local item dependence was assessed to establish whether items were independent of each other. A score of 0.2 for residual correlations compared with the average residual correlation means was used to denote a significant correlation between specific items.[14]
3. Differential item functioning was assessed by age and gender.
4. Item fit to the Rasch model: Statistical significance was assessed using a chi-square test with Bonferroni adjustment. To determine that the final chosen items and persons did not deviate considerably from the Rasch model, the result should be nonsignificant, with a range of ± 2.5 and a mean statistical score of 0 ± 1.[16] An independent *t* test was used to assess the unidimensionality of the measure. Unidimensionality was confirmed when <5% of the *t* tests were significant at $P < 0.05$. If >5% of the *t* tests were significant, then the lower bound of the 95% confidence interval (CI) must be <0.05 to show some evidence of unidimensionality.[17]

**Table I.** Fit to the Rasch model

| | *Item residual* | *Person residual* | *Chi-square* | | *Reliability* | *Unidimensionality* | |
|---|---|---|---|---|---|---|---|
| *Analysis name* | *Mean ± SD* | *Mean ± SD* | *Value (df)* | *P value* | *PSI* | *Proportion of tests >5%* | *Lower 95% CI proportion* |
| Initial analysis | −0.01 ± 1.91 | −0.22 ± 1.11 | 342 (62) | <0.001 | 0.71 | 12.2% | 0.09 |
| Rescored to 3-point scale | −0.11 ± 1.15 | −0.27 ± 1.08 | 96 (62) | 0.004 | 0.85 | 19.9% | 0.17 |
| Remove misfitting/highly correlated items/DIF | −0.21 ± 0.63 | −0.29 ± 0.95 | 36 (42) | 0.73 | 0.81 | 8.29% | 0.05 |
| Ideal | 0 ± 1 | 0 ± 1 | | >0.0005[†] | >0.7 | <5% | ≤0.05 |

*df*, degrees of freedom; *DIF*, differential item functioning.
[†]Bonferroni adjusted for 21 items.

5. Reliability: reliability was assessed by the person separation index (PSI). The PSI is similar to Cronbach α but uses logit value, whereas Cronbach α uses the raw score. The recommended value for the PSI was >0.7.[17,18]

After achieving a unidimensional scale, all further analyses were undertaken by transforming raw scores into interval data.

After the Rasch analysis and item reduction, the scores for the remaining items, excluding the global question, were added to calculate a total OTIQ score. The construct validity of the modified questionnaire was assessed by calculating Spearman's rank correlation between the total OTIQ scores and the global question responses. Internal consistency and reliability were examined using Cronbach α and test-retest reliability using a 2-way mixed effect intraclass correlation coefficient.[19] We determined it was not possible to test criterion validity as there are no similar measures to compare the new measure.

## RESULTS

The questionnaire developed in stage 1 and initially evaluated by the focus groups in stage 2 study 1 can be seen in Supplementary Material 1. The wording, layout, and response format of the questionnaire were tested with 2 focus groups of adolescents and a 1-to-1 interview (n = 7; 4 female, 3 male). All 7 wore maxillary and mandibular fixed appliances (or had just had a fixed appliance removed before the focus group). All participants provided informed consent. The focus groups and interviews were audio-recorded, conducted, transcribed, and analyzed by an experienced qualitative researcher (J.E.K.). After discussion among the research team, minor modifications were undertaken, and the revised questionnaire was tested in a cross-sectional convenience sample in study 2.

A total of 181 participants completed the questionnaire in study 2 (117 females, 65%; 64 males, 35%). Because of the use of electronic data collection, there were no missing data. The mean age was 14.7 ± 1.5 years (range, 11-17 years). The majority (n = 138, 76%) were wearing maxillary and mandibular fixed appliances only, and the remainder were wearing fixed appliances with auxiliary anchorage, such as transpalatal or lingual arch (n = 13, 7%), a removable functional appliance, mostly Twin-blocks (n = 15, 8%), a removable retainer, mostly thermoplastic (n = 5, 3%), various other appliances (n = 6, 4%; removable biteplane: n = 2; RME: n = 2; quad helix: n = 1; facemask: n = 1) and in 4 participants the appliance was not recorded. The mean length of time the participant wore the appliances was 12.4 ± 8.6 months (range, from 1 week to 37 months). According to the IMD, the socioeconomic status of the participants was evenly spread between the lowest quartile (n = 51, 28%), the highest quartile (n = 56, 31%), and the 2 middle quartiles (<50: n = 34, 19%; ≥50%: n = 40, 22%).

