Protocol

GENDER-Q Youth Patient-Reported Outcome Measure: Protocol for an International Cross-Sectional Field Test Study

Anne F Klassen[1], DPhil; Charlene Rae[1], PhD; Shane D Morrison[2], MS, MD; Natasha Johnson[1], MD; Karine Khatchadourian[3], MSc, MD; Jack Graesser[4], BS; Manraj N Kaur[5], PhD

[1]Department of Pediatrics, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[2]Department of Surgery, Division of Plastic Surgery, University of Washington Medical Center, Seattle, WA, United States

[3]Department of Pediatrics, University of Ottawa, Ottawa, ON, Canada

[4]Department of Clinical Research, University of Southern Denmark, Odense, South Denmark, Denmark

[5]Department of Surgery, Harvard Medical School, Patient-Reported Outcomes, Values & Experience Center (PROVE), Mass General Brigham, Boston, MA, United States

Corresponding Author:

Anne F Klassen, DPhil
Department of Pediatrics
Faculty of Health Sciences, McMaster University
3N27, 1280 Main Street West
Hamilton, ON L8N3Z5
Canada
Phone: 1 905-521-2100 ext 73775
Email: aklass@mcmaster.ca

Abstract

Background: The need for gender-affirming care (GAC) for young people has increased rapidly worldwide. GAC includes social, psychological, and medical interventions aimed at alleviating gender-related distress. Given that many outcomes of GAC relate to how young people function and feel, a rigorously developed patient-reported outcome measure (PROM) is needed. To address this need, GENDER-Q Youth was developed with extensive input from youth with lived experience. Concept elicitation interviews were performed with 47 youth from Canada and the United States. A conceptual framework and draft PROM were developed and refined with feedback from 33 experts and 17 youth, and were pilot tested with 406 older youth. The field test version of GENDER-Q Youth includes 16 independently functioning scales (248 items) that measure health-related quality of life, gender practices, voice, and experience of care.

Objective: This paper outlines the protocol for an international cross-sectional study to field test GENDER-Q Youth in a sample of transgender and gender diverse (TGD) adolescents and young adults aged 12 to 25 years receiving GAC.

Methods: This study follows international guidelines for PROM development and validation (eg, US Food and Drug Administration and Consensus-Based Standards for the Selection of Health Measurement Instruments [COSMIN]). GENDER-Q Youth was translated into Danish, Dutch, and German following the Professional Society for Health Economics and Outcomes Research (ISPOR) guidelines for the translation and cultural adaptation of PROMs. For the field test, TGD youth are being recruited using multiple strategies (eg, face-to-face, emails, and so on). A REDCap survey is used to collect sociodemographic and clinical data. Branching logic is used to ensure relevant GENDER-Q Youth scales are completed as some scales are gender-specific. For psychometric analysis, Rasch measurement theory (RMT) analysis will be used to examine fit of the observed data to the Rasch model. A series of tests and criteria will examine item fit and scale reliability and validity. Test-retest reliability will be examined with intraclass correlation coefficients. For construct validity, 147 predefined hypotheses of expected group differences and correlations between GENDER-Q Youth scales will be examined. Acceptance of at least 75% of hypotheses is considered sufficient evidence of construct validity per COSMIN criteria.

Results: The research described in this protocol is funded by a Canadian Institutes of Health Research Sex and Gender Science Chair (April 2020) and supported by a Canada Research Chair in Patient-Reported Outcomes (April 2025). With the Danish, Dutch, and German translations now completed, the international field test is underway in Canada, Europe, Australia, and the United States. Data collection and analysis are expected to be completed by the end of 2026.

Conclusions: This protocol describes the international GENDER-Q Youth field test study. Once developed, GENDER-Q Youth can be used to inform GAC, research, and quality improvement efforts.

International Registered Report Identifier (IRRID): DERR1-10.2196/100420

JMIR Res Protoc 2026;15:e100420; doi: 10.2196/100420

Keywords: gender-affirming care; youth; hormones; mixed methods; patient-reported outcomes; protocol

Introduction

Background

The global need for gender-affirming care (GAC) for young people has increased substantially, especially among those assigned female at birth [1]. GAC encompasses social, psychological, and medical interventions that aim to alleviate gender-related distress. Individual treatment goals vary but can include treatments that modify how the voice sounds and change how the face and body look and function. Studies have shown the benefits of GAC in terms of improved appearance congruence and mental health outcomes with low levels of regret [2-5]. Concurrently, systematic reviews of gender-affirming treatments for youth (eg, puberty blockers, hormones, and mental health care) suggest that long-term outcomes and overall quality of evidence remain limited [6-12].

Since many outcomes associated with GAC affect how people function and feel, rigorously developed patient-reported outcome measures (PROMs) specific to gender-diverse populations are needed. This need is underscored by 6 systematic reviews of PROMs, which converge to conclude that PROMs used in GAC are typically generic (for use in any health condition), designed for cisgender populations, and designed without adherence to international PROM development guidelines [13-18]. One review to date examined PROMs used with transgender and gender diverse (TGD) youth [18] identifying 38 measures that had been used in research with TGD youth. Most PROMs are generic measures of psychological function or quality of life. Such tools were not developed to measure outcomes of GAC and therefore lack content validity for TGD youth. Three gender-specific PROMs that measure gender-specific concepts were found, but published evidence about how these PROMs were developed and validated is lacking. The use of such PROMs in research with TGD youth can result in inaccurate or misleading conclusions by failing to ask the right questions [19]. Also missing are validated scales that measure how TGD youth experience their care [20]. Based on current literature, there is currently no rigorously developed comprehensive PROM that can be used to measure outcomes and experience of care for TGD youth.

To address the lack of validated GAC-specific PROMs, our team developed the GENDER-Q for adults. This comprehensive PROM includes 54 unidimensional scales and 2 checklists covering domains of health-related quality of life, sexual, urination, gender practices, voice, hair, face and neck, body, breasts, genital feminization, chest, genital masculinization, and experience of care [21,22]. GENDER-Q was designed to evaluate any form of GAC for adults, including all forms of gender-affirming surgery. To address the need for a PROM for TGD youth, our focus was to design a comprehensive set of scales to evaluate the multidisciplinary care provided in pediatric gender clinics. As such, GENDER-Q Youth does not include scales to measure surgical outcomes but instead focuses on measurement of outcomes related to social, psychological, voice, and medical treatments (eg, puberty blockers, hormones) provided to TGD youth by pediatricians and other healthcare providers.

Our study adheres to international guidelines for PROM development and validation, including the US Food and Drug Administration, the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN), and ISPOR (The Professional Society for Health Economics and Outcomes Research) Principles of Good Practice for the Translation and Cultural Adaptation of PROMs [23-28]. Figure 1 shows our mixed method approach, which is grounded in the experiences of TGD youth. Our team has completed step 1 of a mixed method study to identify the concepts that GENDER-Q Youth should measure. Full details for the sample, methods, and results for step 1 are published [29]. Briefly, we conducted concept elicitation interviews with 47 youth aged 12 to 19 years from Canada and the United States who were seeking or receiving GAC. Interviews were digitally recorded, transcribed, and coded. Analysis led to the development of a conceptual framework (Figure 2) and an item pool. The items were used to form a set of scales that were refined with input from 33 experts and 17 TGD youth. Content validity was further explored in an online sample of 120 youth aged 18 to 25 years.

