Elevating Equity: Transforming Teaching Evaluation at CU Denver

Executive Summary

Purpose and Context

The Equitable Faculty Course Questionnaire (FCQ) Working Group at the University of Colorado Denver has undertaken a critical examination of current teaching evaluation practices, driven by a fundamental commitment to equity, transparency, and academic excellence. This initiative emerges from a growing recognition of systemic biases inherent in traditional student evaluation of teaching (SET) tools, which have been consistently shown to disadvantage women and marginalized faculty members.

The purpose of this comprehensive review was to examine current practices and perceptions of faculty, students, and administrators regarding the FCQs. By gathering these insights, the review seeks to reimagine and restructure the FCQ process in ways that better align with the university's strategic goal of becoming an equity-serving institution. Rooted in the Faculty Assembly FCQ Resolution of May 2023 (Appendix B), the working group identified significant concerns: existing evaluation mechanisms demonstrate little relationship with actual teaching quality while perpetuating potential discriminatory practices.

Key contextual factors driving this initiative include:

- Alignment with University of Colorado Regent Law, which emphasizes faculty-driven evaluation processes
- Recognition of inherent biases in current student evaluation of teaching methodologies
- A commitment to developing more holistic, fair, and meaningful approaches to assessing teaching effectiveness

The working group's mandate extends beyond merely identifying problems; it seeks to develop actionable, evidence-based strategies that empower faculty, center student learning, and promote a more just and inclusive academic environment. By challenging traditional evaluation paradigms, this initiative represents a forward-thinking approach to professional assessment that respects the complexity and diversity of teaching practices across disciplines.

Through a collaborative effort involving faculty, students, and administrators, the project aims to transform teaching evaluation from a potentially punitive mechanism to a constructive tool for professional growth, institutional improvement, and educational excellence.

Working Group Members (with special thanks)

Faculty

- Lucy Dwight (SPA - Retired)
- Ashley Hoffman (CAM)
- Maren Scull (CLAS)
- Jennifer Camacho Taylor (SEHD)
- Cristina Gillanders (SEHD)
- Courtney Donovan (SEHD)
- Liana Hall (Doctoral Student)
- Abbey Eversole (Doctoral Student)
- Chelsea Sobczak (Doctoral Student)
- Dennis DeBay (Chair - SEHD)

Staff

- Rachel Brown
- Amy McGuire

Key Findings Highlights

This comprehensive analysis of 29,564 courses taught by 3,881 instructors over seven years offers valuable insights into patterns of teaching evaluation and their impact across faculty groups. While the findings reveal areas where bias occurs, they also highlight opportunities to strengthen evaluation practices to better recognize instructional excellence. All measures—including course evaluations, peer feedback, student success indicators, and institutional data—were reviewed to ensure a holistic perspective. To deepen this work, we incorporated focus groups with faculty and students,, along with targeted surveys of administrators, to gather richer, more collaborative insights.

Quantitative Evidence of Systematic Bias

- Black, Indigenous, and People of Color (BIPOC) faculty consistently score 0.12-0.13 points lower on critical evaluation items used for promotion and merit decisions
- First-year faculty face systematic disadvantage with scores 0.13-0.15 points lower than established faculty during their most vulnerable career period
- Larger enrollment courses receive consistently lower ratings, penalizing faculty who teach high-enrollment classes
- STEM courses show systematic disadvantage across most evaluation items, creating particular challenges for faculty in quantitative disciplines
- Adjunct and part-time instructors score significantly lower than full-time faculty (tenure and non-tenure track), reinforcing existing institutional hierarchies

Areas forFurther Exploration (not captured in this study):

- New or recently updated courses

- International faculty experiences
- Condensed course formats (e.g., summer sessions)
- Active learning vs. lecturing (pedagogy)

Institutional Dysfunction and Unintended Consequences

Because FCQs carry significant weight in merit, promotion, and tenure decisions across most colleges and units, their design and use have produced several unintended consequences. While intended to provide meaningful feedback on teaching effectiveness, the current system often generates confusion, discourages innovation, places undue strain on faculty well-being, and overlooks critical contextual factors. The following themes highlight the most pressing concerns identified in this review.

- Widespread confusion exists about FCQ purpose and appropriate use among faculty, students, and administrators.
- Concerns about implementation of the FCQs in terms of timeline, appropriateness, and response rates.
- Innovation penalty: Faculty report that the current FCQ system discourages pedagogical experimentation, as innovative or nontraditional teaching approaches often receive lower scores. Without a shared definition of what constitutes "innovation" in teaching, it is difficult to fully assess the extent of this effect. Nonetheless, the perception of a penalty can create a disincentive for trying new strategies, limiting instructional growth and adaptation.
- Mental health impact: Faculty—especially those who are new, underrepresented, or undergoing promotion—report that the evaluation process can cause significant psychological strain. While some stress and anxiety are expected in any evaluation, the lack of clarity around how these evaluations are used intensifies the pressure. This vagueness is particularly harmful for those in high-stakes career stages, making them more vulnerable to the process's negative effects.
- Contextual blindness: Standardized evaluations fail to account for fundamental differences in course size, content, modality, and student populations.
- Although some faculty found the qualitative comments helpful for improving their teaching, the quantitative feedback was generally not useful which runs counter to the intended purpose of the FCQs.