The initial scale was relatively unidimensional and showed a misfit to the model, but not by strict standards as detailed in the methods. (Table I). Ten of the original 31 items had disordered thresholds, indicating that the response categories were not functioning as expected. For example, the analysis demonstrated that participants could not adequately distinguish between similarly worded response options. Changing the 5-point scale to a 3-point scale by combining the second and third categories (a little bit and a bit) and the fourth and fifth categories (quite a bit and a lot) resulted in the ordered response categories. Even combining the response options in this way resulted in 1 item (forgetting to wear my brace or retainer) still failing to show an ordered response; therefore, this question was removed. Two items (my brace or retainer breaking and my brace or retainer interfering with sports, hobbies, or pastimes [eg, musical instruments]) demonstrated differential item functioning by gender and were also excluded from further analysis. Seven items (the smell of my brace

**Table II.** Item fit statistics ordered by item location

| *Item* | *Location* | *Standard error* | *Fit residual* | $\chi^2$ |
|---|---|---|---|---|
| I feel attractive | −2.957 | 0.136 | −0.346 | 2.883 |
| Food getting stuck | −2.268 | 0.146 | −0.601 | 0.722 |
| Catching the inside of mouth | −1.478 | 0.127 | 0.818 | 1.437 |
| Difficulty eating certain foods | −0.887 | 0.137 | 0.244 | 1.054 |
| Rubbing my gums | −0.726 | 0.133 | −0.948 | 0.421 |
| Feeling tight | −0.588 | 0.149 | −0.377 | 0.844 |
| Worrying about brace breaking | −0.473 | 0.145 | 0.762 | 2.108 |
| I feel negative about my smile | −0.066 | 0.134 | −0.536 | 2.392 |
| Difficulty cleaning my brace | 0.004 | 0.148 | 0.64 | 1.014 |
| Difficulty pronouncing words | 0.179 | 0.141 | 0.149 | 0.488 |
| I feel normal | 0.183 | 0.144 | 0.016 | 2.443 |
| Appearance of brace | 0.293 | 0.145 | 0.335 | 4.405 |
| Making jaw ache | 0.412 | 0.149 | −0.172 | 0.485 |
| Having photograph taken | 0.55 | 0.154 | −0.959 | 2.008 |
| Difficulty chewing or swallowing | 0.743 | 0.157 | −0.385 | 1.309 |
| Difficulty sleeping | 1.01 | 0.173 | 0.604 | 0.399 |
| I feel annoyed | 1.016 | 0.174 | −0.864 | 3.6 |
| I feel ugly | 1.065 | 0.188 | −1.468 | 3.104 |
| Being teased about brace | 1.264 | 0.204 | −0.093 | 0.493 |
| I feel weird | 1.311 | 0.189 | −0.817 | 3.84 |
| I feel shy | 1.412 | 0.203 | −0.367 | 0.668 |
| Ideal | | | $< \pm 2.5$ | >0.0005[†] |

[†]Bonferroni adjusted for 21 items.

or retainer, I feel embarrassed, talking in public [eg, answering questions at school], making my teeth ache, making my mouth sore, I feel confident, and I feel positive about my smile) showed high residual correlation with other items and they were also removed. Removal of these 10 items improved the fit of the model. There remained some residual correlations greater than 0.2. These were paired items in which some correlation might be expected (rubbing on my gums, making my jaw ache, I feel ugly, and I feel negative about my smile). Removal of these items did not improve the fit; therefore, they were retained. None of the participants demonstrated misfit to the model, indicating that participants were completing the measure as expected. The ideal statistics and the finding of each stage of the item reduction process are presented in Table II.

The item fit statistics for the 21 retained items are shown in Table III, which are ordered from easiest (I feel attractive) to most difficult (I feel shy). The mean person location (degree of impact) is 1.45 when the items are centered on 0. This indicates that the measure is targeted at participants with slightly fewer impacts than those in this study. This is not unexpected as these participants were attending a hospital orthodontic department and may be expected to have more impacts than those suitable for treatment in primary care.

