Check for updates

Received: 5 January 2024 | Revised: 20 June 2024 | Accepted: 3 July 2024

DOI: 10.1111/jocd.16465

JCD Journal of Cosmetic Dermatology WILEY

ORIGINAL ARTICLE

# Development and validation of a patient-reported outcome measure for hair loss treatment: The HAIR-Q

Anne F. Klassen DPhil[1] | Jasmine Mansouri BSc[2] | Manraj Kaur PhD[3] | Charlene Rae PhD[1] | Lotte Poulsen MD, PhD[4,5] | Steven Dayan MD[6] | Stefan J. Cano PhD CPsychol AFBPsS[7] | Andrea L. Pusic MD[3]

[1]Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada

[2]Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

[3]Department of Surgery, Patient Reported Outcome, Value, and Experience (PROVE) Center, Brigham and Women's Hospital, Boston, Massachusetts, USA

[4]Research Unit for Plastic Surgery, Odense University Hospital, Odense, Denmark

[5]Løntoft, Nyhøj and Poulsen Plastic Surgery, Odense, Denmark

[6]Dayan Facial Plastic Surgery, Chicago, Illinois, USA

[7]Modus Outcomes (A Division of Thread), Cheltenham, UK

**Correspondence**
Anne F. Klassen, Department of Pediatrics, McMaster University, 3N-27, 1280 Main Street W, Hamilton, ON L8S 4L8, Canada.
Email: aklass@mcmaster.ca

## Abstract

**Background:** Patient-reported outcome measures (PROMs) for hair loss focus mainly on Alopecia Areata. We created a PROM (i.e., HAIR-Q) that is applicable to any hair loss condition. The HAIR-Q measures satisfaction with hair.

**Patients/Methods:** Concept elicitation interviews were conducted and analyzed to develop a draft scale. Content validity was established through multiple rounds of patient and expert input. Psychometric properties of the scale were examined in an online sample (i.e., Prolific) using Rasch measurement theory (RMT) analysis. Test–retest reliability and tests of construct validation were examined.

**Results:** Content validity of a 22-item draft scale was established with input from 11 patients, 12 experts and an online Prolific sample of 59 people who had a variety of hair loss treatments. In the RMT analysis ($n = 390$), 8 items were dropped. Data for the 14-item scale fit the Rasch model ($\chi^2 = 89.85$, df = 70, $p = 0.06$). All 14 items had ordered thresholds and good item fit. Reliability was high with person separation index and Cronbach alpha values ≥0.91, and intraclass correlation coefficient of 0.94 based on a sample of 97 participants. Higher (better) scores on the scale were associated with having more hair, looking younger than ones' age, satisfaction with hair overall, being less bothered by hair loss, and for those who had a hair loss treatment in the past year, being more satisfied with their hair now than before treatment ($p < 0.001$).

**Conclusion:** The HAIR-Q evidenced reliability and validity and can be used in research and to inform clinical care to measure satisfaction with hair from the patient perspective.

KEYWORDS
aesthetic medicine, hair loss, patient-reported outcome, psychometric, Rasch analysis

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Author(s). *Journal of Cosmetic Dermatology* published by Wiley Periodicals LLC.

## 1 | INTRODUCTION

Hair loss is a common dermatological condition affecting various age groups and genders.[1] According to the American Academy of Dermatology Association, common causes of hair loss are genetics, aging, Alopecia Areata, stressors, hair care, and hormonal imbalance.[2] Scalp hair plays an important role in body image.[3] The psychological impact of hair loss on both men and women has been well documented.[4-14] Quality of life has been shown to improve after treatment for hair loss.[15,16]

Patient-reported outcome measures (PROMs) are tools that can be used to measure outcomes that are important to patients.[18] To ensure the usefulness of PROMs, a rigorous development and validation process must be undertaken within the target population.[19,20] The quality of PROM design and the strength of psychometric properties can be assessed using COnsensus-based *Standards* for the selection of health Measurement Instruments (COSMIN) guidelines and checklists.[21,22] Application of PROMs that are incorrectly designed or not properly validated can lead to inconsistency of measurement results, or measurement of an unintended construct, impacting interpretation of study results or clinical decision-making.[23,24]

Several PROMs have been developed for hair loss, with most focused on Alopecia Areata (AA)[25] including: the Scalp Hair Assessment PRO,[26] the Alopecia Areata Symptom Impact Scale,[27] and the AA-QLI.[28] These PROMs address symptoms, severity of hair loss, and impact of AA on daily life, rather than a patient's perspective on the appearance of their hair. Other hair loss PROMS such as The Hair Specific Skindex-29 (a modified version of Skindex-29[29]) measures impact of hair loss on quality life for use in androgenetic alopecia.[30] Another PROM, the hair growth questionnaire, was developed for use in men with hair loss to measure patient-perceived changes in hair growth and appearance.[31] A new PROM is needed because none of these hair-specific PROMs were designed using a modern psychometric approach, nor do they comprehensively measure satisfaction with hair from the patient perspective. A modern psychometric approach that uses item response theory has certain advantages over traditional methods, most notably, the scales produce interval-level measurement properties, which allows for the application of parametric statistics.[32] Also, modern psychometrics can be used to determine if the PROM works the same across important patient characteristics, such as age, gender and condition. Since most current hair-specific PROMs focus on a single hair condition, developing a PROM that is generic would make it possible to measure effectiveness of treatments across different hair conditions.

The specific aims of our study were as follows: (1) to elicit hair-related concepts important to patients with hair loss of any etiology; (2) to use the concepts to create a PROM to measure satisfaction with hair appearance; and (3) to assess the psychometric properties of the new PROM.

## 2 | METHODS

This study was part of a larger mixed methods study to develop PROMs for aesthetic treatments.[33,34] Prior to study initiation, ethics board approval was obtained from the Hamilton Integrated Ethics Board (Canada) (#13603) at McMaster University in Canada. International guidelines for PROM development were followed.[19,35] The mixed methods approach began with a qualitative study that used interpretive description.[36] Concept elicitation interviews were conducted between October 2021 and March 2022. Participants were recruited from six private practice clinics in Canada and the USA. Staff recruited adult patients who varied by age, gender, race and minimally invasive treatment. Interviews took place over an institutionally licensed Zoom platform (Zoom Video Communications, Inc., San Jose, CA, 2003) with an experienced interviewer who followed an interview guide. Hair loss concepts were elicited with questions and probes that asked participants to describe how areas targeted by treatment looked before and after treatment, aspects of how they looked that they were most/least satisfied with, and that were most/least important to them, and how their appearance affected how they felt and interacted with people.

Interviews were audio-recorded and transcribed. Coding involved two coders working independently through each transcript to label each concept with a domain and major/minor theme. Any discrepancies in codes were settled by consensus. Codes were transferred to Microsoft Excel 365 and refined through constant comparison.[37] Interviews continued until saturation for the majority of concepts was reached.[38] Participants were sent an e-gift card equivalent to $100 as a thank you for their time.