Administrative Awareness and Systemic Problems

- **Administrators' Acknowledgment of Bias in FCQs and the Need for Reform**
    - University of Colorado administrators have recognized patterns of bias within the Faculty Course Questionnaires (FCQs), particularly concerning gender and racial disparities, which may influence high-stakes personnel decisions such as tenure and promotion (CU Faculty Council, 2016).
- **Lack of Standardized Interpretation Across Campuses**
    - There is currently no standardized process for interpreting FCQ data across the CU campuses. This lack of uniformity complicates the equitable assessment of faculty performance and undermines the reliability of FCQs as a sole evaluative tool (CU Faculty Council, 2016).
- **Administrative Support for Exploring Alternatives**

    - A significant majority of CU administrators support exploring alternatives to the current FCQ system. This overwhelming interest indicates a collective acknowledgment of the need for reform to ensure fair and effective faculty evaluations (CU Faculty Council, 2016).
- **Student Burden in Faculty Evaluations**
    - The current reliance on FCQs places an undue burden on students by assigning them the responsibility of evaluating faculty performance. This practice not only affects the quality of evaluations but also shifts the accountability for faculty assessment away from institutional mechanisms (CU Faculty Council, 2016).

Recommendations Overview

Based on extensive research and stakeholder input, the working group recommends a fundamental transformation of faculty teaching evaluation that prioritizes equity, meaningful feedback, and pedagogical growth over standardized metrics that perpetuate bias and are not helpful for teaching improvement.

Immediate Action Items

- Suspend high-stakes use of current FCQs for promotion, tenure, and merit decisions until bias patterns are addressed.
- Implement bias awareness training for all administrators involved in faculty evaluation processes.
- Establish clear guidelines for appropriate FCQ use and interpretation across all academic units.
- Develop additional evaluation pathways designed to address the documented inequities experienced by faculty in systematically disadvantaged groups. These pathways should be grounded in clear, evidence-based criteria to ensure they target the specific barriers these faculty face, rather than serving as a general alternative for all.
- Provide support, especially to new faculty in interpreting FCQs and understanding how they can be used to improve teaching.

Long-term Systemic Changes

- Develop context-specific evaluation tools that account for course size, content, modality, and student population differences
- Implement multi-source evaluation systems combining peer review, teaching portfolios, student learning assessment, and self-reflection
- Establish timing reforms to provide more meaningful feedback opportunities throughout the academic term
- Create professional development programs focused on inclusive teaching practices and bias mitigation
- Create mechanisms for continuous improvement in the use of evaluation of teaching tools

Institutional Culture Transformation

- Shift from evaluation to improvement: Redesign systems to prioritize faculty development over punitive assessment
- Support pedagogical innovation through evaluation processes that reward rather than penalize creative teaching approaches
- Promote equity-centered practices that recognize and address systemic barriers faced by underrepresented faculty
- Engage students as partners in meaningful feedback processes rather than high-stakes evaluators

Implementation Strategy

The working group recommends a phased approach beginning with immediate harm reduction measures while developing comprehensive alternatives through collaborative faculty-led processes. This transformation requires sustained institutional commitment, adequate resources, and ongoing assessment to ensure equity goals are achieved.

These recommendations represent a fundamental shift from the current system toward evaluation processes that support all faculty members, promote innovative teaching, and genuinely serve student learning while upholding principles of fairness and institutional excellence.

I. Introduction

Historical Context of FCQ Challenges

The Faculty Course Questionnaire (FCQ) has long been a standard mechanism for evaluating teaching effectiveness in higher education at our institution. However, mounting evidence has revealed significant systemic challenges inherent in these evaluation tools. Extensive research has demonstrated that student evaluations of teaching (SET) are riddled with problematic biases that disproportionately impact women and marginalized faculty members.

These biases manifest in multiple ways, including:

- Scoring discrepancies based on faculty gender
- Differential treatment of faculty from underrepresented groups
- Evaluations that measure student satisfaction rather than actual teaching quality
- Limited correlation between FCQ scores and objective measures of pedagogical effectiveness

The traditional FCQ model has increasingly been recognized as an instrument that potentially reinforces institutional inequities rather than promoting genuine academic excellence.

Faculty Assembly Resolution Background

In May 2023, the Faculty Assembly issued a pivotal resolution that critically examined the existing FCQ process. The resolution explicitly highlighted the inherent biases and inappropriate uses of student evaluations in faculty assessment. Drawing direct connection to the University of Colorado Regent Law,

Article 5.A.1.(B), the resolution emphasized that tenured and tenure-track faculty should have principal responsibility in developing fair evaluation mechanisms (Appendix B).

The resolution's key points included:

- Acknowledging systemic biases in current evaluation methods
- Calling for alternative assessment approaches
- Protecting both faculty and student rights
- Ensuring evaluation processes align with institutional equity goals

Working Group Formation and Mandate

In response to the Faculty Assembly Resolution, the Equitable FCQ Working Group was established as a collaborative, faculty-led initiative. Comprising diverse representatives from across academic disciplines, the working group was charged with a comprehensive review and transformation of faculty evaluation practices.