Figure 1. Methods diagram. CTT: classical test theory; RMT: Rasch measurement theory.

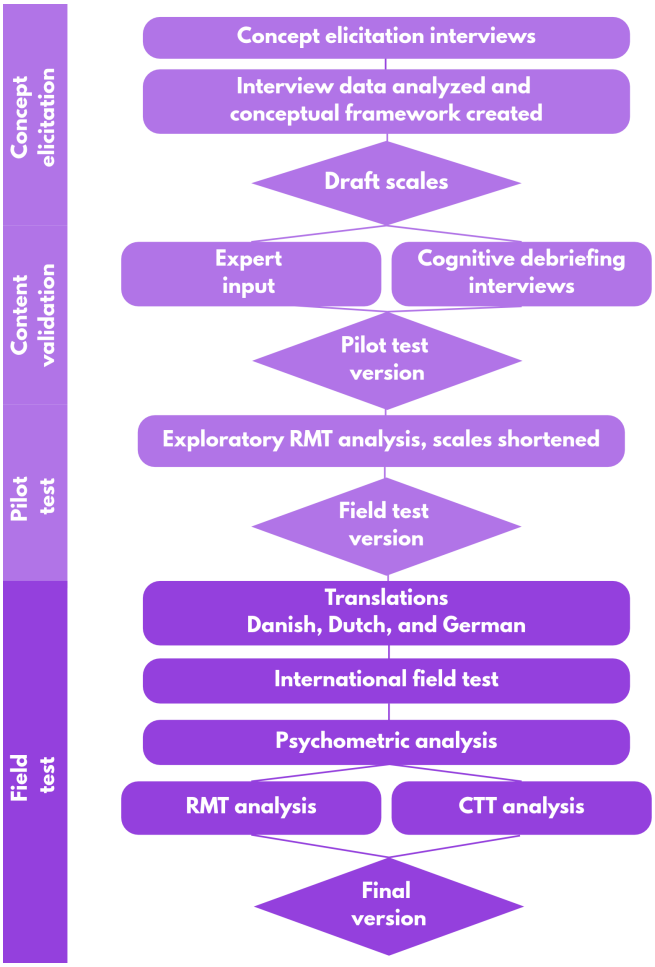
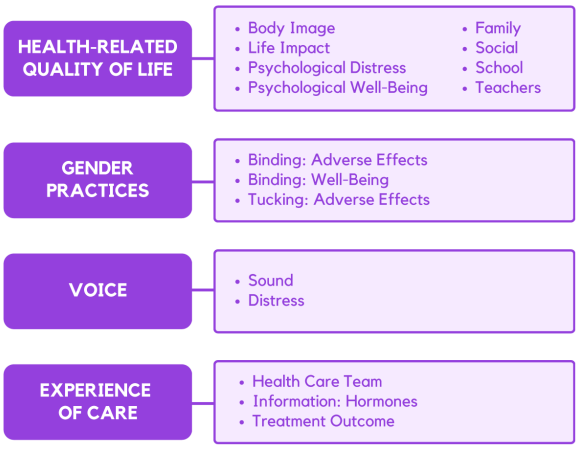


Figure 2. GENDER-Q Youth conceptual framework.



GENDER-Q Youth was pilot tested in an international sample of 406 youth aged 18 to 25 years recruited from the Prolific crowd-sourcing platform. The sample was also used to collect test-retest (TRT) data for GENDER-Q Youth scales. For psychometric analysis, Rasch measurement theory (RMT) analysis [30-33] was used to identify items with extreme misfit to the Rasch model (11 items were dropped from 7 scales). At the end of step 1, the field test version of GENDER-Q Youth included 16 scales (248 items). Concepts used to develop the items for each scale and example quotes to illustrate participants' experiences are published elsewhere [29]. The Flesch-Kincaid Grade Level for GENDER-Q Youth field test scales ranged from 0.5 to 5.0.

Objective

The objective of this study is to field test GENDER-Q Youth in a large international sample of TGD youth aged 12 to 25 years seeking or receiving GAC and to evaluate its psychometric performance in order to provide a comprehensive PROM for use in research, clinical care, and quality improvement.

This objective will be achieved by examining a range of measurement properties outlined by the COSMIN initiative [27,28]. To shorten each scale, we will conduct RMT analysis, a modern psychometric method that uses a probabilistic measurement model to identify the optimal set of items to retain in each scale. In this approach, selected items should measure the full range of the underlying construct with high reliability and validity. A series of statistical tests (eg, ordered thresholds, item fit, local dependency, and differential item functioning) will guide decisions about which items to retain. The resulting scales will be unidimensional, invariant, and easy to interpret. Additional classical test theory (CTT) tests will be performed to examine scale stability (ie, test-retest reliability) and construct validity by testing predefined hypotheses for group differences and correlations between GENDER-Q Youth scales.

Methods

Ethical Considerations

Research ethical approval has been obtained for the McMaster University coordinating site (Hamilton Integrated Research Ethics Board #17427). Ethical approvals and data sharing agreements will be obtained for all participating sites prior to data collection. In Canada, youth under the age of 18 years must complete an assent form and have a parent or primary caregiver provide consent for their participation in the study unless a clinical team member has deemed the individual capable of consenting for themselves. In all other countries, participants younger than 18 years of age must have a parent or primary caregiver provide consent for their participation unless a waiver of parental consent is obtained from their local ethics board. All participants aged 18 years or older may provide consent for themselves. Participant consent in the research study will be obtained in the REDCap survey (Vanderbilt University). The survey is created and hosted at McMaster University or a participating site's server as per each institution's requirements. In the reporting of results, the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) reporting guidelines for cross-sectional studies will be followed [34].

At the start and end of the survey, participants are provided with a list of local resources they can access for support if needed (eg, Trans Lifeline Canada and Kids Help Phone). The survey is designed such that participants are required to answer a limited number of questions needed for branching logic (eg, sex assigned at birth) to ensure they receive the appropriate items and scales (eg, scales measuring binding vs tucking). In addition, demographic and clinical questions include "Prefer to not answer" or "I want to skip this question" as response options. The rest of the survey is set up such that participants can skip any items or scales they are not comfortable answering. At the end of the survey, to measure the impact of the GENDER-Q Youth survey on participants, we ask how answering the survey made them feel (ie, "Very unhappy," "A little unhappy," "Same as usual," "A little happy," and "Very happy"). An open text box is provided for comments. The impact of completing GENDER-Q Youth in relation to demographic and clinical characteristics will be examined in the analysis.

Translations

In preparation for the international field test, GENDER-Q Youth has been translated into Danish, Dutch, and German following the Professional Society for Health Economics and Outcomes Research (ISPOR) guidelines for PROM translation and cultural adaptation [26]. This approach involved two independent forward translations, reconciliation, back-translation into English, and cognitive debriefing interviews with at least 5 youth aged 12 to 25 years who represented different gender identities. Youth were asked to review the translated instructions, response options, and items to determine if they were understandable and culturally relevant. The goal was to develop conceptual and culturally equivalent translations of GENDER-Q Youth. Each translation team also translated the site-specific consent letter and REDCap survey (eg, demographic and clinical questions).

