ACPA
American Cleft Palate Craniofacial Association

Original Article

The Cleft Palate Craniofacial Journal
1-11

DOI: 10.1177/10556656231160392
journals.sagepub.com/home/cpc
SAGE

Translation, Validation, and Cultural Adaptation of CLEFT-Q© for use in Indonesia

Prasetyanugraheni Kreshanti, MD[1], Kasih Rahardjo Djarot, MD[1], Fransiska Kaligis, MD[2], Dewi Friska, MD[3], Jordan W. Swanson, MD, MSc[4], Jessica Blum, MSc[4], Valencia Jane Martin, MD[1], and Kristaninta Bangun, MD[1]

Abstract

Objective: To translate and validate CLEFT-Q©, patient-reported outcome measure for patients with cleft lip and/or palate (CL and/or P), into Indonesian. CLEFT-Q© covers the domains of appearance, facial function, health-related quality of life and consists of scales describing outcomes after cleft surgery.

Design: The CLEFT-Q© instrument was translated according to the International Society of Pharmacoeconomics and Outcomes Research guidelines, including translation, cognitive debriefing, and field-testing.

Setting: Dr. Cipto Mangunkusumo Hospital, Indonesia; independent CL and/or P support groups.

Patients: Patients ages 8-29 with a history of repaired CL and/or P were grouped based on age. Those unable to complete the questionnaire independently were excluded.

Interventions: The primary objective was reliable translation of the CLEFT-Q® instrument. Each scale was assessed for its internal consistency (Cronbach's alpha) and validity (inter-item correlation), and sub-group analyses were performed based on age group.

Results: Forward and back translation revealed 25(13.3%) and 12(6.3%) of items were difficult to translate. Cognitive debriefing revealed 10(5.3%) items were difficult to understand, with the lowest reliability on the facial appearance scale (α=0.27). Other scales demonstrated acceptable to excellent reliability (α=0.53–0.68). Field testing revealed acceptable reliability and validity of the translation (α = 0.74-0.92; 69% ideal range of inter-item correlation). Sub-group analyses revealed patients in the <11y.o. and >18y.o. groups had the lowest scores on the "cleft lip scar" scale while those 11-18y.o. had the lowest scores on the "nostrils" scale.

Conclusion: Iterative translation and cultural adaptation of CLEFT-Q© into Indonesian demonstrated reliability and validity of the tool, supported by acceptable to excellent internal consistency and ideal inter-item correlation.

Keywords

quality of life, counseling, psychosocial adjustment, speech perception, psychological assessment, nonsyndromic clefting

Introduction

Cleft lip and/or palate (CL and/or P) has an estimated worldwide prevalence of 0.3 to 0.45 in 1000 live births,[1] with significant variability between regions and over time.[2,3] In 2017 alone, the global prevalence of orofacial clefts was reportedly 141.56 cases (range 130.17 to 152.53 cases) per 100 000 in the population.[3] Children with CL and/or P are susceptible to issues such as initial feeding difficulties, esthetic differences, hearing loss, dental abnormalities and malocclusion, airway obstruction, velopharyngeal insufficiency, and other issues secondary to their craniofacial differences.[4] Included among the potential sequelae of CL and/or P are serious speech, hearing, nutritional, mental, and social developmental disorders.[1] One study out of the

[1] Cleft and Craniofacial Center Dr. Cipto Mangunkusumo Hospital, Division of Plastic Reconstructive and Aesthetic Surgery, Department of Surgery, Faculty of Medicine Universitas Indonesia, Jakarta, Indonesia
[2] Department of Psychiatry, Dr. Cipto Mangunkusumo Hospital, Faculty of Medicine Universitas Indonesia, Jakarta, Indonesia
[3] Department of Community Medicine, Faculty of Medicine Universitas Indonesia, Jakarta, Indonesia
[4] Division of Plastic, Reconstructive and Oral Surgery, Children's Hospital of Philadelphia, Philadelphia, PA, USA

Corresponding Author:
Prasetyanugraheni Kreshanti, Cleft and Craniofacial Center Dr. Cipto Mangunkusumo Hospital, Division of Plastic Reconstructive and Aesthetic Surgery, Department of Surgery, Faculty of Medicine Universitas Indonesia, Jakarta, Indonesia.
Email: prasetyanugraheni@ui.ac.id

United States revealed patients with CL and/or P typically require an average of 8.6 surgical procedures prior to adulthood to optimize form and function.[5] These surgeries may include but are not limited to lip adhesion, palate repair, lip repair, rhinoplasty, placement of myringotomy tubes, speech surgery, orthognathic surgery, alveolar bone grafts, and fistula revision.[5] In addition to the multitude of surgical interventions indicated, comprehensive cleft care requires multidisciplinary management throughout childhood to address the functional, esthetic, and psychosocial concerns that arise.[6]

A key element of optimizing cleft interventions is taking into account the patient's perspective on the adequacy or limitations of their function and appearance. Patient reported outcome measures (PROMs) complement clinical assessment and interviews, providing key insights into a patient's quality of life beyond what they may be willing or able to verbally share. Clinicians can use PROMs to both track treatment outcomes and guide future interventions.[7,8] The field-test version of CLEFT-Q© consists of 13 scales and 119 items measuring appearance, health-related quality of life (HRQoL), and facial function.[8] Due to its clinical applicability and effectiveness, the CLEFT-Q© questionnaire has been translated and validated in many non-English languages,[9] opening the door for cross-cultural comparison of CL and/or P-related HRQoL.