The group's core mandate included:

- Conducting a comprehensive literature review on faculty evaluation methods
- Exploring alternative evaluation approaches
- Developing strategies to mitigate potential biases
- Creating recommendations for a more equitable assessment process
- Reviewing the impact of FCQs on merit and promotion, across schools and colleges

The working group's composition reflects the university's commitment to inclusive problem-solving, with members representing various schools and departments, including:

- Human Development and Family Relations
- Public Affairs
- Early Childhood Education
- Visual Arts
- Sociology
- Business
- STEM Education

Guided by core values of inclusivity, integrity, empowerment, and a student-centered focus, the group embarked on a mission to reimagine teaching evaluation as a constructive, growth-oriented process that genuinely supports educational excellence.

Throughout their summer work, the group held eight meetings, invited key stakeholders like the Student Government Association and the Director of the Center of Excellence in Teaching and Learning, and began laying the groundwork for a transformative approach to faculty assessment.

II. Literature Review

Biases in Course Evaluation

Extensive research demonstrates that student evaluations of college courses contain significant biases that compromise their validity as measures of teaching effectiveness. Studies consistently show that instructor scores may be loosely correlated or completely uncorrelated with actual teaching quality (Kreitzer & Sweet-Cushman, 2021; Stark & Freishtat, 2014; Uttl et al., 2017). These biases generally fall into two primary categories: measurement bias and equity bias.

Measurement bias encompasses factors unrelated to teaching quality or course content that nonetheless influence evaluation scores. These factors include class size, scheduling time, course modality (online versus in-person), academic discipline, difficulty level, and whether the course is required or elective (Kreitzer & Sweet-Cushman, 2021). Such variables create systematic distortions in evaluation data that obscure meaningful assessment of instructional effectiveness.

Equity bias refers to the influence of students' perceptions of a faculty member's demographic characteristics—such as gender, ethnicity, age, sexuality, and, in limited research —on course evaluations. The evidence presents a complex picture: some studies find that women faculty face more pronounced inequities than other minoritized groups (Kreitzer & Sweet-Cushman, 2021), while others indicate that instructor ethnicity may be a stronger predictor, with faculty of color receiving harsher evaluations than their white colleagues. The role of ability remains largely unexplored, representing a significant gap in the literature.

Student characteristics also influence equity bias patterns. Male students consistently rate female instructors lower than female students do, while female students tend to give higher scores to female instructors (Young et al., 2009). Additionally, research indicates that students provide higher ratings when they share the same ethnicity as their instructor (Centra, 2000).

Faculty Perceptions and Uses of Student Evaluations

While improving teaching quality represents the primary stated goal of course evaluation systems in universities, evidence suggests limited success in achieving this objective. Despite receiving student evaluations year after year, many instructors fail to demonstrate measurable teaching improvement (Marsh, 2007). This disconnect stems partly from faculty concerns about evaluation bias and skepticism regarding their usefulness.

Faculty generally report that student evaluations provide minimal assistance for teaching improvement and express a need for more meaningful feedback (Beran & Rokosh, 2009; Safadi et al., 2013). Many instructors indicate that student evaluations actually discourage creativity and innovation, as concerns about negative feedback that could impact promotion and merit reviews lead to more conservative teaching approaches (Beran & Rokosh, 2009).

Despite widespread criticism, some faculty members do find value in student evaluations, using them to guide teaching preparation, delivery methods, and student interactions (Hammer et al., 2018, Safadi et

al., 2013). Award-winning university faculty often view student feedback as opportunities for self-reflection and continuous improvement (Golding & Adam, 2016).

Faculty trust in student evaluations varies considerably based on their perceived relationship with students. Instructors who feel "underestimated" by their students are more likely to question evaluation validity and find the feedback counterproductive for teaching improvement (Hammer et al., 2018). This suggests that higher education institutions should recognize these attitudes and develop alternative methods for providing constructive feedback to faculty.

Use of Student Evaluations for Promotion Decisions

Student evaluations have become deeply embedded in faculty personnel decisions across higher education institutions, despite mounting evidence questioning their validity and reliability as measures of teaching effectiveness. Even when student evaluations are unbiased, reliable, and valid, they can still create unfair outcomes when used in high-stakes personnel decisions (Esarey & Valdes, 2020). Significant gaps exist between sound performance appraisal practices and typical implementation of student evaluations in academic settings (Lohman, 2021).

Academic units demonstrate considerable variation in how they utilize student evaluations for personnel decisions. Many administrators lack adequate training in statistical interpretation, leading to misuse of evaluation scores (Benton & Cashin, 2012). Research reveals that administrators often rely heavily on student evaluations without fully understanding their limitations (Ferguson et al., 2020).

Faculty members, particularly women and faculty of color, express significant concerns about the impact of student evaluations on career advancement (Davidovitch & Eckhaus, 2019). These groups face documented bias in student evaluations yet see these same evaluations used as evidence of teaching effectiveness in promotion decisions. Substantial disconnects exist between what student evaluations actually measure and how they are interpreted in promotion contexts (Beran et al., 2005).

Lecturers and non-tenure-track faculty face even greater vulnerability, as they often experience heavier reliance on student evaluations in personnel reviews while having fewer opportunities to demonstrate excellence through research or service (Kreitzer & Sweet-Cushman, 2022). This creates a particularly vulnerable population disproportionately affected by evaluation system biases.