### 2.1 | Concept elicitation and scale refinement

This paper is focused on the concepts related to hair appearance. We created a single scale to measure satisfaction with hair that was refined and validated through several steps. In October 2022, to determine content validity for the scale, participants from the qualitative study were invited to provide feedback in a REDCap survey.[39] Specifically, participants were asked to choose one answer for each item in the draft scale: (1) I do not understand the question; (2) I understand the question, but it could be worded better, (3) I understand the question, but it is NOT relevant to me, and (4) I understand the question and it is relevant to me. A comment box was provided to describe any missing items from their perspective. We provided a gift card of $30 to thank participants for their time. Items identified as problematic were dropped or reworded and open-text comments reviewed.

For a subset of participants, an experienced interviewer conducted cognitive debriefing interviews over the Zoom platform. Participants reviewed the scale's instructions, items and response options, and suggested missing items. The interviews were audio-recorded, transcribed, and analyzed. We gave participants $70 as a thank you for their time. Experts in aesthetics and representatives from the aesthetics industry were invited to review the scale and indicate any items they thought would be not relevant to patients, and to suggest missing concepts. Feedback from patients and experts was used to refine the scale.

To establish content validity for hair loss conditions, we used an online crowd working platform, that is, Prolific (www.prolific.co). A screening survey was conducted in December 2022 to identify a

14732165, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jocd.16465, Wiley Online Library on [16/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

sample. On the day we screened, 121 170 Canadian and USA residents fluent in English formed the membership of Prolific. The rate of pay was the equivalent of 10.80 GBP per hour. Participants were asked if they have had one more of the following hair loss treatments: platelet rich plasma (PRP), medication (i.e., finasteride or minoxidil), surgery (i.e., scalp advancement, hair transplant), tattooing, steroid injections or creams, immunotherapy and/or light treatment. Any participant who reported that they had some or all their hair (i.e., not bald), and had one or more of the hair loss treatments in the past year was invited to complete a cognitive survey. Participants were asked to think of their most recent hair loss treatment when answering the following question for each item in the draft scale: (1) I do not understand the question; (2) I understand the question, but it is NOT relevant to me, and (3) I understand the question and it is relevant to me. A comment box was provided to describe missing concepts.

In February 2023 a pilot field was conducted by inviting the Prolific sample described above to complete the scale. Rasch measurement theory (RMT) analysis[40] was performed in RUMM2030 software[41] using the unrestricted Rasch model for polytomous ordered responses. Items with extreme misfit to the Rasch model were removed. Following the pilot survey, a screening survey was conducted to identify a sample of people having aesthetic treatments for the face or body as part of the larger study. The screening survey took place in February 2023. A total of 2500 people were screened. To validate the hair scale, participants were asked how much hair they currently have on their head (none, some, most, all), and if they had any of the eight treatments for hair loss listed above. We identified all participants who reported hair loss for potential inclusion in the sample. A sample was chosen and invited to complete the field-test survey in REDCap[39] in March 2023.

Data from the pilot- and field-test surveys were combined for the psychometric analysis. Table 1 shows the tests conducted to examine the item and scale psychometric performance. Person locations from the Rasch analysis were used to transform scale scores from

**TABLE 1** Psychometric tests performed.

| Test | Description |
|---|---|
| Thresholds for item responses | Item response options measuring satisfaction needed to be ordered on a continuum (e.g., a score of 1 lower than a score of 2). This approach is used to create a hierarchy of items to determine how items are ordered from easiest to hardest to endorse. |
| Item fit | The extent to which observed data fit expected values based on the Rasch model. Item fit was assessed by inspecting fit residuals and chi-square statistics. Fit residuals summarize the observed and expected responses to an item by the sample and should ideally lie within the range −2.5 and +2.5. Only items that had chi-square values that were non-significant after a Bonferroni adjustment for multiple testing were included. |
| Local dependency | The extent of local dependency among items. Residual correlations were examined to identify any greater than 0.20 above the average correlations, and a subtest analysis was performed to determine the impact of local dependency on scale reliability.[42] |
| Scale-to-sample targeting | The spread of person locations (i.e., satisfaction with hair in the sample) and item locations (i.e., range the set of items measured). A scale that is better targeted has more coverage and has the mean person location close to the center of the items.[43] We also computed the proportion of the sample that scored on scale. |
| Differential Item Functioning (DIF) | The extent to which items were invariant across age (i.e., ≤35 and ≥36), gender, treatment (yes, no), and hair loss (yes, no). This test uses analysis of variance to examine estimated person ability differences between class intervals within subgroups. When a significant DIF was identified, variables were split for the relevant items, and the original and split person locations were correlated to examine the impact of DIF on scale scoring.[44] |
| Reliability | 1. Person Separation Index—this statistic determined how well people in the sample were separated by the scale items.[45]<br>2. Cronbach alpha—this statistic was used to examine internal reliability.<br>3. Test–retest reliability—a subset of participants completed the survey twice separated by 7 days. We excluded anyone who: reported an important change in satisfaction with hair, completed the TRT outside of 7–14 days, had change scores that varied by more than 2.5 standard deviations despite reporting no change. Intraclass correlation coefficients (ICC) with a two-way random effects model was used to evaluate consistency. Reliability values should be >0.70.[21,46] |
| Construct validity | Rasch logit scores were transformed into 0 (least satisfied) to 100 (most satisfied). We examined the normality of the data by assessing kurtosis and skewness.[47] Nonparametric statistics were applied to data that exceeded ±2.0,[47] with statistical significance set at a two-tailed $p$-value of <0.05. The following tests of construct validity were performed:<br>1. Participants were asked: "How much hair do you CURRENTLY have on your head.?" Answers—A little (I have lost about 25% of my hair), Some (I have lost about 50% of my hair), Most (I have lost about 25% of my hair), All (I have a full head of hair). We hypothesized scores would be incrementally higher for people with more hair.<br>2. Participants were asked to report how satisfied they are with: "How your hair looks OVERALL?" We hypothesized that scores would be incrementally higher for those who are more satisfied with their hair overall.<br>3. Participants were asked: "Are you bothered by any HAIR LOSS?" Answers—Not bothered, A little bothered, Moderately bothered, Very bothered, Extremely bothered. We hypothesized that scores would be incrementally higher for those less bothered by hair loss. |

14732165, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jocd.16465, Wiley Online Library on [16/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 2** Characteristics for the 26 qualitive interview and Prolific survey participants.