Despite the many available versions of the CLEFT-Q©, a validated version in Indonesian is not yet available. In addition to making CLEFT-Q© available to the fourth most populous country in the world, deploying the instrument and studying the pattern of patient-reported outcomes may offer valuable cross-cultural insight into perceptions of cleft treatment. Given the potentially significant contribution of the CLEFT-Q© to collection and comparison of CL and/or P evidence-based patient outcomes,[10] the primary goal of this study was to translate and validate the CLEFT-Q© for use in Indonesia through forward and backward translation and cognitive debriefing. A secondary aim of this study was to analyse initial responses to the Indonesian CLEFT-Q© through field testing of the translated instrument.

The CLEFT-Q©

The CLEFT-Q© is a PROM designed specifically for patients with CL and/or P ages 8 to 29 years to evaluate their pre- and post-operative appearance, speech, and HRQoL.[10] The CLEFT-Q© is owned by McMaster University (Hamilton, Canada) and the Hospital for Sick Children (Toronto, Canada). The CLEFT-Q© was designed in accordance with guidelines from the United States Food and Drug Administration,[11] the Scientific Advisory Committee of the Medical Outcomes Trust,[12] and the International Society for Pharmacoeconomics and Outcomes Research (ISPOR).[13,14] The methods for development and validation of the tool have been previously described.[8] The CLEFT-Q© is divided into three domains: appearance, facial function, and health-related quality of life, each containing multiple scales. The seven appearance scales evaluate the face, nose, nostrils, teeth, jaws, lips, and cleft lip scar. The two facial function scales evaluate speech function and eating and drinking. Lastly, the four health-related quality of life scales evaluate psychological state, school life, social life, and speech distress. Each scale contains 3-4 Likert-type response options (eg, always-sometimes-never, or not at all-a little bit-quite a bit-very much). The raw scores are converted to a 0-100 scale using a conversion table provided by the CLEFT-Q© developers, with higher scores reflecting a better outcome.[10] The CLEFT-Q© contains 119 items, and a total of 189 components were translated including all instructions, response options, and items.

Methods

Human Subjects Protection

This study was approved by the Ethics Committee of the Faculty of Medicine Universitas Indonesia (KET-427/UN2.F1/ETIK/PPM.00.02/2020). Patients' data are stored in archives that can only be accessed by the investigators.

Selection of Translators

All translators involved were licensed professionals from a medical translation company. The medical translation company recruits and evaluates employees through their standardized process consisting of advanced Test of English as a Foreign Language (TOEFL) or expert International English Language Testing System (IELTS) scores, an interview, and a translation test. Following formal recruitment, translators underwent ongoing qualitative evaluation by proofreaders.

To conduct forward translations, the research team recruited two native-Indonesian certified translators with English proficiency, prior experience in patient-reported outcomes, and medical degrees. For back translation, a third licensed translator whose mother tongue was English and who was fluent in Indonesian was recruited. The back translator had an academic title in English literature. Of note, the back translator had no prior knowledge of the source version of the questionnaire.

Selection of Participants

Study participants for cognitive debriefing and field testing were selected via convenience consecutive sampling from both the Cleft and Craniofacial Center at Dr. Cipto Mangunkusumo Hospital database and independent cleft lip-and-palate communities based in Indonesia. Inclusion criteria included patients with CL and/or P ages 8 to 29 years old who underwent cleft correction surgery. Patients who could not complete the questionnaire independently and/or had intellectual disabilities associated with severe facial clefts and/or syndromes were excluded. All participants who met inclusion criteria and were approached for consent agreed to participate in the study. Participants

were divided into three age groups based on their psychosocial developmental stage: < 11 years old (childhood), 11 −18 years old (adolescence), and >18 years old (adulthood), according to International Association for Child and Adolescent Psychiatry and Allied Professions (IACAPAP).[15] Each stage is characterized by unique cognitive, linguistic, social-emotional, and behavioral criteria with the potential to impact one's reported outcomes.[15] Participant demographics are described in Table 1.

Translation, Validation, and Cultural Adaptation Process

Translation, validation and cultural adaptation of the Indonesian CLEFT-Q© took place between June and August of 2020. Prior to translation, we obtained a license agreement from McMaster University. International guidelines set by McMaster University based on the International Society of Pharmacoeconomics and Outcomes Research (ISPOR) guided the iterative translation process.[13,14] The process of translation contained several stages which include: forward and backward translation of the original questionnaire (Version 1 and Version 2), cognitive debriefing, review of the questionnaire after cognitive debriefing (Version 3), finalization of the questionnaire by CLEFT-Q© developers (final version), and field testing. The validity of the questionnaire was evaluated by considering its content validity, transcultural validity, and construct validity. Content validity was assessed via discussion between the research team and CLEFT-Q© developers. Transcultural validity was assessed using cognitive debriefing following forward and backward translation of the questionnaire. Construct validity was assessed after cognitive debriefing and field testing using inter-item correlation assessed statistically. A project manager oversaw the entirety of the translation and cultural adaptation process.

Version 1

Forward translation into the target language was performed by two translators. Items were worded using vernacular Indonesian to facilitate comprehension across literacy levels. Any inconsistencies in the translations were resolved at a consensus meeting for reconciliation and harmonization of the two translations that included the project manager, the research team, and both forward translators. This process resulted in the Version 1 translation of the tool in the target language. Harmonization was performed to highlight differences between the original version and its derivative translations and to arrive at a consistent approach to translation issues. Reconciliation seeks to rectify any differences in language preference between the project manager and translators, as well as any disputes regarding the forward translation.

Version 2

An additional translator was tasked with back-translating the Version 1 translation into the source language. Comparison of the back translation to the original source language version was performed by the project manager and the research team in collaboration with the CLEFT-Q© developers, resulting in Version 2.[16] Version 2 of the translation was used for cognitive debriefing.

Cognitive Debriefing

Cognitive debriefing was conducted on five participants using Version 2 of the translation. Cognitive debriefing is the process of testing an instrument among representatives of a target population and language group to assess for accurate comprehension.