Student Perceptions of Teaching Effectiveness

Research on student perceptions of teaching effectiveness remains limited in scope and inclusivity. Existing studies typically focus on small samples from specific fields, often excluding graduate and online students. Important student characteristics such as first-generation status, age, race/ethnicity, and gender are rarely considered comprehensively. Comparative studies examining faculty characteristics and years of service are also uncommon.

Multiple studies have documented gender bias in student perceptions, consistently finding that women faculty are perceived as having lower teaching effectiveness (MacNell et al., 2015; Mitchell & Martin,

2018). Additional factors influencing student perceptions include class size, workload, and students' prior subject interest and experience (Hu, 2020).

Current literature indicates that students receive inadequate guidance from college administrations or faculty members regarding the purpose and desired impacts of evaluations (Dana et al., 2023, Hu, 2020). This lack of preparation connects to issues of cognitive engagement, accurate recall, time constraints for assessment, and minimal trust that constructive feedback will be impactful. Consequently, student feedback often lacks depth and thorough reflection.

Research on student perceptions of teaching effectiveness typically centers on teaching methods, content knowledge, instructor approachability, and respectful treatment of students (Alhija, 2017; Hu, 2020; Liu et al., 2016).

Alternative Methods for Evaluating Faculty Teaching

Recognition of student evaluation limitations has prompted exploration of alternative and complementary approaches to assessing teaching effectiveness. These methods aim to provide more comprehensive, fair, and pedagogically meaningful evaluation of faculty teaching performance.

Peer evaluation represents one of the most widely discussed alternatives. Structured peer review processes can provide valuable insights into teaching practices that students cannot assess, focusing on pedagogical expertise, curriculum design, and instructional innovation where faculty colleagues possess necessary disciplinary knowledge (Bell & Mladenovic, 2008).

Teaching portfolios offer a comprehensive alternative by providing a more holistic view of teaching effectiveness than student ratings alone. These portfolios typically include evidence of teaching effectiveness, improvement activities, and reflective statements, allowing faculty to present a more complete picture of their teaching contributions (Murray, 2005, Wheeler et al., 2009).

Self-reflection and assessment mechanisms represent additional evaluation components. Validated instruments for measuring reflective thinking in teaching provide tools for meaningful self-assessment of pedagogical practices (Kember et al., 2000).

Student learning assessment offers a more direct measure of teaching effectiveness than satisfaction-based evaluations. Classroom assessment techniques that measure student learning rather than student opinions provide frameworks for meaningful evaluation (Angelo & Cross, 2011).

Multiple-source evaluation systems combining various assessment methods show particular promise. Comprehensive faculty evaluation models integrating student feedback, peer review, self-evaluation, and learning outcome assessment address limitations of single-source evaluation methods (Arreola, 2004). Technology-enhanced approaches also offer new possibilities, including innovative student evaluation methods leveraging multimedia portfolios and real-time feedback systems (Berk, n.d.).

Psychological Impact of Evaluations on Faculty

The psychological effects of teaching evaluations on faculty members represent an underexplored but critically important dimension of evaluation systems. Research reveals that current student evaluation practices can significantly impact faculty well-being, professional identity, and teaching practices.

Negative evaluations can substantially affect faculty members' professional confidence and teaching identity. Faculty who feel underestimated by students are more likely to question feedback validity and experience decreased motivation for teaching improvement, creating a counterproductive cycle where evaluation systems intended to improve teaching may actually undermine faculty engagement (Hammer et al., 2018).

The high-stakes nature of evaluation usage amplifies psychological impact. Faculty experience significant stress and anxiety related to student evaluations, particularly when used for promotion and tenure decisions (Stupnisky et al., 2018). Faculty from underrepresented groups experience additional psychological burden from potentially biased evaluation systems (Gonzales & Terosky, 2016).

Student evaluations can discourage creativity and innovation in teaching, as faculty become concerned about negative feedback impacting career advancement (Beran & Rokosh, 2009). Overemphasis on student evaluations can create a "teaching to the test" mentality, where faculty focus on maximizing evaluation scores rather than optimizing student learning (McKeachie, 1997).

However, positive psychological impacts are possible when evaluation systems provide meaningful, constructive feedback and are used primarily for improvement rather than judgment (Penny & Coe, 2004). This suggests that psychological impact depends significantly on evaluation system design, implementation, and intended use.

III. Research Methodology

This FCQ transformation initiative employed a comprehensive mixed-methods research design that combined rigorous quantitative analysis with in-depth qualitative investigation and administrative perspectives. The methodology was specifically designed to identify systematic bias patterns and understand the institutional impact of current faculty teaching evaluation practices through multiple complementary approaches.

Our methodology employs a two-pronged approach:

1. Quantitative Analysis: Graduate research assistants conducted extensive statistical analysis of historical FCQ data, examining patterns and correlations across multiple variables. This analysis provided empirical evidence of systemic patterns and potential biases.
2. Qualitative Investigation: Through faculty focus groups, we explored deeper insights into FCQ experiences, concerns about bias, and perspectives on evaluation effectiveness. These sessions were conducted by graduate student assistants to ensure peer-level dialogue and professional comfort. Additionally, these discussions provided an opportunity to identify and evaluate alternate means of assessment, such as peer evaluations, self-assessments, or qualitative feedback tools, that could complement or replace traditional FCQs in fostering a more comprehensive and equitable evaluation framework.