The Figure shows the person-item threshold map, indicating that participants (top) are distributed in a similar pattern to the items, demonstrating that the items (*bottom*) measure the impacts of wearing orthodontic appliances on oral health–related quality of life along with the construct from least to most. The level of impact experienced by most participants aligns with the items' difficulty. As the items fit the Rasch model, a

**Table III.** Raw score to interval score conversion

| *Raw score* | *Interval scale score* | *Raw score* | *Interval scale score* |
|---|---|---|---|
| 0 | 0 | 22 | 24 |
| 1 | 4 | 23 | 24 |
| 2 | 6 | 24 | 25 |
| 3 | 8 | 25 | 25 |
| 4 | 9 | 26 | 26 |
| 5 | 11 | 27 | 26 |
| 6 | 12 | 28 | 27 |
| 7 | 13 | 29 | 27 |
| 8 | 14 | 30 | 28 |
| 9 | 15 | 31 | 28 |
| 10 | 16 | 32 | 29 |
| 11 | 17 | 33 | 30 |
| 12 | 17 | 34 | 30 |
| 13 | 18 | 35 | 31 |
| 14 | 19 | 36 | 32 |
| 15 | 19 | 37 | 33 |
| 16 | 20 | 38 | 34 |
| 17 | 21 | 39 | 35 |
| 18 | 21 | 40 | 36 |
| 19 | 22 | 41 | 39 |
| 20 | 22 | 42 | 42 |
| 21 | 23 | | |

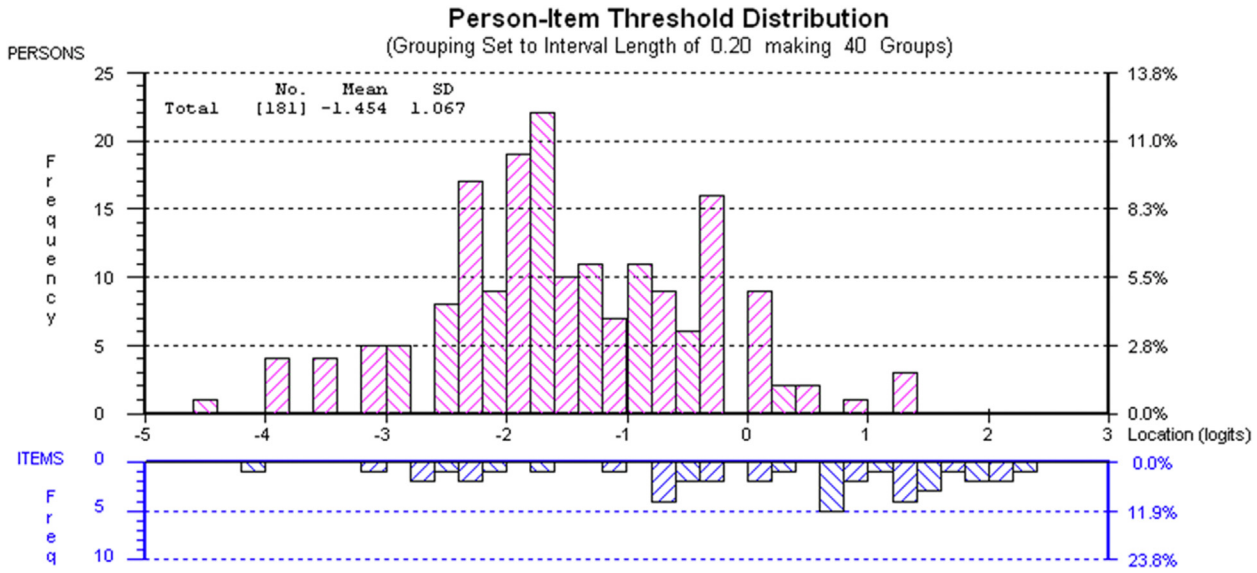


**Fig.** Targeting of the current measure. The *top* shows the distribution of participants, and the *bottom* shows the distributions of thresholds (category transitions) of the items. The x-axis displays the location (severity of impact) of participants and the item location (difficulty) of the item thresholds. The y-axis shows the frequency of item thresholds and the number of participants.