| | Qualitative sample | Cognitive sample | Field test sample |
|---|---|---|---|
| | *N* = 26 | *N* = 59 (%) | *N* = 390 (%) |
| Country | | | |
| Canada | 6 | 4 (6.8) | 50 (13.0) |
| USA | 20 | 55 (93.2) | 335 (87.0) |
| Missing | 0 | 0 (0.0) | 5 (1.5) |
| Age (years) | | | |
| 20–29 | 3 | 31 (52.5) | 120 (30.7) |
| 30–39 | 6 | 16 (27.1) | 115 (29.5) |
| 40–49 | 7 | 11 (18.6) | 96 (24.6) |
| 50–59 | 6 | 1 (1.7) | 35 (9.0) |
| ≥60 | 4 | 0 (0) | 24 (6.2) |
| Gender | | | |
| Woman | 23 | 16 (27.1) | 172 (44.1) |
| Man | 3 | 43 (72.9) | 208 (53.3) |
| Gender diverse | 0 | 0 (0) | 9 (2.3) |
| Prefer to not answer | 0 | 0 (0) | 1 (0.3) |
| Race | | | |
| White | 22 | 31 (52.5) | 237 (60.8) |
| Black | 2 | 9 (15.3) | 36 (9.2) |
| Latin American | 0 | 1 (1.7) | 20 (5.1) |
| East Asian | 0 | 6 (10.2) | 27 (6.9) |
| Middle Eastern | 0 | 0 (0) | 3 (0.8) |
| South Asian | 1 | 5 (8.5) | 17 (4.4) |
| Southeast Asian | 1 | 3 (5.1) | 17 (4.4) |
| Other/Mixed race | 0 | 4 (6.8) | 9 (2.4) |
| Marital status | | | |
| Married/Common Law | 16 | 29 (49.2) | 174 (44.7) |
| Single | 7 | 29 (49.2) | 180 (46.2) |
| Divorced | 2 | 0 (0) | 24 (6.2) |
| Separated | 0 | 1 (1.7) | 3 (0.8) |
| Widowed | 1 | 0 (0) | 3 (0.8) |
| Other/Prefer not to answer | 0 | 0 (0) | 6 (1.6) |
| Fitzpatrick skin type | | | |
| Always burn and never tan | 2 | 5 (8.5) | 31 (7.9) |
| Usually burn and minimally tan | 9 | 7 (11.9) | 99 (25.4) |
| Mild burn and then tan | 9 | 17 (28.8) | 125 (32.1) |
| Rarely burn and always tan | 4 | 8 (13.6) | 73 (18.7) |
| Rarely burn and tan very easily | 1 | 15 (25.4) | 47 (12.1) |
| Never burn and never tan | 1 | 7 (11.9) | 15 (3.8) |
| Highest education | | | |
| Some high school | 0 | 1 (1.7) | 5 (1.3) |
| High school | 1 | 8 (13.6) | 33 (8.5) |
| Some college, trade or university | 4 | 13 (22.0) | 87 (22.3) |
| College, trade or university degree | 9 | 28 (47.5) | 181 (46.4) |
| Some Masters or Doctoral degree | 0 | 1 (1.7) | 19 (4.9) |
| Masters or Doctoral degree | 11 | 8 (13.6) | 64 (16.4) |
| Missing/Prefer to not answer | 1 | 0 (0) | 1 (0.3) |

0 (least satisfied) to 100 (most satisfied). The transformed scores were used in the test–retest (TRT) analysis and tests of construct validation using parametric or nonparametric tests depending on the distribution of the data.

## 3 | RESULTS

### 3.1 | Sample characteristics

Tables 2,3 show patient characteristics and treatment history for the qualitative and Prolific survey participants. Table 4 shows the reasons for hair loss for the field-test sample.

### 3.2 | Concept elicitation

Of the 26 participants that took part in interviews, participants had PRP treatment for hair loss. Coding and analysis identified 122 hair-specific codes of which 67 were appearance-related. Appearance concepts included the location of hair loss [So it's more affecting the top of my head.], the amount of hair loss [If I had not had it done, that 25–30 percent would not have been there.], shedding [I was shedding a lot of hair.], regrowth [I didn't realize that hair would grow back.], fullness [I know it's not completely full, but it's a big difference.], how thick or thin the hair was [It has definitely thickened it.], texture [My texture was OK.], hair health [It looks healthier.], styling concerns [...when I would kind of dye my hair. Or pull it up or try to make any hairstyle.], how the hair looked in photos [Like occasionally I would be playing with my kid and my husband would take a picture and I would look at the picture and think wow, it's thinning.], age concerns [I feel that I'm probably

14732165, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jocd.16465, Wiley Online Library on [16/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 3** Treatments history for the qualitive interviews and Prolific samples.

| | Qualitative | Cognitive sample | Field-test sample[a] |
|---|---|---|---|
| | *N* = 26 | *N* = 59 (%) | *N* = 337 (%) |
| Injectable | | | |
| Platelet Rich Plasma—injections into your scalp | 4 | 33 (55.9) | 13 (3.9) |
| Medication | | | |
| Medication—finasteride or minoxidil, the 2 main treatments for hereditary pattern baldness | 0 | 32 (54.2) | 120 (35.6) |
| Procedure | | | |
| Surgery—hair transplant | 0 | 5 (8.5) | 10 (3.0) |
| Surgery—scalp advancement | 0 | 3 (5.1) | 3 (0.9) |
| Tattooing—to make it look like short hair | 0 | 5 (8.5) | 4 (1.2) |
| Topical | | | |
| Steroids—injections or creams applied to bald patches | 0 | 11 (18.6) | 22 (6.5) |
| Light treatment—shining ultraviolet light on bald patches | 0 | 8 (13.6) | 33 (9.8) |
| Immunotherapy—chemicals applied to bald patches | 0 | 6 (10.2) | 2 (0.6) |

[a]This data was not collected for the pilot field-test sample.

looking about 5 years older than I am right now, still because of the lack of hair more than anything], and comparisons with other people [….I'm in my mid-thirties and most of my friends still have hair and I've always been very self-conscious that I don't have it.]. The appearance concepts were used to develop 22 items that measure satisfaction with hair. The goal was to include no more than 15 items in the final scale.

## 3.3 | Scale development and refinement

Appendix A shows the item-level decisions made in each round. In Round 1, 11 of the original 26 participants completed the 22 items survey providing 242 ratings. All items were easy to understand, with 0 ratings: "I do not understand"; 7 (2.9%) ratings: "I understand this question, but it could be worded better"; 86 (35.5%) ratings: "I understand this question, but it is not relevant to me"; and 149 (61.6%) ratings: "I understand this question and it is relevant to me."