Quantitative Analysis

The quantitative component involved an extensive longitudinal examination of FCQ data spanning seven academic years (2018-2024). The analysis drew from comprehensive institutional records encompassing 29,564 courses taught by 3,881 instructors across the university. Data sources included FCQ scores across all academic departments and programs, comprehensive demographic information of faculty participants, course-specific contextual variables (enrollment size, level, modality, type), institutional records related to faculty positions, rank, and career stage, and department and college affiliation data. The quantitative analysis employed a QuantCrit (Quantitative Critical Race Theory) framework to examine patterns of systematic bias, with systematic bias operationally defined as "consistent and repeatable deviation that consistently skews results in a particular direction" and "is not reduced by sample size." This framework ensured that statistical analysis would specifically identify patterns that disadvantage particular groups rather than random variations.

The quantitative methodology employed sophisticated statistical techniques to comprehensively examine bias patterns. Descriptive analysis provided statistical summarization of FCQ distributions across multiple dimensions and examination of response patterns and completion rates. Comparative analysis utilized correlation analysis to examine relationships between FCQ scores and various factors, analysis of variance (ANOVA) and t-tests to identify patterns in score variance based on course and instructor variables, and factorial ANOVA to examine intersectional patterns across multiple demographic and contextual factors. Advanced modeling incorporated factor analysis to determine whether FCQ items constitute a holistic unidimensional construct suitable for evaluation purposes, structural equation modeling to identify course-level and instructor-level variables that predict FCQ scores, and effect size calculations to determine practical significance of identified differences.

The analysis comprehensively explored multiple dimensions of potential bias through both course-level and instructor-level variables. Faculty demographics examined included gender identity and expression, race and ethnicity (with particular attention to BIPOC faculty experiences), age and career stage, faculty position type (tenure-track, adjunct, instructor, research faculty), and years of teaching experience. Course characteristics analyzed encompassed enrollment size and class composition, course level (undergraduate vs. graduate), delivery modality (in-person, online, hybrid), course type (required vs. elective, foundational vs. advanced), subject matter and disciplinary context, and college and department affiliation. Intersectional analysis examined multiple identity intersections (e.g., women of color in STEM), combined effects of faculty demographics and course characteristics, and temporal patterns affecting new faculty during vulnerable career periods.

Qualitative Investigation

Faculty & Student Focus Groups

The qualitative component utilized structured focus group methodologies to gather in-depth insights into lived experiences with FCQs. The design prioritized creating safe spaces for honest dialogue while maintaining professional confidentiality. Focus groups were conducted separately for different

stakeholder groups to encourage open discussion within peer communities. Focus groups were strategically composed to ensure comprehensive representation, with faculty focus groups including representation across diverse academic disciplines and colleges, inclusion of various faculty ranks, positions, and career stages, deliberate inclusion of faculty with diverse demographic backgrounds, and separate sessions for different faculty types to encourage candid discussion. Student focus groups included students from multiple academic levels and programs, diverse student demographics and backgrounds, and students with varying levels of FCQ completion experience.

Data collection methods incorporated semi-structured focus group discussions with experienced facilitators, individual follow-up interviews where appropriate, and anonymous written reflections to supplement verbal discussions. The analysis approach utilized systematic thematic analysis of all transcripts, constant comparative method to identify patterns across groups, multiple rounds of coding to ensure comprehensive theme development, and member checking with participants to validate interpretations. Primary objectives included exploring faculty and student perspectives on current evaluation processes, identifying systemic challenges and unintended consequences of FCQ mechanisms, understanding the psychological and professional impact of current assessment practices, generating insights into potential alternative evaluation methods, and documenting lived experiences of navigating current assessment frameworks.

Administrative Perspectives

A comprehensive survey was administered to academic administrators including deans, associate deans, and provost office personnel to understand institutional perspectives on FCQ use and effectiveness. Survey components addressed current use of FCQs in decision-making processes, awareness of bias patterns and limitations, institutional policies and procedures related to FCQ interpretation, interest in exploring alternative evaluation methods, and recommendations for system improvements.

Methodological Considerations and Safeguards

The methodology incorporated multiple ethical safeguards to ensure research integrity and participant protection. Participant protection measures included comprehensive informed consent processes, voluntary participation with clear withdrawal options, multiple layers of confidentiality protection, minimal risk assessment and mitigation strategies, and protection against potential retaliation or professional consequences. Data security protocols encompassed secure data storage and transmission protocols, de-identification procedures for sensitive information, limited access protocols for research team members, and transparent data retention and destruction policies. Faculty focus groups were conducted and de-identified externally by the SEHD Evaluation Center. Student focus groups were conducted and de-identified by a PhD student. Administrations were given anonymous surveys instead of focus groups to further protect identities. FCQ data was not provided in an intersectional manner; instructor variables (e.g. gender, race) were provided in separate tabs for further anonymity.