**Table IV.** The global scores before and after combining the categories' responses

| *Before* | *n* | *%* | *After* | *n* | *%* |
|---|---|---|---|---|---|
| Not at all | 69 | 38.1 | Not at all | 69 | 38.1 |
| A little bit | 70 | 38.7 | A little bit/a bit | 98 | 54.1 |
| A bit | 28 | 15.5 | Quite a bit/a lot | 14 | 7.8 |
| Quite a bit | 5 | 2.7 | | | |
| A lot | 9 | 5.0 | | | |
| Total | 181 | 100 | Total | 181 | 100 |

transformation from the raw score to interval scaling is shown in Table III.

The modified questionnaire with 21 items and a 3-point response format is shown in Supplementary Material 2. The mean total OTIQ score was 11.0 ± 5.5 (range, 1 to 29), and there were no floor (minimum = 0) or ceiling effects (maximum = 42).

The frequency distribution for the global scores is shown in Table IV. The correlation between the total OTIQ scores and the global question "How does your brace or retainer affect your life overall?" was moderate (rho = 0.480), indicating acceptable construct validity. According to the global question, only about 8% of the participants reported that their brace affected their overall life a lot, whereas 38% reported that their overall life was not affected at all.

The Cronbach α score was 0.827, indicating that the internal consistency and reliability of the modified measure were high. Of the 181 participants, 129 agreed to complete the questionnaire for a second time; however, only 41 participants (31.8% of the sample) returned their completed second questionnaire. One participant ticked the box claiming a change in their oral and, or orthodontic circumstances; hence, this participant was excluded from the reliability testing. There were no missing items in the remaining 40 returned questionnaires. The interclass correlation coefficient result showed that the test-retest reliability was high, giving a score of 0.85 (95% CI, 0.73-0.92).

Many participants provided free text comments in addition to responses to the closed questions. The most common foods avoided were apples, followed by chewy, hard, or sweet foods. Interestingly several mentioned avoiding chewing gum, even though there is no evidence that this leads to increased breakages of appliances and may reduce the impact of fixed appliances.[6]

Comments about the bad aspects of braces caused pain and discomfort, particularly after adjustment, rubbing of the cheeks and lips, and the need to avoid certain foods. Many thought a good aspect of braces would be the final result of having straight teeth, which concurs with the findings of a recent longitudinal qualitative study.[20]

Most participants (n = 116, 64%) reported not taking any analgesia since the last visit to their orthodontist. The mean total OTIQ score was slightly higher for those

who had taken analgesia (12.9 ± 5.5) compared with those who had not taken analgesia (10.0 ± 5.0), which was statistically significant (95% CI, 1.2-4.7; independent $t$ test, $P = 0.001$). A higher proportion of those who had not taken analgesia reported that the brace did not affect their life (n = 48, 41.4%) compared with those who had taken analgesia (n = 16, 30.8%).

## DISCUSSION

This report is an initial evaluation of a new instrument for assessing the self-reported impacts of adolescents wearing an orthodontic appliance. The results suggest the measure is valid and reliable and can be used as a patient-reported outcome in clinical studies involving orthodontic treatments; however, further testing in different sites and populations would be beneficial to confirm the properties of the measure.

Tsichlaki et al[1] have identified a set of core outcomes that should be collected as part of prospective clinical trials of routine noncleft, nonsurgical orthodontic treatments.[1] The final orthodontic core outcome set includes 7 outcomes categorized into 4 domains, and this measure will provide an important patient perspective as an outcome in the domain entitled delivery of care.

The initial items in the questionnaire were identified through 1-to-1 interviews with patients in the appropriate age groups wearing a range of orthodontic appliances, both removable and fixed.[9] The relevance, completeness, and clarity of the questions; instructions; and response options were then checked with 2 focus groups and a further 1-to-1 interview. The content and face validity were good. The rationale for developing a questionnaire for use with a range of orthodontic appliances is that this will enable researchers to use OTIQ as an outcome measure to evaluate and compare the impacts of any type of orthodontic intervention.