Table 1. Demographic Characteristics of Cognitive Debriefing and Field-Testing Participants.

Characteristics	Cognitive Debriefing (Number of participants (%))	Field Testing (Number of participants (%))		
		A	B	C
Age				
<11 y.o.	1 (20%)	1 (3.3%)	6 (20%)	-
11-18 y.o.	2 (40%)	4 (13.3%)	5 (16.7%)	-
>18 y.o.	2 (40%)	1 (3.3%)	11(36.7%)	2 (6.7%)
Sex				
Male	3 (60%)	3 (10%)	8 (26.7%)	-
Female	2 (40%)	3 (10%)	14 (46.7%)	2 (6.7%)
Cleft Type and Subtypes				
Cleft lip only	1 (20%)	1 (3.3%)	5 (16.7%)	1 (3.3%)
Cleft palate only	-	-	2 (6.7%)	-
Cleft lip and palate	4 (80%)	5 (16.7%)	15 (50%)	1 (3.3%)
Cleft lip and alveolus	-	-	-	-
Total	5	6	22	2

*A: Cleft and Craniofacial Center Dr. Cipto Mangunkusumo Hospital Jakarta, B: Patient support group 1, C: Patient support group 2.

Cognitive debriefing was performed in a face-to-face, one-on-one setting by the investigator. Participants were asked to complete the questionnaire by themselves with the investigator present in the room to elaborate on items when needed. For questions which were misinterpreted, the investigator assisted by explaining the intended question and pointing out available image references, which were already present in the questionnaire, in order to facilitate understanding of the instructions, question stem, and/or responses. Participants <11 years old were accompanied by their parents, but parents were asked not to interfere with the process. Investigators took note of any difficulties that participants encountered. Results of the cognitive interviews were discussed with the CLEFT-Q© developers and resulted in Version 3 of the tool.

Version 3

The cognitive debriefing results were input in the CLEFT-Q© translation report spreadsheet and reviewed by the research teiam. The reviewed results of the cognitive interviews were sent to the CLEFT-Q© developers and discussed via email. For the Cleft Lip Scar, Psychological Function, School Function, Social Function, Speech Distress, and Eating and Drinking scales, some literal translations of the items were not well-interpreted by the participants (see Results). These terms were later discussed and replaced with more common terms used in Indonesian culture and vernacular, resulting in Version 3. This step constituted transcultural validation of the instrument.

Final Version

Version 3 of the translation was finalized and formatted by the CLEFT-Q© developers, resulting in the Final Version. Subsequently, reliability and validity of the translated scales were assessed by the research team using statistical analyses.

Field-Testing

The approved final version of the Indonesian CLEFT-Q© was distributed to 30 participants. Seventy percent of the participants came from one of the largest independent CL and/or P support groups in Indonesia. The male-to-female ratio was 11:19, with seven participants in the <11 y.o. age group, nine participants in the 11-18 y.o. age group, and 14 participants in the >18 y.o. age groups. Twenty-one out of 30 participants had cleft lip and palate (Table 1).

Field testing was performed online through video call by the investigators due to the COVID-19 pandemic. The questionnaire was shown on the screen and participants were asked to read the questions and provide answers which were noted by the investigators.

Statistical Analyses

The cognitive debriefing and field-test raw scores from participants were converted to standardized scores from 0 (worst) to 100 (best) using CLEFT-Q© conversion tables. The converted values were then assessed for reliability and validity. Reliability, or internal consistency, was measured using Cronbach's alpha (α). A Cronbach's alpha (α) value less than 0.5 reflects unacceptable internal consistency, $0.5 < \alpha < 0.6$ is poor, $0.6 < \alpha < 0.7$ is questionable, $0.7 < \alpha < 0.8$ is acceptable, $0.8 < \alpha < 0.9$ is good, and an α value greater than or equal to 0.9 is considered excellent.[17] Statistically, construct validity is measured using inter-item correlation. The ideal range of average inter-item correlation is 0.2 - 0.4. Less than 0.1 is considered to be poorly correlated while more than 0.5 is considered redundant.[18] The mean and standard deviation of raw responses are the two of the measures of descriptive statistics. Data was recorded in Microsoft Excel and analysed using Statistical Package for Social Sciences version 22.0 (SPSS, Inc, Chicago, IL, USA).

Results

Translation and Validation

Version 1. Reconciliation of two forward translations revealed inconsistencies related to the wording or phrasing of items. Out of 189 items, 25 items (13.3%) were found to be discordant. Detailed discordance of the forward translation is described in Table 2. Most commonly, word choices were discordant with the source language, leading to different interpretations of the question. The primary source of difficulty encountered in forward translation was a relatively limited vocabulary of the target language. For example, for the question "how much do you like your nostrils when you smile?" each translator translated the word "nostrils" differently, with one translator using the Indonesian equivalent of "nostrils" and the other using the equivalent of "nasal lobes", each with an obviously different meaning to a native English speaker.

Similar findings were also identified for the "appearance of the jaw" scale. One translator translated "jaw" as "rahang", and the other translated it as "dagu". In English, "dagu" means "chin" – which only represents one part of the whole jaw. We decided to use the word "rahang" as it includes the entirety of the jaw. Other inconsistencies were found in the "speech function" scale. Specifically, for the word "speech", one translator translated it as "pembicaraan", which meant "conversation", while the other translated it as "cara bicara". We decided to use the word "cara bicara" as it means "the way someone speaks".

Version 2. The result of back translation also revealed several discrepancies between the back translation and the source language. Twelve items out of 189 (6.3%) were considered to have a slightly different meaning than the source language. For example, the response option "quite a bit" was back-

Table 2. Translation Discordance During Forward Translation.