Sample

The field test sample includes youth recruited in multiple countries. Eligible TGD youth are aged 12 to 25 years seeking or receiving GAC who are able to complete the survey independently in one of the GENDER-Q Youth translations. The upper age range is used since pediatric hospitals and gender clinics in some countries provide GAC to the young adult population.

Recruitment

Participant recruitment uses a number of strategies and approaches to accrue the sample. In hospitals and clinics (gender, endocrinology, voice therapy, and/or surgery), participants are recruited in-person, virtually (eg, Zoom), by email or phone call, or through electronic medical record messaging (eg, MyChart). Youth who completed a concept elicitation interview in the GENDER-Q Youth step 1 study [29], and eligible participants from the GENDER-Q adult field test study who agreed to be recontacted [21], are invited to complete the field test survey. In the community, TGD youth aged 18 years and older are recruited via email and social media posts through 2SLGBTQI+ (Two-Spirit, lesbian, gay, bisexual, transgender, queer or questioning, and intersex) community groups, advocacy groups, and professional/clinical services organizations.

Data Collection

An online REDCap questionnaire is used to collect self-reported demographics (eg, age, gender identity, race, education, and country), health status (eg, health conditions including neurodivergent diagnoses), and health care utilization variables including number of visits to the hospital or clinic for GAC; treatments participants have had or want to have in the future such as puberty blockers, hormones, voice or speech therapy, and mental health care and surgery. For gender identity, participants are first provided with an open text box with the question: "What word or words would you use to describe your gender?" This question is followed with the fixed choice question: "If you had to pick one of the following, would you say that you are" with answer options: "Boy or Male," "Girl or Female," "Nonbinary," and "Another gender." Participants that choose "Another gender" receive an open text box that asks: "Can you please tell us what other gender you identify as?" Participants are also asked the following question: "Do you identify as transgender?" Answer options include the following: "No," "Yes," and "Not sure." After the gender questions, sex at birth is asked and used with branching logic to ensure that participants receive the correct gender practices scales (ie, for tucking or binding). This question is worded as follows: "What was your assigned sex at birth?" The options include the following: "Male," "Female," and "Not sure." Those who select not sure are asked: "Are you intersex?." Answers include: "No," "Yes," and "Not Sure." Additional questions were asked to measure construct validity for GENDER-Q Youth scales. An example is the question "How happy are you with the people you see in the clinic (eg, doctor, nurse, social worker)?" Answers could be "Not happy," "A little happy," "Quite happy," and "Very happy." This question will be used to examine group differences for the hypothesis that scores would be higher on the GENDER-Q Youth Health Care Team scale for those who reported being happier.

The REDCap survey is hosted on McMaster University's secure server or at collaborating sites if required by their institution. Youth in clinics have the option of completing the REDCap survey using an iPad provided in the clinic or their own device. Youth who are emailed a link to the survey can use their own device. Participants who complete the survey receive a US $18 gift card equivalent to thank them for their time if allowed by their local ethics board.

Data Analysis

Sociodemographic and clinical characteristics for the study sample, and the question asking about the impact of completing the GENDER-Q Youth will be summarized using descriptive statistics. Data from the pilot test will be combined with the field test data for the following psychometric analyses.

RMT Analysis

Overview

RMT analysis is an established modern psychometric method used in rating scale design and psychometric evaluation [30-33]. This approach uses a probabilistic framework to model the relationship between a person's latent trait level and item difficulty. Rasch analysis orders items hierarchically from hardest to easiest to endorse ("Rasch ruler"). When observed data fit the expectations of the Rasch model, the resulting scale supports invariant and interpretable interval-level measurement. RMT analysis will be used to examine how items function together within each scale and to identify redundant or poorly functioning items. Our goal is to identify a subset of items for each scale that best represents the underlying construct, resulting in shorter, clinically meaningful scales with strong measurement properties. Analyses will be conducted using RUMM2030 software (RUMM Laboratory) with the unrestricted partial credit model for polytomous data. We will examine the following item and scale properties described below.

Thresholds for Item Response Options

Category thresholds between adjacent response options (eg, "Never" and "Rarely") will be examined to determine whether response categories function as they are intended. Ordered thresholds indicate whether participants with higher levels of the construct are more likely to endorse higher response categories. Disordered thresholds indicate that respondents may have trouble distinguishing between adjacent response categories. Items with disordered thresholds can be collapsed (rescored) to improve ordering or removed from the scale.

Item Fit Statistics

Three fit indicators will be examined: fit residuals, item characteristic curves, and chi-square statistics. Fit residuals are expected to fall within the range of –2.5 to +2.5, chi-square values should be nonsignificant after Bonferroni adjustment, and intraclass correlation curves should depict agreement between observed and expected responses. We will interpret fit statistics together and in relation to their clinical importance.

Dependency

We will examine residual correlations between item pairs to assess local item dependence. Residual correlations greater than 0.30 above the average residual correlation will indicate potential dependency, which can artificially inflate reliability [35]. For item pairs showing dependency, subtests will be created to evaluate their impact on scale reliability.

Targeting

We will examine the distribution of item locations relative to person locations to assess targeting. Good targeting is indicated when item thresholds are spread across the range of the construct and align with the distribution of participants. Poor targeting may result in floor and/or ceiling effects.

Differential Item Functioning

We will examine differential item functioning (DIF) to determine if subgroups (ie, gender, age, country, and language) respond differently to items despite having the same level of underlying trait. For gender, DIF will be examined by comparing three groups as follows: boy/male, girl/female, and nonbinary. Where subgroup sizes are unequal, random samples will be selected to create equal-sized groups. Items with chi-square values that are significant will be considered for deletion during item reduction. We will also examine the impact of DIF on each scale by splitting items on the variable demonstrating DIF, and correlating the person locations from the original and split solutions.

Reliability

We will compute the Person Separation Index for each scale to assess the reliability of each scale and the ability to distinguish between individuals of different levels of the trait for each scale. We will also compute Cronbach α to examine the internal consistency of items within each scale [36]. Reliability coefficients of ≥0.70 are considered acceptable [28].

Unidimensionality

The Rasch model requires that scales measure a single unidimensional construct [33]. This assumption can be tested using the Smith test [37]. In this test, items are split into 2 subsets based on a principal component analysis of the residuals and person estimates are compared. If fewer than 5% of tests show significantly different estimates between the 2 subsets, the scale is considered unidimensional [32].

CTT Analysis

After the RMT analysis is completed, the item-reduced scales will be scored by transforming person locations into scores that range from 0 to 100. For all scales, higher scores will indicate a better outcome. The following psychometric tests will be performed.

TRT Reliability

We will examine TRT reliability in stable participants using a 2-way mixed effect model intraclass correlation coefficient for consistency. A reliability coefficient of ≥0.70 is considered acceptable [28].

Smallest Detectable Change

TRT data will also be used to compute the smallest detectable change (SDC) beyond measurement error at the individual and group levels [28]. Change scores greater than the SDC suggest real change rather than change due to chance. For a scale to be useful, the SDC should be smaller than the minimally important difference (MID) [28]. Calculating an MID value is beyond the scope of the current study.