The target age for a questionnaire is an important consideration during the development of an instrument. This adolescent group, aged 11-17 years, was chosen as a common age range when orthodontic treatment is experienced. In addition, these adolescents are generally considered to have similar cognitive development, ability, and function compared with adults or younger children, and researchers can develop questionnaires to specifically cover these years.[21]

We used a Rasch model for item reduction and validity testing. Based on IRT, the Rasch model is increasingly being used to develop and validate[22,23] It is claimed that a Rasch analysis increases the precision and quality of a measure by enabling the use of a respondent's raw test or scale scores to be expressed on a linear scale.[24] A proposed advantage of IRT is that it is superior to classical test theory in the assessment of individual change, and there is evidence for this in questionnaires with at least 20 items.[25]

The Rasch analysis suggested that our initial measure showed some misfit to the model, which was not unidimensional. The benefit of a unidimensional instrument is that it suggests all the items are measuring the same underlying construct, which makes it easier to explain and interpret individual differences.[26] Accordingly, we decided to eliminate 10 items and shrink the response options to 3 instead of 5, and the fit of items to the model unidimensionality improved.

The criterion validity of the measure was not assessed because of the lack of comparable validated measures designed to assess the impacts of all types of orthodontic appliances. The correlation between the total OTIQ scores and the global scores is similar to the Child Perceptions Questionnaire, suggesting that the measure has acceptable construct validity[27]; however, the correlation is not as high as the Malocclusion Impact Questionnaire.[23] We observed that, whereas 69% of participants with malocclusion were somewhat, quite a bit, or very much bothered about their teeth when the Malocclusion Impact Questionnaire was validated, only 23% of participants in this study responded that their brace affected them a bit, quite a bit, or a lot. One explanation might be that the mean length of time the participants had been wearing their appliance was 12 months. Longstaff et al[20] showed that adolescents quickly adapt to their appliances, which soon become a new normal.[20] This finding also emerged in the 1-to-1 interviews conducted for stage 1 and focus groups conducted for stage 2, study 1. It would be useful to test the questionnaire early in treatment and examine how impacts change longitudinally with time. Another possible explanation is the prevalence of a particular appliance type within a peer group. For example, the interviews and focus groups indicated that the impact of an appliance might be lessened if the appliance is perceived as common among a year group at school.

The internal consistency and test-retest reliability of the current questionnaire was good and consistent with other questionnaires measuring oral health–related quality of life[23,27,28] or orthodontic pain.[29] These properties of the current questionnaire were better than that of the Impact of Fixed Appliances on Daily Life questionnaire.[7]

The initial questionnaire was developed using a combination of negatively and positively worded items, which is considered valuable in developing patient-reported measures, particularly for assessing the quality of life.[30] Four positively worded items were included in the initial measure; however, 2 of these were removed after the Rasch analysis and item reduction (I feel confident and I feel positive about my smile). It is important

to remember to reverse the score for the 2 remaining positive items (I feel normal and I feel attractive).

The questionnaire was administered through an electronic device, which has several advantages over paper administration. The data can be downloaded directly into a spreadsheet for analysis, reducing the risk of transcription errors. Responses can be made compulsory for progression through the questionnaire, reducing the potential for missing items. In addition, adolescents are very used to using electronic devices for various activities, and completing the questionnaire could be more interactive and fun, increasing response levels. Bjorner et al[31,32] found that the method of administration did not influence the measurement characteristics of various items chosen from the Patient-Reported Outcomes Measurement Information System.

There was a higher proportion of female than male participants in this study. This gender imbalance in those seeking and undergoing orthodontic treatment concurs with previous studies.[33,34] Longstaff[35] noted that gender influenced both the motivation for and experience of undergoing orthodontic treatment, whereas female participants spoke freely about the extent, duration, and impact of the pain and discomfort, as well as their use of analgesia, because of their brace, many of the boys downplayed these aspects. Longstaff[35] explained that Western culture encourages stoicism in males, and her position as a female interviewer might have inflated these brave narratives.

The cross-sectional validation was carried out in 1 hospital (Charles Clifford Dental Hospital), which will affect the generalizability of the measure, and it is important to test the measure further in different settings to assess the external validity. The number of returned questionnaires was also relatively low, similar to other studies.[7,23] Only 1 repeat questionnaire was excluded as the respondent indicated an event had occurred after completing the first questionnaire, which might affect how they answered the repeated questions.