Seven of the 11 participants took part in a cognitive debriefing interview. Round 1 also included three aesthetic plastic surgeons and one plastic surgery resident from Canada. Based on feedback, 22 items were retained, and 3 new items were added. Round 2 included five plastic surgeons, one dermatologist and two industry experts from Denmark, Canada, Sweden and the USA. Based on this round, 24 items were retained and 1 item was revised. In round 3, 138 Prolific participants accessed the screening survey for hair loss. We invited 70 people who met the study criteria to complete the survey and 63 did. Of the 63, 59 met the study inclusion criteria. For the 25 items tested, the option "I do not understand the question" was chosen 12 (0.8%) times, the option "I understand the question, but it was not relevant to me before or after treatment" was chosen 422 (28.6%) times and the option

**TABLE 4** Cause of hair loss reported by the sample of 337 in Prolific field-test.

| Cause of hair loss | *N* (%) |
|---|---|
| Family history of pattern baldness | 144 (42.7) |
| Aging | 139 (41.2) |
| Stress | 64 (19.0) |
| Hormonal imbalance | 53 (15.7) |
| Diet—lack of iron, protein, zinc, biotin in the diet | 47 (13.9) |
| Hairstyles—pulling on your scalp such as ponytails | 30 (8.9) |
| Hair care—colors, perms, or relaxing hair | 28 (8.3) |
| Medications | 20 (5.9) |
| Childbirth | 18 (5.3) |
| Thyroid disease | 18 (5.3) |
| Alopecia areata | 15 (4.5) |
| Psoriasis on the scalp | 15 (4.5) |
| Recovering from an illness or surgery | 10 (3.0) |
| Pulling your hair to relieve stress | 7 (2.1) |
| Scarring Alopecia | 3 (0.9) |
| Cancer treatments | 3 (0.9) |
| Scalp infection | 2 (0.6) |
| Other cause | 10 (3.0) |
| Do not know | 47 (13.9) |

*Note*: More than one cause could be selected.

14732165, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jocd.16465, Wiley Online Library on [16/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

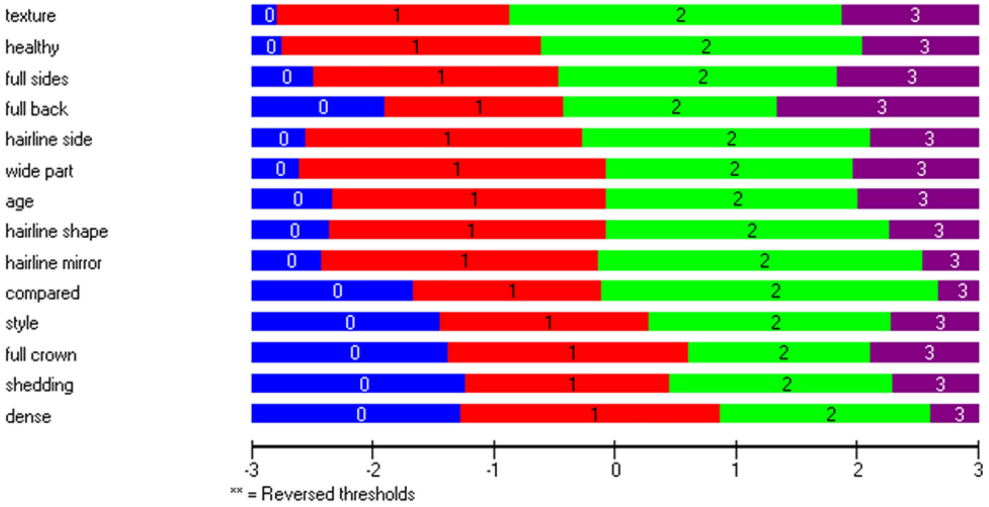


FIGURE 1 Threshold map for the 14 items that form HAIR-Q scale.

TABLE 5 RMT item level fit statistics and differential item function results.

| | RMT Statistics | | | | | | | DIF | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Scale | Location | SE | Fit residual | DF | $\chi^2$ | DF | *p*-value | Age | Gender | Treatment | Hair loss |
| 1. Texture | −0.59 | 0.08 | 2.56 | 345 | 11.89 | 5 | 0.04 | No | No | No | No |
| 2. Healthy | −0.44 | 0.08 | 0.30 | 345 | 3.62 | 5 | 0.60 | No | No | No | No |
| 3. Full—sides | −0.37 | 0.08 | −0.26 | 345 | 5.54 | 5 | 0.35 | No | Yes | No | No |
| 4. Full—back | −0.32 | 0.08 | 2.54 | 345 | 11.04 | 5 | 0.05 | No | No | No | No |
| 5. Hairline side | −0.24 | 0.08 | −0.73 | 345 | 3.66 | 5 | 0.60 | No | No | No | No |
| 6. Wide part | −0.24 | 0.08 | 0.31 | 345 | 7.14 | 5 | 0.21 | No | No | No | No |
| 7. Age | −0.13 | 0.08 | −0.72 | 345 | 6.01 | 5 | 0.31 | No | Yes | No | No |
| 8. Hairline shape | −0.05 | 0.08 | 1.27 | 345 | 2.47 | 5 | 0.78 | No | No | No | No |
| 9. Hairline mirror | 0.00 | 0.08 | −0.41 | 345 | 2.50 | 5 | 0.78 | No | No | No | No |
| 10. Compared | 0.30 | 0.08 | −3.41 | 345 | 12.49 | 5 | 0.03 | Yes | No | No | Yes |
| 11. Style | 0.38 | 0.08 | −1.33 | 345 | 3.54 | 5 | 0.62 | No | No | No | No |
| 12. Full crown | 0.45 | 0.08 | −1.29 | 345 | 3.10 | 5 | 0.68 | No | Yes | No | No |
| 13. Shedding | 0.51 | 0.08 | 1.97 | 345 | 6.55 | 5 | 0.26 | No | Yes | No | No |
| 14. Dense | 0.74 | 0.08 | −2.62 | 345 | 10.31 | 5 | 0.07 | No | No | No | No |

Abbreviations: DF, degrees of freedom; DIF, differential item functioning; RMT, Rasch measurement theory; SE, standard error; $\chi^2$, Chi-square.

"I understand the question and it was relevant to me before or after treatment" was chosen 1041 (70.6%) times. Based on this round, 21 items were retained, 1 item was revised, and 3 items were dropped. The pilot field test then included 22 items.

## 3.4 | Scale testing

The 59 participants who took part in the cognitive survey were invited to complete the pilot field test and 53 participants responded. Based on the pilot RMT analysis, no items were dropped.

For the field-test study, we screened a new sample of 2500 Prolific participants. After removing duplicates and incompletes, 2419 remained. Of these, 1713 did not experience hair loss, leaving 703 potential participants. Of those with hair loss, we invited all of those with a history of hair loss treatment to take part in the survey ($N = 178$). Of the remainder, we invited all participants who identified as a man or gender diverse ($N = 178$) and selected a random sample ($N = 175$) from 350 women. The total number invited to the survey was 531 and 397 responded. Of these, 4 were incomplete, 37 did not have treatment, and 19 provided suspicious data (inconsistent responses between screen and full survey) leaving 337 participants.