Construct Validity

COSMIN guidelines informed our plan for construct validity testing [27,28,38]. Construct validity examines whether an

instrument measures the construct it was designed to assess by comparing differences between groups and associations between scales. To this end, we created a set of questions to test the constructs measured by GENDER-Q Youth scales. These hypotheses examine a wide range of group differences as well as correlations. We will test hypotheses regarding group differences and correlations between GENDER-Q Youth scales. This approach to construct validity requires specification of both the direction and magnitude of predetermined hypotheses [27]. COSMIN considers construct validity adequate if at least 75% hypotheses are confirmed, although guidance on the number of hypotheses to test is lacking [28]. Multimedia Appendix 1 presents the predefined hypotheses for the GENDER-Q Youth study, with directionality specified for each hypothesis. We formulated 147 hypotheses (3-17 per scale) to test. The goal for construct validation is to examine the overall acceptance of the tests rather than focus on individual test results.

Group Hypotheses

We will test 69 group hypotheses that use demographic information, treatment information, and Likert questions included in the survey. For dichotomous variables, mean differences with 95% CIs will be computed and interpreted using Cohen *d* (0.20 small, 0.50 medium, and 0.80 large) [39]. For variables with more than 2 response options, effect sizes will be calculated using η^2 with 95% CIs and interpreted as follows: <0.01 negligible, 0.01 ≤η^2<0.06 small, 0.06 ≤η^2<0.14 medium, and ≥0.14 large [39]. All group differences are hypothesized to demonstrate at least a small effect size (d≥0.2 or η^2≥0.01).

Correlational Hypotheses

We will test 78 correlational hypotheses. Associations between scales will be examined using Pearson correlation coefficients (*r*). For correlations between GENDER-Q Youth scales, COSMIN-based criteria are applied: scales measuring similar constructs are expected to correlate *r*≥0.5; scales measuring related but dissimilar constructs are expected to correlate 0.3 to <0.5, and scales measuring unrelated constructs are expected to correlate *r*<0.3 [39]. Some scales in the GENDER-Q adult module (ie, Body Image, Gender Dysphoria, Psychological Well-being, Psychological Distress, Social, Voice Sound, Voice Distress, and Treatment Outcome) measure similar constructs as scales in the youth module. For construct validation, the magnitude of correlations between GENDER-Q Youth scales is expected to be similar to the correlations between the 8 GENDER-Q adult module scales . For GENDER-Q Youth scales without an adult counterpart, we hypothesized scales would correlate 0.30, indicating the scales measure related constructs.

Results

This study is funded by a Canadian Institutes of Health Research (CIHR) Sex and Gender Science Chair Award received in April 2020. This award provided grant funds to develop and field test GENDER-Q Youth with TGD youth receiving GAC in Canada and the United States. In April 2025, AFK was awarded a Tier 1 Canada Research Chair in Patient-Reported Outcomes. This award provided funds that made it possible to translate GENDER-Q Youth into multiple languages and expand the field test to include TGD youth in Europe, and Australia. Including youth in multiple countries makes it possible to capture a much broader range of experiences in the dataset that will be used to item-reduce each scale and develop the scoring algorithms.

Step 1, which included concept elicitation interviews, conceptual framework development, scale item generation, content validation, and pilot testing, has been completed [29,40,41]. Elsewhere we report that the sample of youth in the pilot testing found GENDER-Q Youth to be easy to understand (94%), thorough (92%), safe to complete (91%), and that it asked important questions (80%) in a respectful way (91%) [29].

In preparation for step 2, GENDER-Q Youth has been translated into Dutch, Danish, and German. Field testing is taking place at 11 participating sites in Canada (n=2), the United States (n=4), Denmark (n=1), the Netherlands (n=1), England (n=1), Germany (n=1), and Australia (n=1). We anticipate that data collection and analysis will be completed by the end of 2026 and we expect to publish results in the first half of 2027. Once GENDER-Q Youth is finalized, our team will examine psychometric properties for the item-reduced final scales as well as score interpretability.

Discussion

Principal Findings

There is a rising international need for GAC for TGD young people and for more high-quality evidence for the effectiveness of gender-affirming treatment. GENDER-Q Youth will provide a tool that researchers and clinicians can use to collect evidence-based outcomes data from youth aged 12 to 25 years who identify as TGD. In step 1, data from 47 youth were collected and analyzed to form a conceptual framework, items, and preliminary scales [29]. Content validity was demonstrated with feedback from 33 experts, and cognitive debriefing interviews with 17 youth [29]. In step 2, our international field test study will provide important psychometric evidence about the measurement properties of GENDER-Q Youth. Our study will item-reduce GENDER-Q Youth scales and provide evidence of reliability and validity. Such information is needed before GENDER-Q Youth can be used in research, clinical care, and health care policy. Validation is an ongoing process with evidence accumulating over time and across multiple studies. Future research will be needed to examine other measurement properties, such as responsiveness, and to enhance interpretability of GENDER-Q Youth scale scores by computing MIDs.

TGD people experience substantial barriers to accessing care and continue to have unmet health care needs [42]. Given the rising demand for GAC for TGD young people, and the limited evidence for the effectiveness of treatment, a robust comprehensive PROM is needed to capture the patient perspective in outcome assessment. GENDER-Q Youth will provide a rigorously developed and validated PROM that can be used in research studies and clinical care to evaluate outcomes that matter most to TGD youth. Once complete, the GENDER-Q Youth can be used in the evaluation of GAC or in clinics to compare outcomes across studies, monitor change in outcomes over time, as well as to inform shared decision-making. Gender-specific PROMs for youth are important as they can help inform evidence-based care and improve delivery of GAC.

Strengths and Limitations

This study has strengths and limitations. A key strength is the meaningful patient involvement that is taking place throughout all stages of the development and validation of GENDER-Q Youth. Such input is essential to ensuring that GENDER-Q Youth measures concepts relevant to patients in language that is appropriate and respectful. Second, the inclusion of a large heterogeneous international sample that varies by age, gender identity, country, language, and treatment history will enable the capture of a broad range of experiences. Third, the use of a modern psychometric approach (RMT analysis) and modular design means that each scale is unidimensional and independently functioning, which will allow end users to choose only the most relevant scales for their purpose or needs, thereby reducing respondent burden.

A limitation of our study is the reliance on self-reported data and online data collection. Participants are not blinded to the study's aims, which may influence their responses due to concerns about how their answers might be used. We recognize that the survey is lengthy, which may contribute to participant fatigue. To address this limitation, participants are provided with the option to save their responses and complete the study later. Another limitation includes factors outside the control of the research team that may affect timing for recruitment and data collection, such as the time taken to obtain ethics approval and data sharing agreements at collaborating sites. At the end of data collection, the sample size may be insufficient to include all subgroups in some psychometric tests (eg, DIF analyses by country or language). Finally, younger adolescents (aged 12-14 y) and those assigned male at birth may be underrepresented in the sample, which may limit the generalizability of findings to these subgroups.

Acknowledgments

Generative AI tools were not used to write the manuscript.

Funding

The research described in this protocol is funded by a Canadian Institutes of Health Research Sex and Gender Science Chair Award (FRN#: 171376) and a Tier 1 Canada Research Chair in Patient-Reported Outcomes, both awarded to AFK.