Methodological rigor was maintained through bias mitigation strategies including multiple researcher perspectives in data collection and analysis, triangulation across quantitative and qualitative findings, external review of methodological approaches, and transparent reporting of limitations and potential

biases in the research process. Validity and reliability were ensured through multiple data sources to ensure comprehensive understanding, systematic approach to identifying and addressing potential confounding variables, clear operational definitions for all key constructs, and replication of statistical analyses with different approaches to ensure robustness.

Resource Allocation and Team Structure

The project leveraged a diverse research team including senior faculty researchers with expertise in quantitative and qualitative methodologies, graduate student research assistants with specialized training, professional focus group facilitators, statistical analysis consultants, and administrative support for data management and participant coordination. Technology and tools for quantitative analysis included advanced statistical software packages for complex modeling, secure database systems for large-scale data management, and visualization tools for pattern identification and presentation. Qualitative analysis utilized professional transcription services with confidentiality agreements, qualitative data analysis software for systematic coding, and secure platforms for focus group facilitation and recording.

The methodology was designed to enable comprehensive integration of findings across all components. Regular research team meetings ensured ongoing dialogue between quantitative and qualitative findings, allowing for iterative refinement of analysis approaches and emerging insights to inform subsequent data collection phases. This multi-method approach provides a robust foundation for understanding both the measurable impacts and lived experiences related to FCQ use, creating a comprehensive evidence base for informed institutional decision-making about faculty evaluation practices.

IV. Findings

This comprehensive analysis of Faculty Course Questionnaires (FCQs) at CU Denver reveals systematic biases that consistently disadvantage certain faculty groups while creating institutional barriers to effective teaching and innovation. Through quantitative analysis of 29,564 courses taught by 3,881 instructors over seven years, combined with qualitative focus groups and administrative surveys, this study exposes how FCQs function as flawed instruments that perpetuate inequality rather than improve educational outcomes.

Quantitative Findings: The Data Speaks to Systematic Bias

Methodology and Scope

The quantitative analysis examined FCQ patterns across 29,564 courses taught by 3,881 instructors from 2018-2024, incorporating both course-level and instructor-level variables. Using a QuantCrit framework, researchers identified systematic bias as "consistent and repeatable deviation that consistently skews results in a particular direction" and "is not reduced by sample size."

Key Patterns of Systematic Disadvantage

Higher enrollment courses consistently receive lower FCQ ratings across nearly all items (correlations between -0.10 and -0.13), suggesting that faculty teaching larger classes are systematically penalized regardless of teaching quality.

Undergraduate courses show consistently lower ratings than graduate courses across most measures, with the pattern showing "Undergrad was lower than both Masters and Doc" on critical items including respect for diverse students, critical thinking assignments, and instructor availability.

STEM courses face systematic disadvantage, receiving lower ratings on all items except one - the use of numbers, graphs, and statistics. This creates a particularly troubling pattern where faculty in disciplines requiring rigorous quantitative work are penalized, with items 10, 16, 18, and 19 (Appendix A) showing medium effect sizes (the highest effects in this study).

The Faculty Identity Penalty

The most striking findings reveal how faculty identity characteristics systematically influence ratings:

BIPOC faculty consistently score 0.12-0.13 points lower on the critical omnibus questions used for merit and promotion decisions:

- Instructor effectiveness: White faculty average 4.21 vs. BIPOC faculty 4.08
- Course effectiveness: White faculty 4.09 vs. BIPOC faculty 3.97
- Course learning: White faculty 4.10 vs. BIPOC faculty 3.97

Male faculty receive systematically lower ratings than female faculty across most items, though with smaller differences between groups.

Adjunct and part-time instructors score significantly lower than tenure-track and Instructional, Research, and Clinical (IRC) faculty on nearly all measures, creating a two-tiered system that reinforces existing hierarchies.

First-year faculty face systematic disadvantage, scoring 0.13-0.15 points lower on omnibus items compared to established faculty, potentially derailing early careers during the most vulnerable period.

Most Problematic Items

Five FCQ items show the most concerning patterns of systematic bias (See Appendix A):

Q11 - "Connect my learning to 'real world' issues or life experiences" demonstrates bias across 12 of 13 categories examined, making it the most problematic single item in regards to bias.

Q3, Q10, Q13, and Q19 each show significant differences across 10 different categories, indicating concerning bias in student ratings.

Qualitative Findings: The Human Impact of Systematic Bias

Major Themes from Faculty and Student Focus Groups

Theme 1: Institutional Confusion and Misalignment

The most fundamental finding from focus groups reveals widespread confusion about FCQ purpose and use. One faculty member captured this confusion perfectly:

"What is that leading to exactly? Is it leading to greater trust in your ability to be a teacher? Is it leading to higher pay? Is it leading to financial compensation?... Is it to evaluate the ability of the teacher? Is it to evaluate how fun or good the course is, or how it meets student needs? Or is this tied to your own professional and career development as an instructor? How are our instructors guided, and how to use their results? Are they guided at all or is this just a formality that everybody has to do because we have to? These are questions that don't seem to be satisfactorily answered."

Students echoed this confusion, noting that questions were "pretty vague and generalized" and expressing uncertainty about "what exactly the questions were measuring."