The RMT analysis combined the pilot and field-test datasets (total = 390). Participants ranged in age from 20 to 85 years. The average age was 37.7 (SD = 12). Of the 390 participants, 53.3% identified as men, 44.1 as women, and 2.3% as gender diverse. We

14732165, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jocd.16465, Wiley Online Library on [16/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

shortened the scale by eight items based on poor item fit to the Rasch model and redundant content. Data for the remaining 14 items fit the Rasch model (Chi-square = 89.85, df = 70, $p = 0.06$). All items had ordered thresholds (Figure 1). Table 5 shows the item fit statistics. All items had non-significant $p$-values after Bonferroni correction, and 10 items had fit residuals within +2.5. Potential DIF was identified for 5 items. One item had significant DIF for age-group and amount of hair loss, and 4 items had significant DIF for gender. When the items that evidenced DIF were split, the person locations for the original and split items correlated ≥0.99 indicating little impact on scoring. Figure 2 presents the person-item threshold distribution. The top histogram shows the sample, and the bottom histogram shows the scale. The sample was well targeted to the scale; 375 (96.2%) participants scored within the range of measurement provided by the scale. Floor and Ceiling effects were low at 1.8% and 2.1% respectively. Reliability was high with PSI (0.92, 0.91) and Cronbach alpha (0.94, 0.92) values with and without extremes respectively. Three pairs of items had residual that correlated above 0.20 suggestive of local dependency. When subtests were performed for the item pairs, PSI and Cronbach alpha values with and without extremes dropped slightly (i.e., PSI = 0.92, 0.91; Cronbach alpha = 0.92, 0.90).

### 3.5 | Construct validity

The mean HAIR-Q scale score for the sample of 390 participants was 48.7 (SD = 17.1). The mean score differed ($p = 0.012$) between women (mean = 46.1, SD = 17.6) and men (mean 50.6, SD = 16.6). No relationships were found for age or race (White vs. other).

Figures 3–5 show the construct validation tests. Scores were incrementally higher for participants who reported having more hair ($p \leq 0.001$), were more satisfied with their hair overall ($p \leq 0.001$), and were less bothered by hair loss ($p \leq 0.001$).

### 3.6 | Test retest reliability

Of the 124 participants in the TRT, 27 reported that their satisfaction with their hair had changed and were excluded from analysis. The ICC based on 97 participants was 0.94, with a 95% lower bound of 0.92 and upper bound of 0.96.

## 4 | DISCUSSION

Our team followed international guidelines to develop the HAIR-Q, a new PROM that measures satisfaction with hair. The HAIR-Q items were based on concepts that participants found important to them. Through multiple rounds of refinement and testing, we ensured the items in the HAIR-Q were easy to understand and relevant to participants, providing evidence of content validity. Psychometric properties were demonstrated including internal consistency, test-reliability, and construct validity.

These findings add to the published literature on PROMs for hair loss by providing a means to measure satisfaction with hair that is applicable across hair loss-related conditions and genders. Currently available PROMs for hair loss are specific to a condition, making comparison across etiologies difficult.[25,29] The HAIR-Q is more generic and measures a single latent construct, which differs from other available PROMs such as AA-QLI (symptoms, relationship; objective signs)[28] or Hair-specific Skindex-29 (symptom; function; emotion).[27] These PROMs are multi-dimensional and create an overall score based on two or more constructs.

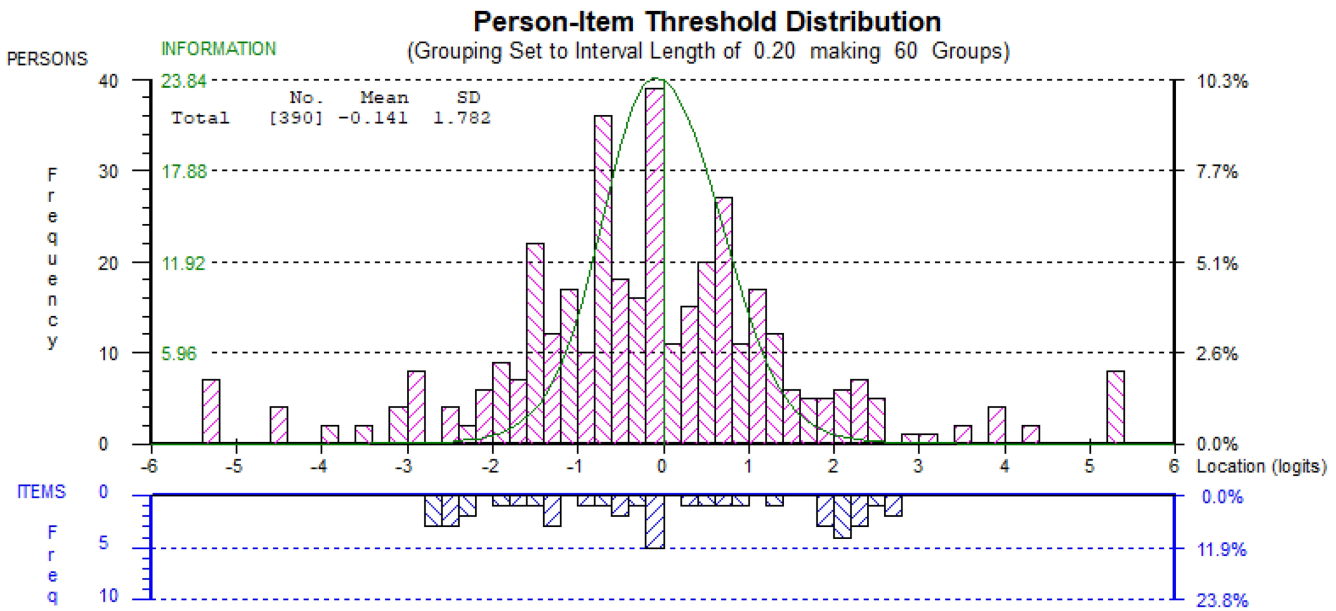


FIGURE 2 Person-item threshold distribution for the sample (upper histogram) and items (lower histogram).