In addition to recruiting participants wearing different orthodontic appliances, we also recruited patients at different stages of their treatment. This was to ensure that the full range of impacts from wearing an orthodontic appliance was captured. As noted previously, the mean length of time the participants had been wearing their appliance was 12 months, whereas previous work suggests that pain and discomfort worsen in the first 2 weeks.[36] The responsiveness of the measure to change over time needs assessing.

Information about the ethnicity of the participants was not collected, but the majority were White and British, as identified in a previous study.[23] Potential participants who did not understand English or could not complete the questionnaire without considerable assistance were excluded. This is because we felt that translating or explaining large parts of the questionnaire might affect the interpretation of the questions. The questionnaire should be tested in further diverse samples to test cross-cultural validity.

## CONCLUSIONS

The initial OTIQ was tested and modified using a Rasch analysis. The modified questionnaire demonstrated good construct validity, reliability, and internal consistency. Further testing to assess generalisability to other settings and longitudinal responsiveness is required. If the properties of the questionnaire are confirmed, then the measure can be used as an assessment tool in the delivery of care domain of the orthodontic core outcome set.

## AUTHOR CREDIT STATEMENT

Philip E. Benson contributed to conceptualization, methodology, validation, formal analysis, investigation, resources, data curation, original draft manuscript, visualization, supervision, project administration, and funding acquisition; Ebrahim Alshawy contributed to software, formal analysis, investigation, resources, data curation, original draft preparation, manuscript review and editing, and project administration; Jennifer Kettle contributed to conceptualization, methodology, validation, formal analysis, investigation, resources, data curation, and project administration; and Fiona Gilchrist contributed to formal analysis and manuscript review and editing.

## SUPPLEMENTARY DATA

Supplementary data associated with this article can be found, in the online version, at http://doi.org/10.1016/j.ajodo.2022.06.018.

## REFERENCES

1. Tsichlaki A, O'Brien K, Benson PE, Marshman Z, Johal A, Colonio-Salazar FB, et al. Development of a core outcome set for use in routine orthodontic clinical trials. Am J Orthod Dentofacial Orthop 2020;158:650-60.
2. Liu Z, McGrath C, Hägg U. Changes in oral health-related quality of life during fixed orthodontic appliance therapy: an 18-month prospective longitudinal study. Am J Orthod Dentofacial Orthop 2011;139:214-9.
3. O'Brien K, Wright J, Conboy F, Sanjie Y, Mandall N, Chadwick S, et al. Effectiveness of treatment for Class II malocclusion with the Herbst or Twin-block appliances: a randomized, controlled trial. Am J Orthod Dentofacial Orthop 2003;124:128-37.
4. Feldmann I, List T, John MT, Bondemark L. Reliability of a questionnaire assessing experiences of adolescents in orthodontic treatment. Angle Orthod 2007;77:311-7.