Scale	Item	Difficult to translate (specific word or phrase)	Description	Solution
Lips	…how full your lips look?	Full	Fullness of the lips translates as "thickness" in Indonesian	Replace "full" with "thick"
Nose	…how well both sides of your nose match?	Match	Match translates as "similarity" in Indonesian	Replace "match" with "similarity"
Nostrils	Instruction, items 1-6	Nostrils	Nostrils was translated as "cuping hidung" and "lubang hidung"	"Lubang hidung" is used because it is more common terminology
Jaws	Instruction, items 1-7	Jaw	Jaw was translated as "rahang" and "dagu"	"Rahang" was used because it includes the whole part of jaw, while "dagu" means chin
Teeth	…how close together your teeth are	Close together	Close together was translated as "rapat" and "dekat"	"Rapat" was used because it is a more common term
Teeth	…how your top and bottom teeth meet when you bite?	Meet	Meet was translated as "bertemu" and "beradu"	"Bertemu" was used because it is a more common term
School function	I like seeing my friends at school	Seeing	Seeing was translated as "melihat" and "bertemu"	"Bertemu" was chosen because it represent the meaning in Indonesian
Speech function	Instruction, items 1,2,6,7	Speech	Speech was translated as "cara bicara" and "pembicaraan"	"Cara bicara" was chosen because it means the way someone speaks, while "pembicaraan" means conversation

translated as "enough", yielding a different interpretation. In the forward translation, "quite a bit" was translated as "cukup", which means "quite a bit" and "enough" in Indonesian, resulting in a discrepancy in the back translation. Other differences were also found in the appearance of the lips scale—"How much do you like how full your lips look?" was back-translated as "how much do you like the thickness of your lips?". The source of this difference can be traced to the term used in Indonesian to describe fullness of the lips, which is "thick" ("tebal" in Indonesian) rather than "full".

Differences in terms used in the source and target language pose a problem in translation and cultural adaptation of this questionnaire, as described by the previous examples. Furthermore, the limited vocabulary offered by the target language proved to be troublesome. This was especially relevant when it came to describing feelings on the psychological function scale. For example, "great" was back-translated as "amazing" as there are limited words in Indonesian that delineate the degree of greatness. The "I feel good about how I look" item is meant to assess the patient's satisfaction with their appearance, but since "how I look" was translated to a word meaning "appearance" ("penampilan" in Indonesian), the back-translation yielded the item, "I feel my appearance is good".

Details of the discordance are described in Table 3. Following review of the back translation and reconciliation between the project manager and research team, all discrepancies were resolved and resulted in Version 2 of the tool.

Version 3. Participants from all age groups were represented in the cognitive debriefing phase, with one participant in the <11 y.o. group; two in the 11-18 y.o. group; and two in the >18 y.o. group. The male-to-female ratio was 3:2, including one participant with isolated cleft lip and four participants with cleft lip and palate. Debriefing revealed 10 (5.3%) items from the questionnaire were not easily understood by participants. For example, on the "appearance of the face" scale, two younger participants aged 9 and 13 y.o. had trouble understanding the question "how much do you like the shape of your face?" as many had never thought about this before. On the "appearance of the cleft lip scar" scale, the item, "how much do you like the size of your cleft lip scar?" item created confusion for participants, requiring the investigator to explain that "size" refers to the length and width together. The items that caused confusion among participants are delineated further in Table 4.

The internal consistency and validity from cognitive debriefing can be found in Table 5. The scales pertaining to appearance of the face demonstrated unacceptable internal consistency ($\alpha = 0.27$). Other scales with poor to questionable α values included social function, school function, and eating and drinking with α values of 0.54, 0.68, and 0.53 respectively. The remainder of the scales showed acceptable to excellent internal consistency ($\alpha = 0.70–0.96$).

Furthermore, the validity measured by the inter-item correlational analyses revealed two scales with an average

Table 3. Translation Discordance During Back Translation.

Scale	Item	Difficult to translate (specific word or phrase)	Description	Changed meaning post back translation	Solution
Response	Quite a bit	Quite a bit	Limited vocabulary in target language leads to differences in translation. "Quite a bit" translates the same as "enough" in Indonesian	Yes	Discussion with CLEFT-Q team to reach agreement that although it might translates differently in back translation, it does not change its meaning in target language
Lips	…how full your lips look	How full	Indonesians tend to describe lips by its thickness, not fullness	No	n/a
Teeth	…how close together your teeth are	How close together	Used "density" instead of "closeness"	No	n/a
Teeth	…how straight your teeth look	How straight	Used "even" instead of "straight"	Yes	Discussion with CLEFT-Q team to reach agreement that although it might translates differently in back translation, it does not change its meaning in target language
Psychological Function	I feel okay about myself	Okay	In Indonesian "okay" and "fine" has the same terminology	No	n/a
Psychological Function	I believe in myself	I believe in myself	It's not common for Indonesians to say "I believe in myself", but we use "I believe in my own abilities"	Yes	Change to "I believe in my own abilities"
Psychological Function	I feel good about how I look	How I look	In Indonesian, "how I look" describes as "penampilan" (appearance)	No	n/a
Psychological Function	I feel great about myself	Great	"Great" translates the same as "amazing" in Indonesian. Limited words in Indonesian that delineate greatness	Yes	Change to "I am feeling in a very good condition"
School Function	I like seeing my friends at school	Like	"Like" translates the same as "happy" in Indonesian	No	n/a
Social Function	It's okay when people look at my face	Okay	"Okay" in this context translates better with "comfortable" in Indonesian	Yes	Discussion with CLEFT-Q team that the intended meaning in target language does not change
Speech Distress	I avoid going out because of my speech (like to a party)	Going out	In Indonesian, "meeting people" is more common to use instead of "going out"	No	n/a
Eating and drinking	I have to take small bites when I eat	Small bites	"Small pieces" is more common in Indonesian compared to "small bites"	No	n/a

Table 4. Items Causing Difficulties among Participants During Cognitive Debriefing.