14732165, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jocd.16465, Wiley Online Library on [16/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

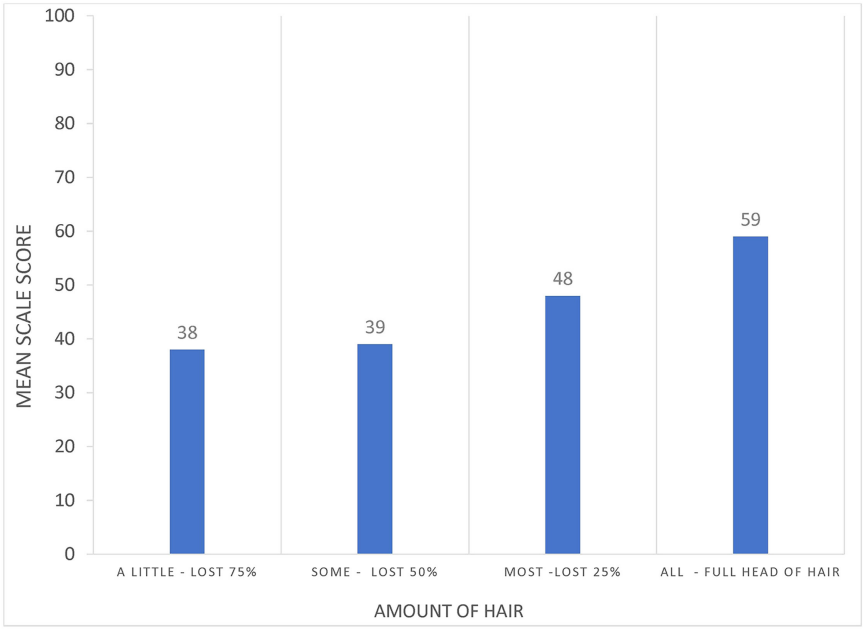


**FIGURE 3** Mean scores for HAIR-Q based on self-reported amount of hair on head. Based on field-test sample only ($n = 337$); Sample size: A little = 13, Some = 55, Most = 203, All = 66.

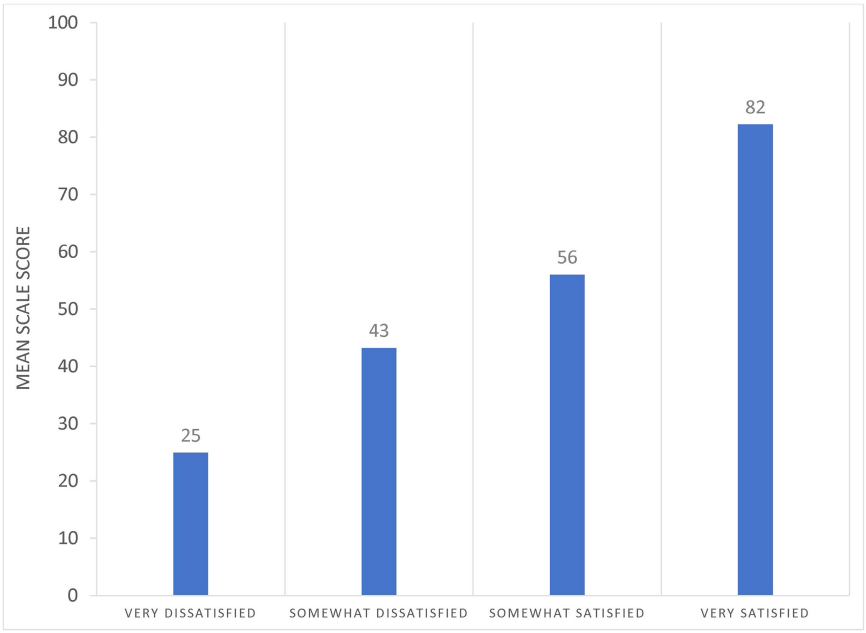


**FIGURE 4** Mean scores for HAIR-Q based participant level of satisfaction with hair overall. Based on full sample ($n = 390$); Sample size: Very dissatisfied = 57, Somewhat dissatisfied = 147, Somewhat satisfied = 156, Very satisfied = 30.

Challenges with multi-dimensional scales have been highlighted in the literature.[32,48] The COSMIN criteria notes that total scores from multiple scales should not be used unless there is evidence of unidimensionality.[49] Another advantage to the HAIR-Q was that it was developed using a modern psychometric approach and therefore has interval (rather than ordinal) level measurement properties.[50] Interval-level measures on a PROM scale works like a ruler, where the distance between each value is known and equal. Such a scale provides data that can be used in parametric statistics to measure differences between groups or changes over time.[48]

There are many ways PROMs, such as HAIR-Q, can be applied in clinical practice to help improve patient care.[51,52] By using a PROM routinely in care, data can be collected to evaluate practices and policies, inform shared decision-making, as well as to monitor a patient's condition over time.[53,54] When relevant, PROM data can also be collected as part of national registries, to help evaluate treatment effectiveness or assess quality of care,[55,56] or to inform value-based healthcare.[17] The inclusion of PROMs within clinical trials as a primary or secondary endpoint can provide valuable information regarding outcomes from the patient perspective, especially for those outcomes not readily observable.[57]

## 4.1 | Limitations

The initial sample of participants in the concept elicitation interviews for hair loss treatment was small and focused on aesthetic treatment.

14732165, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jocd.16465, Wiley Online Library on [16/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

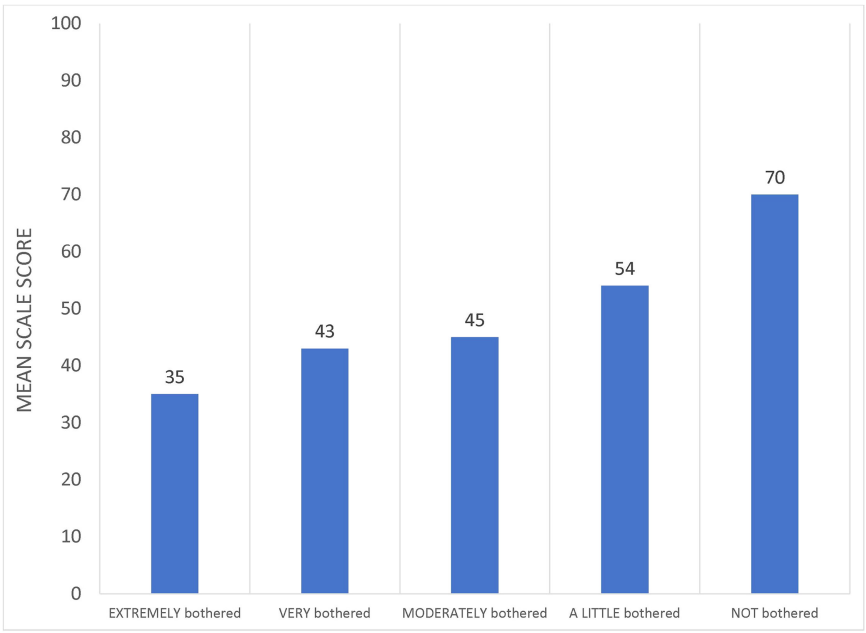


**FIGURE 5** Mean scores for HAIR-Q based on how bothered participants were by hair loss. Based on field-test sample only ($n = 337$); Sample size: Extremely bothered = 44, Very bothered = 63, Moderately bothered = 89, A little bothered = 115, Not bothered = 26.