Data Availability

The data supporting this study's findings are restricted due to privacy and ethical regulations protecting participant confidentiality.

Authors' Contributions

Conceptualization: AFK, MNK
Funding acquisition: AFK
Investigation: AFK, CR, SDM, NJ, KK, JG, MNK
Methodology: AFK, CR, SDM, MNK
Supervision: AFK
Writing – original draft: AFK
Writing – review & editing: AFK, CR, SDM, NJ, KK, JG, MNK

Conflicts of Interest

AFK and MNK reported being co-developers of GENDER-Q, which is owned by McMaster University and would receive a share of license revenues as royalties for its use in for-profit research based on their institution's inventor sharing policy. AFK provides research consulting services to the pharmaceutical industry through EVENTUM Research for research that is not related to this study. KK reports participation in advisory board meetings for Tolmar. All other authors declare no conflicts of interest.

Multimedia Appendix 1

A priori construct validity hypotheses.
[DOCX File (Microsoft Word File), 48 KB-Multimedia Appendix 1]

Peer Review Report 1

Granting agency peer review report.
[PDF File (Adobe File), 57 KB-Peer Review Report 1]

References

1. Taylor J, Hall R, Langton T, Fraser L, Hewitt CE. Characteristics of children and adolescents referred to specialist gender services: a systematic review. Arch Dis Child. Oct 30, 2024;109(Suppl 2):s3-s11. [doi: 10.1136/archdischild-2023-326681] [Medline: 38594046]
2. Chen D, Berona J, Chan YM, et al. Psychosocial functioning in transgender youth after 2 years of hormones. N Engl J Med. Jan 19, 2023;388(3):240-250. [doi: 10.1056/NEJMoa2206297] [Medline: 36652355]
3. Tordoff DM, Wanta JW, Collin A, Stepney C, Inwards-Breland DJ, Ahrens K. Mental health outcomes in transgender and nonbinary youths receiving gender-affirming care. JAMA Netw Open. Feb 1, 2022;5(2):e220978. [doi: 10.1001/jamanetworkopen.2022.0978] [Medline: 35212746]
4. Olson KR, Durwood L, Horton R, Gallagher NM, Devor A. Gender identity 5 years after social transition. Pediatrics. Aug 1, 2022;150(2):e2021056082. [doi: 10.1542/peds.2021-056082] [Medline: 35505568]
5. Olson KR, Raber GF, Gallagher NM. Levels of satisfaction and regret with gender-affirming medical care in adolescence. JAMA Pediatr. Dec 1, 2024;178(12):1354-1361. [doi: 10.1001/jamapediatrics.2024.4527] [Medline: 39432272]
6. Thompson L, Sarovic D, Wilson P, et al. A PRISMA systematic review of adolescent gender dysphoria literature: 3) treatment. PLOS Glob Public Health. 2023;3(8):e0001478. [doi: 10.1371/journal.pgph.0001478] [Medline: 37552651]
7. Dopp A, Peipert A, Buss J, De Jesús-Romero R, Palmer K, Lorenzo-Luaces L. Interventions for gender dysphoria and related health problems in transgender and gender-expansive youth: a systematic review of benefits and risks to inform practice, policy, and research. Rand Health Q. 2025;12(2):2. [doi: 10.7249/rhq-12-02-02]
8. Taylor J, Mitchell A, Hall R, Langton T, Fraser L, Hewitt CE. Masculinising and feminising hormone interventions for adolescents experiencing gender dysphoria or incongruence: a systematic review. Arch Dis Child. Oct 30, 2024;109(Suppl 2):s48-s56. [doi: 10.1136/archdischild-2023-326670] [Medline: 38594053]
9. Ludvigsson JF, Adolfsson J, Höistad M, Rydelius PA, Kriström B, Landén M. A systematic review of hormone treatment for children with gender dysphoria and recommendations for research. Acta Paediatr. Nov 2023;112(11):2279-2292. [doi: 10.1111/apa.16791] [Medline: 37069492]
10. Miroshnychenko A, Ibrahim S, Roldan Y, et al. Gender affirming hormone therapy for individuals with gender dysphoria aged <26 years: a systematic review and meta-analysis. Arch Dis Child. May 16, 2025;110(6):437-445. [doi: 10.1136/archdischild-2024-327921] [Medline: 39855725]
11. Taylor J, Mitchell A, Hall R, et al. Interventions to suppress puberty in adolescents experiencing gender dysphoria or incongruence: a systematic review. Arch Dis Child. Oct 30, 2024;109(Suppl 2):s33-s47. [doi: 10.1136/archdischild-2023-326669] [Medline: 38594047]
12. Heathcote C, Taylor J, Hall R, et al. Psychosocial support interventions for children and adolescents experiencing gender dysphoria or incongruence: a systematic review. Arch Dis Child. Oct 30, 2024;109(Suppl 2):s19-s32. [doi: 10.1136/archdischild-2023-326347] [Medline: 38594045]
13. Oles N, Darrach H, Landford W, et al. Gender affirming surgery: a comprehensive, systematic review of all peer-reviewed literature and methods of assessing patient-centered outcomes (part 1: breast/chest, face, and voice). Ann Surg. Jan 1, 2022;275(1):e52-e66. [doi: 10.1097/SLA.0000000000004728] [Medline: 33443903]
14. Oles N, Darrach H, Landford W, et al. Gender affirming surgery: a comprehensive, systematic review of all peer-reviewed literature and methods of assessing patient-centered outcomes (part 2: genital reconstruction). Ann Surg. Jan 1, 2022;275(1):e67-e74. [doi: 10.1097/SLA.0000000000004717] [Medline: 34914663]
15. Dy GW, Nolan IT, Hotaling J, Myers JB. Patient reported outcome measures and quality of life assessment in genital gender confirming surgery. Transl Androl Urol. Jun 2019;8(3):228-240. [doi: 10.21037/tau.2019.05.04] [Medline: 31380229]
16. Andréasson M, Georgas K, Elander A, Selvaggi G. Patient-reported outcome measures used in gender confirmation surgery: a systematic review. Plast Reconstr Surg. Apr 2018;141(4):1026-1039. [doi: 10.1097/PRS.0000000000004254] [Medline: 29595738]
17. Barone M, Cogliandro A, Di Stefano N, Tambone V, Persichetti P. A systematic review of patient-reported outcome measures following transsexual surgery. Aesthetic Plast Surg. Jun 2017;41(3):700-713. [doi: 10.1007/s00266-017-0812-4] [Medline: 28204933]
18. Jackman L, Chan C, Jacklin C, et al. Patient-reported outcome measures for paediatric gender-affirming care: a systematic review. Paediatr Child Health. Dec 2024;29(8):514-527. [doi: 10.1093/pch/pxae019] [Medline: 39990034]
19. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. May 2018;27(5):1159-1170. [doi: 10.1007/s11136-018-1829-0] [Medline: 29550964]