5. Yassir YA, McIntyre GT, Bearn DR. Three questionnaires to assess the perception of fixed orthodontic therapy before, during and after treatment: validity and reliability. Eur J Orthod 2017;39:402-10.
6. Bradley E, Shelton A, Hodge T, Morris D, Bekker H, Fletcher S, et al. Patient-reported experience and outcomes from orthodontic treatment. J Orthod 2020;47:107-15.
7. Mandall NA, Vine S, Hulland R, Worthington HV. The impact of fixed orthodontic appliances on daily life. Community Dent Health 2006;23:69-74.
8. Benson PE, Razi RM, Al-Bloushi RJ. The effect of chewing gum on the impact, pain and breakages associated with fixed orthodontic appliances: a randomized clinical trial. Orthod Craniofac Res 2012;15:178-87.
9. Kettle JE, Hyde AC, Frawley T, Granger C, Longstaff SJ, Benson PE. Managing orthodontic appliances in everyday life: A qualitative study of young people's experiences with removable functional appliances, fixed appliances and retainers. J Orthod 2020;47:47-54.
10. Federation FWD. FDI's Definition of Oral. Available at: https://www.fdiworlddental.org/fdis-definition-oral-health. Accessed December 8, 2021.
11. Guyatt GH, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. CMAJ 1986;134:889-95.
12. Ministry of Housing, Communities & Local Government. National statistics: English indices of deprivation. Available at: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019. Accessed August 9, 2022.
13. Shrive FM, Stuart H, Quan H, Ghali WA. Dealing with missing data in a multi-question depression scale: a comparison of imputation methods. BMC Med Res Methodol 2006;6:57.
14. Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, de Vet HCW, et al. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. 2019 [cited 2021 3-8-2021]. Available at: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf. Accessed March 8, 2021.
15. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007;57:1358-62.
16. Kersten P, White PJ, Tennant A. Is the pain visual analogue scale linear and responsive to change? An exploration using Rasch analysis. PLoS One 2014;9:e99485.
17. Vincent JI, MacDermid JC, King GJ, Grewal R. Rasch analysis of the Patient Rated Elbow Evaluation questionnaire. Health Qual Life Outcomes 2015;13:84.
18. Gilchrist F, Rodd HD, Deery C, Marshman Z. Development and evaluation of CARIES-QC: a caries-specific measure of quality of life for children. BMC Oral Health 2018;18:202.
19. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155-63.
20. Longstaff S, Davies K, Benson P. Exploring 10-15-year-old patients' perspectives of fixed orthodontic treatment. J Orthod 2021;48:110-7.
21. Omrani A, Wakefield-Scurr J, Smith J, Brown N. Survey development for adolescents aged 11-16 years: A developmental science based guide. Adolesc Res Rev 2019;4:329-40.
22. Shelton AT, Houghton NY, Morris DO, Latchford GL, Bekker HL, Munyombwe T. The development and validation of a psychological questionnaire for patients undergoing orthognathic treatment. Orthod Craniofac Res 2015;18:51-64.
23. Benson PE, Cunningham SJ, Shah N, Gilchrist F, Baker SR, Hodges SJ, et al. Development of the Malocclusion Impact Questionnaire (MIQ) to measure the oral health-related quality of life of young people with malocclusion: part 2 - Cross-sectional validation. J Orthod 2016;43:14-23.
24. Boone WJ. Rasch analysis for instrument development: Why, When, and How? CBE Life Sci Educ 2016;15:rm4.
25. Jabrayilov R, Emons WHM, Sijtsma K. Comparison of classical test theory and item response theory in individual change assessment. Appl Psychol Meas 2016;40:559-72.
26. Hattie J. An empirical study of various indices for determining unidimensionality. Multivariate Behav Res 1984;19:49-78.
27. Jokovic A, Locker D, Stephens M, Kenny D, Tompson B, Guyatt G. Validity and reliability of a questionnaire for measuring child oral-health-related quality of life. J Dent Res 2002;81:459-63.
28. Jokovic A, Locker D, Guyatt G. Short forms of the Child Perceptions Questionnaire for 11-14-year-old children (CPQ11-14): development and initial evaluation. Health Qual Life Outcomes 2006;4:4.
29. Iwasaki LR, Freytag LE, Schumacher CA, Walker MP, Williams KB. Validation of a modified McGill Pain Questionnaire for orthodontic patients. Angle Orthod 2013;83:906-12.
30. Broder HL, McGrath C, Cisneros GJ. Questionnaire development: face validity and item impact testing of the Child Oral Health Impact Profile. Community Dent Oral Epidemiol 2007;35(Suppl 1):8-19.
31. Bjorner JB, Rose M, Gandek B, Stone AA, Junghaenel DU, Ware JE Jr. Method of administration of PROMIS scales did not significantly impact score level, reliability, or validity. J Clin Epidemiol 2014;67:108-13.
32. Bjorner JB, Rose M, Gandek B, Stone AA, Junghaenel DU, Ware JE Jr. Difference in method of administration did not significantly impact item response: an IRT-based analysis from the Patient-Reported Outcomes Measurement Information System (PROMIS) initiative. Qual Life Res 2014;23:217-27.
33. Bernabé E, Sheiham A, de Oliveira CM. Impacts on daily performances related to wearing orthodontic appliances. Angle Orthod 2008;78:482-6.
34. Chen M, Wang DW, Wu LP. Fixed orthodontic appliance therapy and its impact on oral health-related quality of life in Chinese patients. Angle Orthod 2010;80:49-53.
35. Longstaff SJ. Exploring young people's perspectives of fixed orthodontic treatment. Sheffield: University of Sheffield; 2017.
36. Erdinç AM, Dinçer B. Perception of pain during orthodontic treatment with fixed appliances. Eur J Orthod 2004;26:79-85.