This limitation was addressed by thoroughly investigating content validity using cognitive interviews, expert input and an online survey of patients whose hair loss had many causes. Some survey questions were not asked in the pilot field test (cause of hair loss, degree of bother by hair loss, and percentage of hair loss). However, the sample size for the field test sample was robust ($n = 390$). Data were collected from participants of an online platform, who self-select and are paid to take part in the research. Data provided through Prolific has been shown elsewhere to be of high quality in comparison to other similar platforms.[58,59]

## 5 | CONCLUSION

The HAIR-Q evidenced both validity and reliability and can be used pre or post treatment in adults who have hair loss as a result of any condition to measure the patient's perspective of satisfaction with hair. Further work is needed to assess psychometric properties not examined in this paper such as responsiveness and calculating minimally important differences to aid in interpretability. The HAIR-Q can be accessed at www.qportfolio.org.

### AUTHOR CONTRIBUTIONS

**Dr. Anne F. Klassen:** Conceptualization, Methodology, Data Curation, Formal analysis, Writing—Original Draft. **Jasmine Mansouri:** Investigation, Writing—Review & Editing. **Dr. Manraj Kaur:** Conceptualization, Methodology, Investigation, Writing—Review & Editing. **Dr. Charlene Rae:** Data Curation, Formal analysis, Writing—Review & Editing. **Dr. Lotte Poulsen:** Resources, Writing—Review & Editing. **Dr. Steven Dayan** Resources, Writing—Review & Editing. **Dr. Stefan Cano:** Conceptualization, Methodology, Formal analysis, Writing—Review & Editing. **Dr. Andrea L. Pusic:** Conceptualization, Methodology, Writing—Review & Editing.

### ACKNOWLEDGMENTS

Drs Klassen, Cano and Pusic are co-developers of the HAIR-Q scale and as such would receive a share of any license revenues as royalties based on their institutions' inventor sharing policy.

### DATA AVAILABILITY STATEMENT

The data are not publicly available due to privacy or ethical restrictions.

### ETHICS STATEMENT

Ethics board approval was obtained from the Hamilton Integrated Ethics Board (Canada) (#13603) at McMaster University in Canada on October 12, 2021. Participants in the qualitative interviews provided written and oral consent. Participants to the online survey were provided with a consent form to read, and electronic consent was obtained before participants started the survey.

### ORCID

*Jasmine Mansouri* https://orcid.org/0000-0002-9690-2943

### REFERENCES

1. Al Aboud AM, Zito PM. Alopecia. *StatPearls [Internet]*. StatPearls Publishing; 2023 [Accessed April 16, 2023]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK538178/
2. https://www.aad.org/public/diseases/hair-loss/causes/18-causes
3. Dhami L. Psychology of hair loss patients and importance of counseling. *Indian J Plast Surg*. 2021;54(4):411-415. doi:10.1055/s-0041-1741037
4. Moattari CR, Jafferany M. Psychological aspects of hair disorders: consideration for dermatologists, cosmetologists, aesthetic, and plastic surgeons. *Skin Appendage Disord*. 2022;8(3):186-194.