Scale	Item	Participants' Difficulties	Management
Face	…the shape of your face (eg, round or oval)	For participants aged 9 and 13 they rarely thought about this, so it took time for them to understand the question	n/a
Cleft lip scar	…the width of your cleft lip scar	"Width" refers to the thickness of the scar, which would have translated differently in Indonesian	Replace "width" with "thickness"
Cleft lip scar	…the size of your cleft lip scar	Size refers to length and width altogether. Participants needed to be shown illustration	Added more description: "how much do you like the size of your cleft lip scar (length and width)?"
Psychological Function	I feel great about myself	"Great" translates similarly with "amazing" in Indonesian but tends to have bad connotation. Participants understand better with "feeling in a very good condition"	Change to "I am feeling in a very good condition"
School Function	I feel safe at school (not bullied)	Participants understand better the term "being called names", "mocked"	Replace "bullied" with "being called names"
Social Function	I like being with other people	Participants understand better with "being around other people"	Change to "I like being around other people"
Speech Distress	I get teased about my speech	Participants age 9 and 13 understand better the term "being made fun of"	Change to "I get made fun of about my speech"
Speech Distress	I get frustrated when I speak	Participants (age 9 and 13) had trouble understanding "frustrated". The term "stressed" is understood better	Change "frustrated" to "stressed"
Speech Function	It's hard for new people to understand my speech	A 13 y.o. participant first thought of it as "it's hard for new people to understand him as a person", not referring to the way he speaks	n/a
Eating and Drinking	Food falls out of my mouth when I eat	It was difficult for participants to understand "makanan jatuh dari mulut saya ketika saya makan"; they thought that the whole food fall from their mouth. We explained about how sometimes small parts of the food falls out of their mouth when they have a set of small lips or mouth, or when they aren't able to retain some of the food in their mouth due to problems in lip muscles, etc.	Change into "makanan ada yang keluar dari mulut saya ketika saya makan". We use "leaking out" instead of "falls out"

correlation less than 0.10 (15%, range −0.03-0.02), three scales with average correlation between 0.20 and 0.40 (23%, range 0.22-0.24), and six scales with an average correlation more than 0.50 (46%, range 0.54-0.67).

Final Version. Version 3 of the translation was finalized and proof-read, yielding the final version. Field-testing was performed using the final version of the translated instrument on 30 participants who were predominantly male (n = 19, 63%), > 18 years old (n = 14, 47%), and had a diagnosis of cleft lip and palate (n = 21, 70%) (Table 1). The internal consistency for each scale ranged from acceptable to excellent with Cronbach's α values from 0.74 to 0.92, with special attention to the incremental increase in the α value of the face scale from 0.27 (poor internal consistency) in previous cognitive debriefing to 0.81 (good internal consistency) in this field testing (Table 6). Validity values for each scale showed 31% (four scales) were within the ideal range of average inter-item correlation (0.24 - 0.40), thirty-one percent (four scales) were above the ideal range (0.58 - 0.63), and none of the scales were below the ideal range (Table 6).

Scale scores were further divided based on age cohort (<11, 11-18, > 18 years old) (Table 6). The lowest mean scale score was on the cleft lip scar scale for both the <11 y.o. and >18 y.o. groups (53.9 and 50.4 respectively), while the 11-18 y.o. group had the lowest score on the nostril scale (53.1). This demonstrates different primary concerns with facial appearance among age groups. Participants younger than ten years old demonstrated a slight difference in physical and functional scale scores (62.8 and 70.5, respectively). Patients in the 11–18 y.o cohort demonstrated lower scores in the physical appearance scale relative to functional scale (57.6 vs. 79.2). Additionally, the >18 y.o age group showed slight differences between physical and functional scale scores (57.8 vs 67.3).

Discussion

This study aims to develop a conceptually and cross-culturally equivalent Indonesian translation of CLEFT-Q© as a standardized tool for evaluating the quality of life in CL and/or P patients. We followed the CLEFT-Q© translation protocol, based on ISPOR guidelines for the validation study, utilizing

Table 5. Internal Consistency (Cronbach's α) and Validity (Inter-Item Correlation) for Each CLEFT-Q© Scale from Cognitive Debriefing.

Scales	Total Items	Cronbach's α	Average Inter-item Total Correlation
Face	9	0.27	0.02
Lips	9	0.70	0.22
Nose	12	0.89	0.42
Nostrils	6	0.90	0.65
Jaw	7	0.90	0.54
Teeth	8	0.90	0.59
Cleft Lip Scar	7	0.78	0.24
Psychological Function	10	0.93	0.66
Social Function	10	0.54	0.16
School Function	10	0.68	0.23
Speech Function	12	0.96	0.67
Speech Distress	10	0.96	0.63
Eating and Drinking	9	0.53	−0.03

Table 6. Internal Consistency (Cronbach's α) and Validity (Inter-Item Correlation) for Each CLEFT-Q© Scale from Field-Testing.

Scales	Total Items	Cronbach's α	Average Inter-item Total Correlation
Face	9	0.81	0.33
Lips	9	0.74	0.24
Nose	12	0.92	0.48
Nostrils	6	0.90	0.60
Jaw*	7	0.90	0.58
Teeth	8	0.87	0.46
Cleft Lip Scar	7	0.92	0.63
Psychological Function	10	0.90	0.50
Social Function	10	0.86	0.40
School Function**	10	0.86	0.39
Speech Function	12	0.89	0.41
Speech Distress	10	0.87	0.40
Eating and Drinking	9	0.92	0.58

* completed by participants aged 12 to 29 years old.
** completed by participants aged 8 to 18 years old.

two forward translators and one back translator, reconciliation, cognitive debriefing, and preliminary field-testing. In its initial application, our final version of the translated questionnaire demonstrated acceptable to excellent internal consistency and 69% of the scales were within the ideal range of average inter-item correlation.