Theme 2: The Innovation Penalty

FCQs actively discourage pedagogical innovation, creating a risk-averse teaching environment. Faculty consistently reported that innovative teaching practices result in lower FCQ scores, with one noting:

"I've read enough research to know that if you try something innovative in your classroom, you will get a dip in your FCQs for at least the next three semesters."

This creates a particularly insidious problem where faculty described seeing "new faculty doing all these great, wonderful things, and then they get these [negative] comments. Maybe it's one or two from a student that didn't particularly like that [innovation], and then they change it. They don't do it again."

Theme 3: Mental Health Crisis

The psychological impact of FCQs on faculty emerged as a significant concern. Faculty described needing emotional preparation to review results:

"It is quite damaging to faculty—new faculty in particular—that come in wanting to do this amazing job with their students. They're trying all these wonderful new things and then they get the FCQs, and they are—they read the comments, they see the scores, and it is just—it's damaging. It's so damaging."

Another faculty member explained their coping mechanism: "I can only look at FCQ results after I've steeled myself and given myself some emotional preparation."

Theme 4: The Evaluation Paradox

Faculty experience a devastating contradiction where they're told FCQs don't matter while simultaneously being evaluated on them:

"The other thing that really impacts you is when your chair or your evaluators tell you that FCQs do not matter. They tell you they do not matter, and they do. They come up all the time and it matters not just on a piece of paper. It matters in your paycheck, and it matters to you, your psychology and how you're teaching."

Theme 5: Contextual Blindness

FCQs fail to account for crucial contextual differences that fundamentally affect the learning experience. Faculty and students identified four critical areas where standardized evaluation creates unfair comparisons:

Course Size: Faculty noted the impossibility of providing identical experiences when "some faculty teach 500 students a semester while others teach 15." The return of assignments and personal feedback looks fundamentally different at scale.

Course Content: Faculty teaching required foundational courses or diversity-focused content face systematic disadvantage. One faculty member explained:

"I teach a class focused on race, power, and privilege. The entire class is about pushing concepts and reflection... It's a required class, and so I have comments on my FCQs that are more related to the individual and the challenge that they experienced with the content, than related to me as a teacher."

Course Modality: The failure to differentiate between online and in-person instruction creates unfair comparisons, with faculty noting that "questions on the FCQ seem to assume a traditional in-person classroom experience."

Disciplinary Context: Faculty in STEM fields face systematic disadvantage, with quantitative analysis showing lower ratings across nearly all FCQ items except those related to use of numbers and graphs,

suggesting that disciplines requiring rigorous analytical or mathematical content are penalized regardless of teaching quality.

Theme 6: Underrepresented Faculty Burden

Faculty from underrepresented groups face additional burdens in defending their scores during merit and promotion processes:

"As a woman in STEM I am going to have a certain percentage lower than my male counterparts just across the board. This is research I cite in my dossier when I go up for review... I think it is good to remind us —when it comes up to dossier time or merit review time— that that's really plainly said."

Theme 7: Student Advocacy and Frustration

Students revealed using FCQs primarily for extreme reactions - either very positive or very negative experiences. They expressed frustration with the tool's limitations:

"I think my biggest frustration with the FCQs is talking with current and past cohorts, and finding for certain classes...the needs of the students and near constant repetitive information... is not being addressed... What's the point?"

Administrative Perspectives: Leadership Acknowledges the Problems

Survey Results from Deans and Provost Office

The administrative survey revealed significant awareness of FCQ limitations while simultaneously showing continued reliance on these flawed measures.

Theme 8: Administrative Ambivalence

Administrators expressed divided opinions about FCQ effectiveness, with responses showing both recognition of problems and continued dependence on the system. One dean acknowledged that FCQs are "part of teaching effectiveness determinations in annual merit and promotions" while simultaneously noting they should only be considered as "an indication of effectiveness, not a definitive measure."

Theme 9: Recognized Bias Patterns

Seven administrators acknowledged seeing "challenges and biases mainly among minorities, women, challenging graders, and course format." One wrote explicitly:

"Yes. I am aware of some patterns suggesting that female or women instructors of color are most likely to be negatively scored or average below their counterparts. This is not likely the result of a pattern of poor instruction."

Theme 10: The Standardization Problem

Administrators acknowledged a fundamental lack of standardization in how FCQs are used across campus, with responses indicating "there is no standardized process for how FCQs are used" and processes varying significantly by department.

Theme 11: Overwhelming Desire for Alternatives

The most telling finding was that all but one administrator believed alternatives should be explored, with one responding enthusiastically: "PLEASE PLEASE PLEASE PLEASE PLEASE PLEASE (yes. 200%)."

Recommendations from administrators focused on:

- Earlier timing of evaluations
- Better feedback mechanisms for instructors
- Pre-test/post-test approaches
- Holistic assessment combining multiple evaluation methods
- Course-specific surveys rather than standardized instruments

Major Institutional Themes

Theme 12: The Equity Crisis

The convergence of quantitative and qualitative data reveals FCQs as instruments that systematically perpetuate inequality. BIPOC faculty, women in STEM, first-year faculty, adjunct instructors, and those teaching required or challenging content face systematic disadvantage in a system that directly affects their pay, promotion, and job security.