20. Bailey S, Zbukvic I, Dalton A, et al. Co-development and validation of a patient-reported experience measure for trans and gender diverse young people: a multi-stage participatory study protocol. Health Expect. Jun 2026;29(3):e70678. [doi: 10.1111/hex.70678] [Medline: 42076827]
21. Kaur MN, Morrison SD, Kennedy SL, et al. International study to develop a patient-reported outcome measure to evaluate outcomes of gender-affirming care - the GENDER-Q. J Patient Rep Outcomes. Nov 19, 2024;8(1):134. [doi: 10.1186/s41687-024-00785-x] [Medline: 39560846]
22. Kaur MN, Rae C, Morrison SD, et al. Development and assessment of a patient-reported outcome instrument for gender-affirming care. JAMA Netw Open. Apr 1, 2025;8(4):e254708. [doi: 10.1001/jamanetworkopen.2025.4708] [Medline: 40249619]
23. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. US Food and Drug Administration; 2009. URL: https://www.fda.gov/media/77832/download [Accessed 2026-09-17]
24. Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res. Oct 2013;22(8):1889-1905. [doi: 10.1007/s11136-012-0344-y] [Medline: 23288613]
25. Matza LS, Patrick DL, Riley AW, et al. Pediatric patient-reported outcome instruments for research to support medical product labeling: report of the ISPOR PRO good research practices for the assessment of children and adolescents task force. Value Health. Jun 2013;16(4):461-479. [doi: 10.1016/j.jval.2013.04.004] [Medline: 23796280]
26. 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 Health. 2005;8(2):94-104. [doi: 10.1111/j.1524-4733.2005.04054.x] [Medline: 15804318]
27. Mokkink LB, Elsman E, Terwee CB. COSMIN manual: conducting a systematic review of patient-reported outcome measures (PROMs), version 2.0. COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments); 2024. URL: https://www.cosmin.nl/wp-content/uploads/COSMIN-manual-V2_final.pdf [Accessed 2026-09-17]
28. Mokkink LB, Elsman EBM, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. Qual Life Res. Nov 2024;33(11):2929-2939. [doi: 10.1007/s11136-024-03761-6] [Medline: 39198348]
29. Kennedy SL, Cornacchi SD, Kaur MN, et al. Development of a patient-reported outcome measure for youth receiving gender-affirming care: the GENDER-Q Youth Module. Transgend Health. Sep 2025;10(4):375-385. [doi: 10.1089/trgh.2024.0203] [Medline: 40959407]
30. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Danmarks Paedagogiske Institut; 1960. ISBN: 9780598554512
31. Hobart JC, Cano SJ. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess. Feb 2009;13(12). [doi: 10.3310/hta13120] [Medline: 19216837]
32. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum. Dec 15, 2007;57(8):1358-1362. [doi: 10.1002/art.23108] [Medline: 18050173]
33. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol. Mar 2007;46(Pt 1):1-18. [doi: 10.1348/014466506x96931] [Medline: 17472198]
34. Knottnerus A, Tugwell P. STROBE--a checklist to Strengthen the Reporting of Observational Studies in Epidemiology. J Clin Epidemiol. Apr 2008;61(4):323. [doi: 10.1016/j.jclinepi.2007.11.006] [Medline: 18313555]
35. 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. May 2017;41(3):178-194. [doi: 10.1177/0146621616677520] [Medline: 29881087]
36. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. Sep 1951;16(3):297-334. [doi: 10.1007/BF02310555]
37. Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal components analysis of residuals. J Appl Meas. 2002;3(2):205-231. [Medline: 12011501]
38. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. May 2018;27(5):1147-1157. [doi: 10.1007/s11136-018-1798-3] [Medline: 29435801]
39. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Routledge; 1988. [doi: 10.4324/9780203771587]
40. Kennedy SL, Jack SM, Johnson N, Couturier J, Rae C, Klassen AF. Using a timeline activity in virtual concept elicitation interviews with youth: methods and lessons learned from an interpretive descriptive study to develop a

patient-reported outcome measure (GENDER-Q Youth). Patient. Jan 2026;19(1):5-20. [doi: 10.1007/s40271-025-00770-6] [Medline: 40952443]
41. Kennedy SL, Jack SM, Johnson N, Couturier J, Rae C, Klassen AF. Prioritizing participant and research team emotional safety during data generation and analysis in qualitative patient-reported outcome measure research: development of a framework informed by the GENDER-Q Youth study. Patient. Nov 2025;18(6):597-608. [doi: 10.1007/s40271-025-00756-4] [Medline: 40694295]
42. Chong LSH, Kerklaan J, Clarke S, et al. Experiences and perspectives of transgender youths in accessing health care: a systematic review. JAMA Pediatr. Nov 1, 2021;175(11):1159-1173. [doi: 10.1001/jamapediatrics.2021.2061] [Medline: 34279538]

Abbreviations

2SLGBTQI+: Two-Spirit, lesbian, gay, bisexual, transgender, queer or questioning, and intersex
CIHR: Canadian Institutes of Health Research
COSMIN: Consensus-Based Standards for the Selection of Health Measurement Instruments
CTT: classical test theory
DIF: differential item functioning
GAC: gender-affirming care
ISPOR: The Professional Society for Health Economics and Outcomes Research
MID: minimally important difference
PROM: patient-reported outcome measure
RMT : Rasch measurement theory
SDC: smallest detectable change
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TGD: transgender and gender diverse
TRT: test-retest

Edited by Amy Schwartz; The proposal for this study was peer-reviewed by the committee reviewing applicants to the Sex and Gender Science Chair in Gender Methods and Measures competition, Canadian Institutes of Health Research (CIHR). See Peer Review Report for details; submitted 14.May.2026; final revised version received 04.Aug.2026; accepted 09.Sep.2026; published 05.Oct.2026

Please cite as:
Klassen AF, Rae C, Morrison SD, Johnson N, Khatchadourian K, Graesser J, Kaur MN
GENDER-Q Youth Patient-Reported Outcome Measure: Protocol for an International Cross-Sectional Field Test Study
JMIR Res Protoc 2026;15:e100420
URL: https://www.researchprotocols.org/2026/1/e100420
doi: 10.2196/100420

© Anne F Klassen, Charlene Rae, Shane D Morrison, Natasha Johnson, Karine Khatchadourian, Jack Graesser, Manraj N Kaur. Originally published in JMIR Research Protocols (https://www.researchprotocols.org), 05.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Research Protocols, is properly cited. The complete bibliographic information, a link to the original publication on https://www.researchprotocols.org, as well as this copyright and license information must be included.

Review Type/Type d'évaluation:	Committee Member 1/Membre de comité 1
Name of Applicant/Nom du chercheur:	Klassen, Anne Frances
Application No./Numéro de demande:	427202
Agency/Agence:	CIHR/IRSC
Competition/Concours:	2019-10-08 Chair: CIHR Sex and Gender Science Chair/Chaire: Chaire science des sexe et du genre des IRSC
Committee/Comité:	Sex and Gender Science Chair/Chaire en science du sexe et du genre
Title/Titre:	Developing a Patient-Reported Outcome Measure for Children and Youth Seeking Gender-Affirming Health Care (GENDER-Q Kids)

Assessment/Évaluation:

Brief Summary of Proposal
The specific aims of the chair are to develop and validate Patient Reported Outcome Measures or PROMs for trans and gender non-binary children and adolescents related to gender affirming health care; the research involves a qualitative conceptual development phase, which will include the development of candidate items for PROMs, which will then be followed by field testing the items and scales with a lrge international sample of children and youth seeking gender affirming care.