14732165, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jocd.16465, Wiley Online Library on [16/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5. Schmitt JV, Ribeiro CF, Souza FH, Siqueira EB, Bebber FR. Hair loss perception and symptoms of depression in female outpatients attending a general dermatology clinic. *An Bras Dermatol*. 2012;87:412-417.
6. Hunt N, McHale S. The psychological impact of alopecia. *BMJ*. 2005;331(7522):951-953.
7. Ahluwalia J, Fabi SG. The psychological and aesthetic impact of age-related hair changes in females. *J Cosmet Dermatol*. 2019;18(4):1161-1169.
8. Hadshiew IM, Foitzik K, Arck PC, Paus R. Burden of hair loss: stress and the underestimated psychosocial impact of telogen effluvium and androgenetic alopecia. *J Invest Dermatol*. 2004;123(3):455-457.
9. Alfonso M, Richter-Appelt H, Tosti A, Viera MS, García M. The psychosocial impact of hair loss among men: a multinational European study. *Curr Med Res Opin*. 2005;21(11):1829-1836.
10. Aldhouse NV, Kitchen H, Knight S, et al. "'You lose your hair, what's the big deal?'I was so embarrassed, I was so self-conscious, I was so depressed:" a qualitative interview study to understand the psychosocial burden of alopecia areata. *J Patient Rep Outcomes*. 2020;4(1):1-2.
11. Cash TF. The psychology of hair loss and its implications for patient care. *Clin Dermatol*. 2001;19(2):161-166.
12. Poot F. Psychological consequences of chronic hair diseases. *Rev Med Brux*. 2004;25(4):A286-A288.
13. Farrant P, McHale S. Psychological impact of hair loss. *Pract Psychodermatol*. 2014;20:79-89.
14. Mohamed NE, Soltan MR, Galal SA, El Sayed HS, Hassan HM, Khatery BH. Female pattern hair loss and negative psychological impact: possible role of brain-derived neurotrophic factor (BDNF). *Dermatol Pract Concept*. 2023;13(3):e2023139. doi:10.5826/dpc.1303a139
15. Meyers A, Jin A, Kwiecien GJ, Gatherwright J, Khetarpal S, Zins JE. Platelet-rich plasma for treatment of hair loss improves patient-reported quality of life. *Aesth Plast Surg*. 2023;13:1-7.
16. Zhuang XS, Zheng YY, Xu JJ, Fan WX. Quality of life in women with female pattern hair loss and the impact of topical minoxidil treatment on quality of life in these patients. *Exp Ther Med*. 2013;6(2):542-546.
17. Squitieri L, Bozic KJ, Pusic AL. The role of patient-reported outcome measures in value-based payment reform. *Value Health*. 2017;20(6):834-836.
18. Meadows KA. Patient-reported outcome measures: an overview. *Br J Community Nurs*. 2011;16(3):146-151.
19. Food and Drug Administration. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims. Silver Spring, MD. 2009. Accessed: May 20, 2023 https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf
20. Mokkink LB, Prinsen CA, Patrick DL, et al. *COSMIN Study Design Checklist for Patient-Reported Outcome Measurement Instruments*. The Netherlands; 2019:1-32.
21. Prinsen CA, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27:1147-1157.
22. Terwee CB, Prinsen CA, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27:1159-1170.
23. Meadows K. Patient-reported outcome measures: an overview. *Br J Community Nurs*. 2011;16:146-151.
24. Arguelles G, Shin M, Lebrun D, Kocher M, Baldwin K, Patel N. The majority of patient-reported outcome measures in pediatric Orthopaedic research are used without validation. *J Pediatr Orthop*. 2021;41(1):e74-e79. doi:10.1097/BPO.0000000000001659
25. Xia E, Li SJ, Drake L, et al. An assessment of current clinician-reported and patient-reported outcome measures for alopecia Areata: a scoping review. *J Invest Dermatol*. 2023;143:1133-1137.e12.
26. Wyrwich KW, Kitchen H, Knight S, et al. Development of the scalp hair assessment PRO™ measure for alopecia areata. *Br J Dermatol*. 2020;183(6):1065-1072. doi:10.1111/bjd.19024
27. Mendoza TR, Osei JS, Shi Q, Duvic M. Development of the alopecia areata symptom impact scale. *J Investig Dermatol Symp Proc*. 2013;16(1):S51-S52.
28. Fabbrocini G, Panariello L, De Vita V, et al. Quality of life in alopecia areata: a disease-specific questionnaire. *J Eur Acad Dermatol Venereol*. 2013;27:e276.
29. Chren MM. The Skindex instruments to measure the effects of skin disease on quality of life. *Dermatol Clin*. 2012;30(2):231-236. doi:10.1016/j.det.2011.11.003
30. Han SH, Byun JW, Lee WS, et al. Quality of life assessment in male patients with androgenetic alopecia: result of a prospective Multicenter Study. *Ann Dermatol*. 2012;24(3):311-318.
31. Barber BI, Kaufman KD, Kozloff RC, Girman CJ, Guess HA. A hair growth questionnaire for use in the evaluation of therapeutic effects in men. *J Dermatol Treat*. 1998;9(3):181-186.
32. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess*. 2009;2009:214.
33. Klassen AF, Pusic A, Kaur M, et al. The SKIN-Q: an innovative patient-reported outcome measure for evaluating minimally invasive skin treatments for the face and body. *Facial Plast Surg Aesthet Med*. 2024;26(3):247-255. doi:10.1089/fpsam.2023.0204
34. Klassen AF, Pusic A, Kaur M, et al. Extending the range of measurement for minimally invasive treatments by adding new concepts to FACE-Q aesthetics scales. *Plast Reconstr Surg Glob Open*. 2024;12(4):e5736. doi:10.1097/GOX.0000000000005736
35. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1—eliciting concepts for a new PRO instrument. *Value Health*. 2011;14(8):967-977.
36. Thorne S, Kirkham SR, MacDonald-Emes J. Interpretive description: a noncategorical qualitative alternative for developing nursing knowledge. *Res Nurs Health*. 1997;20(2):169-177.
37. Pope C, Ziebland S, Mays N. Qualitative research in health care analysing qualitative data. *Br Med J*. 2000;320(7227):114-116.
38. Sandelowski M. Theoretical saturation. In: Given LM, ed. *The Sage Encyclopedia of Qualitative Methods*. Vol 1. Sage; 2008:875-876.
39. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381.
40. Rasch G. Studies in mathematical psychology: I. Probabilistic Models for some Intelligence and Attainment Tests. 1960.
41. Andrich D, Sheridan BS, Luo G. *RUMM2030Plus: Rasch Unidimensional Models for Measurement*. RUMM Laboratory; 2021. [Accessed: October 10, 2023]. www.rummlab.com.au
42. Christensen KB, Makransky G, Horton M. Critical values for Yen's Q 3: identification of local dependence in the rasch model using residual correlations. *Appl Psychol Meas*. 2017;41(3):178-194.
43. Cleanthous S, Bongardt S, Marquis P, Stach C, Cano S, Morel T. Psychometric analysis from EMBODY1 and 2 clinical trials to help select suitable fatigue pro scales for future systemic lupus erythematosus studies. *Rheumatol Ther*. 2021;8(3):1287-1301.
44. Andrich D, Hagquist C. Real and artificial differential item functioning. *J Educ Behav Stat*. 2012;37(3):387-416.
45. Andrich D. An index of person separation in latent trait theory, the traditional KR.20 index, and the Guttman scale response pattern. *Educ Res Perspect*. 1982;9(1):95-104. Mary Ann Liebert, Inc. Page 22 of 49.


14732165, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jocd.16465, Wiley Online Library on [16/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

46. Nunnally JC. *Psychometric Theory*. 3rd ed. McGraw-Hill; 1994.
47. Kim HY. Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod*. 2013;38(1):52-54. doi:10.5395/rde.2013.38.1.52
48. McKenna SP, Heaney A, Wilburn J, Stenner AJ. Measurement of patient-reported outcomes. 1: the search for the holy grail. *J Med Econ*. 2019;22(6):516-522.
49. Mokkink LB, Terwee CB, Patrick DL, et al. *COSMIN Checklist Manual*. University Medical Center; 2012.
50. McAllister S. Introduction to the use of Rasch analysis to assess patient performance. *Int J Ther Rehabil*. 2008;15(11):482-490.
51. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. *BMJ*. 2015;10:350.
52. Browne JP, Cano SJ, Smith S. Using patient-reported outcome measures to improve health care. *Med Care*. 2017;55(10):901-904.
53. Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;28(346):f167. doi:10.1136/bmj.f167
54. Damman OC, Jani A, de Jong BA, et al. The use of PROMs and shared decision-making in medical encounters with patients: an opportunity to deliver value-based health care to patients. *J Eval Clin Pract*. 2020;26(2):524-540.
55. Ruseckaite R, Maharaj AD, Krysinska K, Dean J, Ahern S. Developing a preliminary conceptual framework for guidelines on inclusion of patient reported-outcome measures (PROMs) in clinical quality registries. *Patient Relat Outcome Meas*. 2019;10:355-372.
56. Prodinger B, Taylor P. Improving quality of care through patient-reported outcome measures (PROMs): expert interviews using the NHS PROMs Programme and the Swedish quality registers for knee and hip arthroplasty as examples. *BMC Health Serv Res*. 2018;18:1-3.
57. Mercieca-Bebber R, King MT, Calvert MJ, Stockler MR, Friedlander M. The importance of patient-reported outcomes in clinical trials and strategies for future optimization. *Patient Relat Outcome Meas*. 2018;1:353-367.
58. Peer E, Rothschild D, Gordon A, Evernden Z, Damer E. Data quality of platforms and panels for online behavioral research. *Behav Res Methods*. 2022;54(4):1643-1662.
59. Douglas BD, Ewell PJ, Brauer M. Data quality in online human-subjects research: comparisons between MTurk, prolific, CloudResearch, Qualtrics, and SONA. *PLoS One*. 2023;18(3):e0279720. doi:10.1371/journal.pone.0279720

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Klassen AF, Mansouri J, Kaur M, et al. Development and validation of a patient-reported outcome measure for hair loss treatment: The HAIR-Q. *J Cosmet Dermatol*. 2024;00:1-11. doi:10.1111/jocd.16465

14732165, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jocd.16465, Wiley Online Library on [16/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License