The main challenges encountered during the translation process were limited vocabulary in the target language and cultural differences which resulted in discrepancies in forward and backward translations. Literal translation was not applicable for some words. Therefore, word choices between the two forward translators were sometimes slightly different, yet arrived at the same intended meaning. This led to discrepancies between the back-translation and the intended meaning in the source language. The discrepancies were reviewed and resolved after reconciliation between the project manager and research team, resulting in the Version 2 translation which was used in cognitive debriefing.

Cognitive debriefing is an essential component of any tool development effort to ensure comprehensibility and relevancy to the target population.[19] The primary issues identified during cognitive debriefing in this study included unfamiliar vocabulary, especially in younger patients. Misinterpretation of questions across several scales was attributed to limited vocabulary among younger participants. Furthermore, some questions had not previously been considered by patients, such as "how do you like the shape of your jaw?". This was especially relevant for younger patients, perhaps due to less awareness or preoccupation with their physical differences[20] as demonstrated by our initial findings. Interpretation of items was aided by image references included in the questionnaire, ensuring similar and accurate understanding. Measurement of internal consistency is critical in tool development to ensure that the items that propose to measure the same general construct produce similar scores.[21] During cognitive debriefing, there were four scales with unacceptable to questionable internal consistency. Low α levels can represent poor reliability, suggesting that some items are not representative of the domain or behavior of interest. Other factors that may contribute to a low α are survey fatigue, if the respondent is not consistently answering questions; or difficulty comprehending questions, leading to disparate responses to similar concepts. Through thorough cognitive debriefing supervised by the project manager, participants had the opportunity to question statements they were not familiar with or did not find relevant. Thus, difficulties in comprehension can be appreciated even with a small sample size as long as each age group is represented in the sample.

Field-testing of our final version of the translation allowed for reliability and validity testing, both which are critical to demonstrate effective translation of the tool. The items that were revised following cognitive debriefing demonstrated improvement of internal consistency and validity in the field testing. All scales were deemed to have an acceptable or excellent degree of internal consistency. There was also improvement in the validity of the tool as the number of scales with the ideal range of inter-item correlation increased from three to four following the previous cognitive debriefing. Furthermore, the scales with an above average inter-item correlation score decreased from six to four on the final version, demonstrating the lack of redundancy among the scales in the questionnaire. No scale had an inter-item correlation value below the ideal range on the final version compared to two scales prior to cognitive debriefing. This can be interpreted as a success of the iterative process of translation, resulting in a more intelligible and culturally acceptable questionnaire.

Responses by Age Group

Field-testing revealed participants younger than 11 years old had good functional outcomes and rated their appearance more positively. In contrast, those in the 11–18-year-old

Table 7. Field-Testing Scale Scores by age Group.

	<11 y.o.				11-18 y.o.				>18 y.o.			
Scale	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD
Face	56	100	71	14.4	34	100	61.7	17.2	52	76	63.1	6.21
Lips	48	100	68.1	17.3	46	74	57.7	8.52	51	74	58.7	7.35
Nose	41	100	66.3	18.5	38	75	54.6	11.5	45	100	61.3	15.9
Nostrils	32	100	61.1	21.6	36	69	53.1	13.4	0	100	51.4	21.1
Jaw*					49	100	62.9	15.9	29	100	69.4	13.0
Teeth	37	83	56.4	16.1	40	90	54.7	21.1	31	77	50.9	16.6
Cleft lip scar	30	75	53.9	16.0	21	100	59.1	17.3	0	91	50.4	16.7
Average appearance score			62.8				57.7				57.8	
Psychological Function	63	100	78	11.7	54	100	88.1	19.1	49	100	74.2	13.6
School Function**	56	100	68.4	15.2	56	100	84.3	19.4				
Social Function	45	84	64.9	12.3	54	100	78.6	24.5	0	100	67.7	20.5
Speech Distress	52	100	73.1	15.7	19	100	70.8	17.4	0	100	65.2	23.5
Speech Function	47	85	68.3	13.8	53	100	74.4	9.06	0	100	62.4	6.59
Average function score			70.5				79.2				67.4	
Eating and Drinking***	13	36	29.7	7.67	11	36	26.8	15.3	14	36	30.6	20.5

* completed by participants aged 12 to 29 years old.
** completed by participants aged 8 to 18 years old.
*** total score is ranging from 9-36.

age group were very conscious about their physical outcomes. This was reflected by their relatively lower mean appearance-related score (57.7) compared to the functional outcomes score (79.2). The >18 y.o group showed a narrower gap between physical and functional outcomes (57.8 and 67.4, respectively).

Our preliminary results reveal that children less than eleven years old were the least bothered by their physical appearance (Table 7). By the age of 6 to 11, children begin to develop mastery of self and others, including self-esteem, moral development, and peer relationships, and they start to compare themselves to others.[15] This may cause the participants to rate themselves more positively. Our results were in line with those of *Shaffer and Kipp*, who found that during this period, self-esteem increases based on the child's self-perceived competence, peer group status, and through identification with primary caregivers, teachers, and other youth.[22] Furthermore, children's need for self-control increases as they engage in purposeful actions in larger groups, such as formal schooling.[15] This considered, some children in the <11 y.o. age group were reportedly worried that their "different" physical appearance and function might affect how they are perceived by their peers. Some patients reported difficulties making friends due to a speech impediment or facial differences; however, all subjects in this age cohort denied being treated differently by teachers due to their facial difference.