Applicant Track Record
The applicant has a strong track record of developing PROMs for a variety of health issues, and a highly productive career to date. The proposed chair builds on a recent Foundation grant to develop PROMs for adult transgender health care outcomes.

The researcher has only relatively recent experience working with trans populations, but does have some significant funding (Foundation grant) and is working with recognized international collaborations, with an important focus, and high likelihood of success.

Research Program Proposal
This is a strong well-written, clear proposal, with clear identification of the project aims and methods, as well as ways that it leverages important prior work. The proposed research has the potential to create important patient-focused clinical outcome measures for evaluating care,

The applicant has well-described international engagement and potential for wide-ranging international impact, solid cross-cultural approaches. Methodology is well-developed and clear, quite appropriate. The collaborating partners are an equally impressive international team.

The timeframe for the research is quite feasible, and aligns with the stated goals.

One area that is underdeveloped in the proposal is articulating how this work amplifies understanding of gender and/or sex—although the work is focused on transgender children and youth, there was minimal acknowledgment of the diversity of gender identities among trans and non-binary young people. It seems highly likely that one size will not fit all, and that specific PROMs will vary widely by gender identities, but there were no discussion of whether a PROM set might need to be tailored to specific gendered experiences, or gender groups. Indeed, there appeared to be an assumption that one could create a singular PROM that would be relevant for transfeminine and transmasculine and non-binary children and youth, without attending to the distinctly different gendered spaces they occupy. The gender and sex section of the application was perfunctory, and so this lack of discussion of gender within the proposal itself left me wondering to what extent this researcher had actually incorporated gender into prior PROM work, despite having some very specific primarily gendered health outcomes to develop PROMs around, such as the BREAST-Q. That reduces my enthusiasm somewhat for what is otherwise a very strong proposal.

Review Type/Type d'évaluation:	Committee Member 1/Membre de comité 1
Name of Applicant/Nom du chercheur:	Klassen, Anne Frances
Application No./Numéro de demande:	427202
Agency/Agence:	CIHR/IRSC
Competition/Concours:	2019-10-08 Chair: CIHR Sex and Gender Science Chair/Chaire: Chaire science des sexe et du genre des IRSC
Committee/Comité:	Sex and Gender Science Chair/Chaire en science du sexe et du genre
Title/Titre:	Developing a Patient-Reported Outcome Measure for Children and Youth Seeking Gender-Affirming Health Care (GENDER-Q Kids)

Assessment/Évaluation:

Capacity Building Proposal
The applicant has solid evidence of success in recruiting and mentoring a number of trainees from around the world, and helping launch their careers with effective co-authorship of a large number of papers. The current proposal builds on those successes by incorporating clinician scientists and graduate trainees, some with lived experience, in the research.

The applicant described a solid set of approaches to mentoring a large number of trainees in the highly effective methods, with extensive opportunities to develop important skills in conducting research, writing and presenting. There is high likelihood of success in building a further cadre of clinician scientists able to develop and evaluate outcome measures with pediatric patients for a variety of key health issues, although how the researcher would provide training to further their expertise in gender- and sex-based approaches to healthresearch, and thus advance the field, was not made explicit.

Knowledge Translation Plan
The KTE proposal is workmanlike, with little that is innovative, but there is a clear plan with a variety of strategies that are feasible and realistic approaches, and clearly the productivity is high.

Environment and Support
There is excellent evidence of support for this work, at a top tier research intensive university.

Comments on the budget requested and a formal recommendation.
No concerns about budget, well justified.

Review Type/Type d'évaluation:	Committee Member 2/Membre de comité 2
Name of Applicant/Nom du chercheur:	Klassen, Anne Frances
Application No./Numéro de demande:	427202
Agency/Agence:	CIHR/IRSC
Competition/Concours:	2019-10-08 Chair: CIHR Sex and Gender Science Chair/Chaire: Chaire science des sexe et du genre des IRSC
Committee/Comité:	Sex and Gender Science Chair/Chaire en science du sexe et du genre
Title/Titre:	Developing a Patient-Reported Outcome Measure for Children and Youth Seeking Gender-Affirming Health Care (GENDER-Q Kids)

Assessment/Évaluation:

Applicant:

The applicant is a highly productive, high-impact (h-index of 47-57), senior-career researcher in health outcomes (Professor of Pediatrics, Health Sciences at McMaster) with expertise in development of patient-reported outcome measures (PROMs). Applicant is CIHR Foundation Grant holder for a related research developing and refining the Gender-Q PROMs for transgender adults receiving gender affirming treatments, and also currently holds numerous project/operating grants as PI. Applicant has extensive experience with proposed methodology, and has clearly demonstrated success in the development, refinement, and dissemination of PROMs for various conditions where plastic surgeries are indicated, e.g., breast reconstruction, cleft palate. Only possible weakness is a relatively smaller involvement in student research supervision than is typical for a researcher at this career stage (graduate two PhDs, two MScs).

Research Proposal:

Proposed research to develop a PROMs for children and youth seeking gender affirming health care. The proposed methodology, qualitative item development followed by Rasch analysis, is well-established, empirically-supported, and highly rigorous and the applicant has extensive experience with this methodology. International field test of the measure will follow. Extensive and high-profile network of collaborators working on the adult Gender-Q. Chair award would capitalized on existing expertise, networks, infrastructure, and funding to develop patient outcome measure for a growing sector of health-care seeking children and youth.

Capacity Building:

Capacity building is focused on trainees, with funding to support two graduate students and two post-doctoral fellows. The training environment is excellent with significant opportunities to network with international leaders in patient-outcome research. Applicant has relatively limited student supervision experience.

Knowledge Translation:

The major highlight here is the free dissemination of the Gender-Kids PROMs to nonprofit users through the established website. Promotion of the measure is planned through academic channels.

Environment and Support:

The environment is exceptional and well-established with demonstrated success, supportive research infrastructure. Noted that the applicant is a member of CanChild centre of excellence with KT experts to assist.

Review Type/Type d'évaluation:	Committee Member 2/Membre de comité 2
Name of Applicant/Nom du chercheur:	Klassen, Anne Frances
Application No./Numéro de demande:	427202
Agency/Agence:	CIHR/IRSC
Competition/Concours:	2019-10-08 Chair: CIHR Sex and Gender Science Chair/Chaire: Chaire science des sexe et du genre des IRSC
Committee/Comité:	Sex and Gender Science Chair/Chaire en science du sexe et du genre
Title/Titre:	Developing a Patient-Reported Outcome Measure for Children and Youth Seeking Gender-Affirming Health Care (GENDER-Q Kids)

Assessment/Évaluation:

Budget:

Participant stipends for qualitative interviews is entered twice, so -$4K. Otherwise, no concerns.