On the other hand, the 11-to-18-year-old stage is critical in the formulation of self-identity, as adolescents become more self-aware and self-conscious.[15] As a consequence, fluctuations in their self-esteem, emotional regulation, and overall identity formulation occur.[15] This line of thinking is consistent with our data, as most participants in the 11-to-18-year-old age group were distressed by their physical differences and overall appearance. A theme that repeatedly emerged was fear of their appearance interfering with normal peer relationships and social life. This fear often resulted in the decision to refrain from socialization and engage in self-isolation.

Participants aged older than 18 years were similarly concerned with their physical appearance, primarily expressing dissatisfaction with their cleft lip scar. Additionally, speech related concerns frequently arose in this age group. Interestingly, despite this persistent dissatisfaction, not all patients expressed interest in revisions or further interventions to address the respective issue. This finding is contrary to previous studies by *Broder* et al., *Ranganathan* et al., and *Klassen* et al. which found patients with lower health-related quality of life endorsed a desire for revision surgery.[9,23,24] Our finding may be attributable to the fact that many patients were in college (35%) and had stable careers, thereby finding fulfilment in other aspects of their lives which minimized the relative role of their physical dissatisfaction. Our findings are in concert with patient's anecdotal statements suggesting an embracement or acceptance of their condition, with increased prioritization of other elements of their day-to-day lives.

Overall, most participants in our field testing tended towards negative self-evaluation, with relatively low scores on appearance and functional scales. The reasoning for this is likely multifactorial, with personal traits, family environment, cultural milieu, and social experiences all playing a role in self-perception. *Schmitt* et al. administered the Rosenberg Self-Esteem scale to 16 998 participants across 53 nations and reported that people from East Asia scored lower on self-report measures of global self-esteem compared to Westerners.[25] Their outcomes revealed a similarly positive self-perception between regions, but less tendency towards a positive external expression of confidence among Eastern

individuals[23] which they attribute to key differences in collectivist versus individualistic cultures.

Global self-esteem scales measure two related but conceptually distinct aspects of self-worth: cognitively based self-evaluations and affectively based feelings of self-regard.[26] Self-evaluation refers to judgment about one's competencies, talents, and attributes (eg, I am smart), while self-regard refers to how people feel about themselves (eg, I am not proud of myself).[27] *Cai et al*. tried to determine whether the origins of this difference are primarily cognitive or affective.[27] For example, Chinese participants are inclined to appraise themselves less favorably than Americans on a cognitive measure of self-evaluation. On the other hand, this pattern did not emerge on an affective measure of self-regard. Overall, the Chinese participants received higher scores on modesty measures than American participants, suggesting cultural norms of modesty that temper cognitive self-evaluations. Given Indonesia's situation within the broader Eastern culture, the theme of increased modesty in participant self-evaluation likely impacted participants' responses, deflating them relative to those of their Western counterparts due to a tendency towards modest self-evaluation.

A potential limitation of this study was the use of the CLEFT-Q© Likert-type scale. For physical outcome scales, the response options were "not at all", "a little bit", "quite a bit", and "very much". For the functional outcome scales, the response options were "never", "sometimes", "often", and "always". Previous research has identified the influence of social desirability bias in studies that employ Likert and Likert-type scales.[28] Social desirability bias is the desire of the sample population to respond to questions in a way that will be interpreted favorably by others, under-reporting unsatisfactory statements and over-reporting more acceptable statements. However, the CLEFT-Q© has undergone robust psychometric validation and reliability testing, and we expect the effect of social desirability bias to be negligible, even when it is present.[8,29,30]

Amongst the obstacles to completion of this investigation was the COVID-19 pandemic, which precluded our ability to arrange in-person meetings with many participants. Although meetings could be arranged online, unstable internet connections and low technological literacy in some regions pose an ongoing barrier to integration of this approach. However, we successfully completed all interviews to obtain the necessary data from each participant. In addition, our use of five participants for cognitive debriefing was consistent with ISPOR recommendations, with each representing a different age group. Age variations of these participants allowed us to glean insights from each age group and their interpretation of the measure, despite their relatively small number.

Conclusion

PROMs are an essential metric of surgical success and a key variable to guide future interventions for patients with CL and/or P. The data obtained from PROMs can assist clinicians in learning about and improving the patient experience. Additionally, PROMs allow for a universal method of outcome reporting, opening more doors to cross-cultural collaboration between cleft care teams worldwide. This study demonstrates the successful translation and cultural adaptation of CLEFT-Q© in a new language and cultural setting with adherence to established guidelines and support of the CLEFT-Q© team.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is funded by Directorate of Research and Development, Universitas Indonesia under Hibah PUTI 2023 (Grant No. NKB-1251/UN.2RST/HKP.05.00/2022)".

ORCID iDs

Prasetyanugraheni Kreshanti https://orcid.org/0000-0003-4468-5788
Jordan W. Swanson https://orcid.org/0000-0001-7112-8114
Kristaninta Bangun https://orcid.org/0000-0002-6312-9261

References

1. Salari N, Darvishi N, Heydari M, Bokaee S, Darvishi F, Mohammadi M. Global prevalence of cleft palate, cleft lip and cleft palate and lip: A comprehensive systematic review and meta-analysis. *J Stomatol Oral Maxillofac Surg*. 2022;123(2):110-120. doi:10.1016/j.jormas.2021.05.008
2. , Institute for Health Metrics Evaluation. Data from: Global Burden of Disease. *Orofacial clefts*. *USA*.
3. Massenburg BB, Hopper RA, Crowe CS, et al. Global burden of orofacial clefts and the world surgical workforce. *Plast Reconstr Surg*. 2021;148(4):568e-580e. doi:10.1097/prs.0000000000008334
4. Kummer AW. A pediatrician's guide to communication disorders secondary to cleft lip/palate. *Pediatr Clin North Am*. 2018;65(1):31-46. doi:10.1016/j.pcl.2017.08.019
5. McIntyre JK, Sethi H, Schönbrunner A, Proudfoot J, Jones M, Gosman A. Number of surgical procedures for patients with cleft lip and palate from birth to 21 years old at a single children's hospital. *Ann Plast Surg*. 2016;76(Suppl 3):S205-S208. doi:10.1097/sap.0000000000000765
6. Campbell A, Kreshanti P. Comprehensive cleft care centers: scalable, sustainable, and cost effective surgical care. In: Swanson JW, ed. *Global cleft care in low-resource settings*. Springer International Publishing; 2021:517-529.
7. Eckstein DA, Wu RL, Akinbiyi T, Silver L, Taub PJ. Measuring quality of life in cleft lip and palate patients: Currently available patient-reported outcomes measures. *Plast Reconstr Surg*. 2011;128(5):518e-526e.
8. Tsangaris E, Wong Riff KWY, Goodacre T, et al. Establishing content validity of the CLEFT-Q: A new patient-reported outcome instrument for cleft lip/palate. *Plastic and Reconstructive Surgery – Global Open*. 2017;5(4):e1305.

9. Klassen AF, Riff KWW, Longmire NM, et al. Psychometric findings and normative values for the CLEFT-Q based on 2434 children and young adult patients with cleft lip and/or palate from 12 countries. *CMAJ*. 2018;190(15):E455-E462. doi:10.1503/cmaj.170289
10. Klassen AF, Wong Riff KWY. CLEFT-Q User's Guide. McMaster University; 2018:9.
11. , U. S. Department of Health Human Services. Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims: Draft guidance. *Health Qual Life Outcomes*. 2006;4(1):79. doi:10.1186/1477-7525-4-79
12. Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*. 2002;11(3):193-205. doi:10.1023/a:1015291021312
13. 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 2—assessing respondent understanding. *Value Health*. 2011;14(8):978-988. doi:10.1016/j.jval.2011.06.013
14. 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 in Health : The Journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2011;14(8):967-977. doi:10.1016/j.jval.2011.06.014
15. Guerra N, Williamson A, Lucas-Molina B. Normal development infancy, childhood, and adolescence. In: Rey J ed. ,*International handbook of child and adolescent mental health*. International Association for Child and Adolescent Psychiatry and Allied Professions; 2012:6.
16. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value in Health : The Journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2005;8(2):94-104. doi:10.1111/j.1524-4733.2005.04054.x
17. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res Sci Educ*. 2018;48(6):1273-1296. doi:10.1007/s11165-016-9602-2
18. Piedmont R, Hyland M. Inter-Item correlation frequency distribution analysis: A method for evaluating scale dimensionality. *Educational and Psychological Measurement - EDUC PSYCHOL MEAS*. 1993;53:369-378. doi:10.1177/0013164493053002006
19. Wong Riff KWY, Tsangaris E, Forrest CR, et al. CLEFT-Q: Detecting differences in outcomes among 2434 patients with varying cleft types. *Plast Reconstr Surg*. 2019;144(1):78e-88e. doi:10.1097/prs.0000000000005723
20. Wong Riff KWY, Tsangaris E, Goodacre TEE, et al. What matters to patients with cleft lip and/or palate: An international qualitative study informing the development of the CLEFT-Q. *Cleft Palate Craniofac J*. 2018;55(3):442-450. doi:10.1177/1055665617732854. 10.1177/1055665617732854</p>.
21. Sullivan GM. A primer on the validity of assessment instruments. *J Grad Med Educ*. 2011;3(2):119-120. doi:10.4300/jgme-d-11-00075.1
22. Shaffer D, Kipp K. *Development Psychology: Childhood and Adolescence*. 2009.
23. Broder HL, Wilson-Genderson M, Sischo L. Oral health-related quality of life in youth receiving cleft-related surgery: Self-report and proxy ratings. *Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*. 2017;26(4):859-867. doi:10.1007/s11136-016-1420-5
24. Ranganathan K, Shapiro D, Aliu O, et al. Health-Related quality of life and the desire for revision surgery among children with cleft lip and palate. *J Craniofac Surg*. 2016;27(7):1689-1693. doi:10.1097/scs.0000000000002924
25. Schmitt DP, Allik J. Simultaneous administration of the Rosenberg self-esteem scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *J Pers Soc Psychol*. 2005;89(4):623-642. doi:10.1037/0022-3514.89.4.623
26. Tafarodi RW, Swann WBJr. Self-liking and self-competence as dimensions of global self-esteem: Initial validation of a measure. *J Pers Assess*. 1995;65(2):322-342. doi:10.1207/s15327752jpa6502_8
27. Cai H, Brown JD, Deng C, Oakes MA. Self-esteem and culture: Differences in cognitive self-evaluations or affective self-regard? *Asian J Soc Psychol*. 2007;10(3):162-170. doi:10.1111/j.1467-839X.2007.00222.x
28. Garland R. The Mid-Point on a Rating Scale: Is it Desirable? 1991.
29. Miroshnychenko A, Rae C, Riff KW, et al. Psychometric validation of the CLEFT-Q patient reported outcome measure: A prospective study to examine cross-sectional construct validity. *Cleft Palate Craniofac J*. 2021:10556656211062837. doi:10.1177/10556656211062837
30. Wong Riff KW, Tsangaris E, Goodacre T, et al. International multiphase mixed methods study protocol to develop a cross-cultural patient-reported outcome instrument for children and young adults with cleft lip and/or palate (CLEFT-Q). *BMJ Open*. 2017;7(1):e015467. doi:10.1136/bmjopen-2016-015467