Review Type/Type d'évaluation:	Committee Member 3/Membre de comité 3
Name of Applicant/Nom du chercheur:	Klassen, Anne Frances
Application No./Numéro de demande:	427202
Agency/Agence:	CIHR/IRSC
Competition/Concours:	2019-10-08 Chair: CIHR Sex and Gender Science Chair/Chaire: Chaire science des sexe et du genre des IRSC
Committee/Comité:	Sex and Gender Science Chair/Chaire en science du sexe et du genre
Title/Titre:	Developing a Patient-Reported Outcome Measure for Children and Youth Seeking Gender-Affirming Health Care (GENDER-Q Kids)

Assessment/Évaluation:

The applicant has the required expertise and achievements record.
The applicant proposes developing a GENDER-Q for pediatric patients. The proposal includes more data on previous PROMs than the methods used to develop the new one.
Timeline is missing.
KT strategy is a strength.
A disclosure on budgetary overlaps is required. Also the PI's total time allocated for this chair to finalize the objectives in 4 years.
Budget can be improved (The amount allocated to personnel vs trainees/capacity building)

Review Type/Type d'évaluation:	SO Notes /Notes de l'agent scientifique
Name of Applicant/Nom du chercheur:	Klassen, Anne Frances
Application No./Numéro de demande:	427202
Agency/Agence:	CIHR/IRSC
Competition/Concours:	2019-10-08 Chair: CIHR Sex and Gender Science Chair/Chaire: Chaire science des sexe et du genre des IRSC
Committee/Comité:	Sex and Gender Science Chair/Chaire en science du sexe et du genre
Title/Titre:	Developing a Patient-Reported Outcome Measure for Children and Youth Seeking Gender-Affirming Health Care (GENDER-Q Kids)

Assessment/Évaluation:

Strengths:

- The objective of this program is to develop and validate gender affirming Patient Reported Outcome Measures or PROMs for pediatric patients. The applicant has significant experience in developing PROMs for a variety of health issues and the proposed methodology.
- The applicant has had a productive career to date. The researcher has significant funding and recent experience with transgender individuals
- This was a well written and clear proposal that will produce an important clinically focused measure. They provide solid cross-cultural approaches and have assembled strong collaborations.
- The applicant has a strong history of trainee mentorship and promotion. This is an excellent environment for the development of trainees.
- excellent environmental support.

Weaknesses:

- There was minimal acknowledgement of the diversity of gender minorities.
- They treat the PROMS measure with assumptions of the homogeneity of outcomes and this was viewed a shortcoming of the grant. The applicants could proactively deal with this diversity in the proposal and could address how these PROMS could be tailored to specific trans and non-binary groups.
- the KT proposal was clearly defined but was not necessarily innovative.
- It was acknowledged that there was youth engagement in the development of the PROMS. However, there was not youth engagement in conduct and direction of the research and this was perceived as a weakness.

Budget:

No issues discussed – one noted in a reviewer comment.

Multimedia Appendix 1: A priori construct validity hypotheses

	Body Image	Psych Distress	Psych WB	Life Impact	Family	School	Social	Teacher	Voice Sound	Voice Distress	Binding WB	Binding AE	Tucking AE	Health Care Team	Info Hormones	Treatment Outcome
GROUP HYPOTHESES: SCORES HIGHER FOR THE FOLLOWING:																
1. Those who live (vs. do not) in their gender identity	✓	✓	✓	✓	✓											
2. Those who like (vs. do not) how their face looks overall	✓															
3. Those not teased or bullied (vs. are) in past month		✓	✓			✓	✓									
4. Those in a school that support (vs. do not) TGD students						✓		✓								
5. Those happy (vs. not happy) with their teachers at school						✓		✓								
6. Those who rate their mental health higher (poor to excellent)		✓	✓	✓												
7. Those who feel happy (vs. sad)		✓	✓	✓												
8. Those happy (vs. not happy) with their life overall		✓	✓	✓			✓									
9. Those with ≥1 close friend (vs. none)							✓									
10. Those with ≥1 close friend (vs. none) who knows they are TGD							✓									
11. Those who spend time (vs. do not) with friends							✓									
12. Those happy (vs. not) with their social life				✓			✓									
13. Those who rate their social life higher (poor to excellent)				✓			✓									
14. Those happy (vs. not happy) with their family life					✓											
15. Those living with ≥1 person (vs. none)					✓											

who supports their TGD identity																
16. Those living with supportive family (vs. not) of their TGD identity					✓											
17. Those happy (vs. not happy) with how they look overall	✓	✓	✓	✓							✓					
18. Those who voice sounds right (vs. too high or too low)									✓	✓						
19. Those happy (vs. not happy) with how their voice sounds									✓	✓						
20. Those who like (vs. dislike) how their voice sounds overall									✓	✓						
21. Those happy (vs. not happy) with how they look when they bind	✓										✓					
22. Those whose chest looks completely flat (vs. not) when they bind	✓										✓					
23. Those who bind fewer days/hours in a week												✓				
24. Those with did not see (vs. did see) HCT for binding problems												✓				
25. Those without (vs. with) a binding problem												✓				
26. Those who tuck fewer days/hours in a week													✓			
27. Those without (vs. with) a tucking problem													✓			
28. Those who did not see (vs. did see) HCT for tucking problems													✓			
29. Those on longer duration of hormone treatment	✓													✓	✓	✓
30. Those happy (vs. not happy) with treatment in past month														✓		✓

31. Those who do not want (vs. want) top surgery	✓															
32. Those who do not want (vs. want) voice therapy									✓	✓						
33. Those happy (vs. not happy) with the staff at gender clinic														✓		✓
34. Those with more visits (vs. less) to the gender clinic														✓		
35. Those happy (vs. not happy) with knowledge of hormones															✓	
36. Those who know more (vs. less) about hormones															✓	
GROUP HYPOTHESES TOTAL	**7**	**6**	**6**	**7**	**4**	**3**	**7**	**2**	**4**	**4**	**3**	**3**	**3**	**4**	**3**	**3**
CORRELATIONAL HYPOTHESES: see footnote at the bottom of the table to interpret magnitude of association																
37. Positive correlation Body Image		3	3	3			2				1					2
38. Positive correlation Psych Distress	3		3	3	1	1	3	1	2	2						2
39. Positive correlation Psych Well-being	3	3		3	1	1	3	1	2	2						2
40. Positive correlation Life Impact	3	3	3		1	1	3	1	2	2						2
41. Positive correlation Family		1	1	1												
42. Positive correlation School		1	1	1				1								
43. Positive correlation Social	2	3	3	3					2	2						
44. Positive correlation Teacher		1	1	1		1										
45. Positive correlation Voice Sound		2	2	2			2			3						1
46. Positive correlation Voice Distress		2	2	2			2		3							1
47. Positive correlation Binding Well-being	1															
48. Positive correlation Binding Adverse Effects																
49. Positive correlation Tucking Adverse Effects																
50. Positive correlation Health Care Team															1	1

51. Positive correlation Information: Hormones														1		1
52. Positive correlation Treatment Outcome	2	2	2	2					1	1				1	1	
CORRELATIONAL HYPOTHESES TOTAL	**6**	**10**	**10**	**10**	**3**	**4**	**6**	**4**	**6**	**6**	**1**	**0**	**0**	**2**	**2**	**8**
TOTAL (GROUP + CORRELATIONAL)	**14**	**16**	**16**	**17**	**7**	**7**	**14**	**6**	**10**	**10**	**4**	**3**	**3**	**6**	**5**	**11**

AE = Adverse Effects; WB = Well-being; HCP = Health Care Team; 1=>0.3; 2=0.3-0.5; 3=>0.5. Note: 2 and 3 are based on correlations between scales from the GENDER-Q adult module analysis