Theme 13: The Assessment Paradox

FCQs function as high-stakes assessments for faculty while providing minimal useful feedback for improvement. As one faculty member noted:

"I think that the system that we have right now is really not set up to help faculty at all."

Theme 14: The Student Burden Transfer

The system inappropriately places the burden of faculty evaluation on students while providing them with inadequate tools and context. One faculty member captured this ethical concern:

"I think that faculty have been put in a position where they're trying to triangulate creating a learning experience for their students with this tool with their promotional and paycheck issues. It's not fair to subject student learning to those kinds of pressures and it's not fair to the faculty to have their promotion and tenure rely so extensively on such a corrupt and inaccurate tool."

Conclusions and Implications

This comprehensive analysis reveals FCQs at CU Denver as fundamentally flawed instruments that systematically disadvantage specific faculty groups while failing to serve their purported purposes of

improving instruction or fairly evaluating teaching effectiveness. The quantitative data demonstrates clear patterns of bias, the qualitative findings reveal significant institutional dysfunction, and administrative perspectives acknowledge these problems while maintaining dependence on the flawed system.

Key Findings Summary:

1. **Systematic bias is pervasive and measurable**, consistently disadvantaging BIPOC faculty, adjunct instructors, first-year faculty, and those teaching larger, undergraduate, or STEM courses.
2. **FCQs actively harm pedagogical innovation** by creating disincentives for trying new teaching approaches.
3. **The mental health impact on faculty is significant**, particularly for new and underrepresented faculty members.
4. **Institutional confusion about purpose and use** undermines any potential benefits of the system.
5. **Contextual factors are ignored**, creating unfair comparisons between fundamentally different teaching situations.
6. **Students are inappropriately burdened** with high-stakes faculty evaluation responsibilities without adequate tools or training.
7. **Administrative awareness of problems exists** alongside continued systemic dependence on flawed measures.

The evidence overwhelmingly suggests that FCQs in their current form cause more harm than good, functioning as instruments of institutional inequality rather than tools for educational improvement. The unanimous administrative interest in exploring alternatives, combined with the documented patterns of bias and institutional dysfunction, creates a compelling case for fundamental reform or replacement of the current system.

Rather than serving as measures of teaching effectiveness, FCQs appear to function as measures of student satisfaction with course content, instructor demographics, and institutional context - factors largely beyond individual faculty control. This misalignment between stated purpose and actual function represents a fundamental failure that demands immediate institutional attention and reform.

V. Bias Mitigation Strategies

A. Statistical Adjustment Techniques

Statistical adjustments are a critical step in addressing documented biases within FCQ systems. Effective models should account for course-level variables, instructor demographics, and contextual factors that influence student ratings but are beyond faculty control. Our analysis also indicates that FCQ items can be reliably grouped into different composite measures—such as an overall score, an “instructor effectiveness” mean (teacher-focused items), and a “course effectiveness” mean (course-focused items)—to provide more nuanced and equitable evaluations.

B. Contextual Interpretation Frameworks

Developing robust contextual interpretation frameworks will enable administrators to better understand FCQ results within their appropriate context. These frameworks must consider class size, course level, discipline-specific factors, and institutional context when interpreting evaluation data.

C. Recommendation for Question Redesign

Current FCQ questions require fundamental redesign to focus on pedagogical effectiveness rather than student satisfaction. Questions should be developed through evidence-based approaches that prioritize learning outcomes and teaching practices over subjective preference measures.

VI. Proposed FCQ Transformation Model

A. Holistic Evaluation Approach

The transformation model proposes a comprehensive approach that moves beyond single-metric evaluation systems toward multifaceted assessment methods that capture the complexity of effective teaching.

B. Complementary Assessment Methods

1. Peer Observations

Structured peer observation programs provide valuable insights into teaching effectiveness through professional colleague review. These observations should follow standardized protocols while allowing for discipline-specific considerations.

2. Teaching Portfolios

Teaching portfolios offer faculty the opportunity to present evidence of their pedagogical approach, innovation, and effectiveness through curated documentation of their teaching practice and student learning outcomes.

3. Self-Reflection Mechanisms

Incorporating structured self-reflection processes enables faculty to engage in continuous improvement while providing valuable insights into their teaching philosophy and development goals.

C. Implementation Roadmap

The implementation roadmap outlines a systematic approach to transforming current evaluation systems, including timeline considerations, resource requirements, and stakeholder engagement strategies.

VII. Alignment with Institutional Goals

A. Equity-Serving Institution Principles

The proposed transformation aligns with institutional commitments to equity by addressing documented biases that disproportionately impact BIPOC faculty, adjunct instructors, and other underrepresented groups.

B. Supporting Faculty Development

Rather than functioning solely as evaluative tools, the reformed system prioritizes faculty development and pedagogical growth through constructive feedback and professional support.

C. Enhancing Student Learning Outcomes

The ultimate goal of any evaluation system should be the improvement of student learning outcomes through enhanced teaching effectiveness and pedagogical innovation.

VIII. Implementation Plan

A. Phased Rollout Strategy

Implementation will proceed through carefully planned phases, beginning with pilot programs and gradually expanding to full institutional adoption based on evidence and feedback.

B. Pilot Program Design

Pilot programs will test proposed alternatives in controlled environments, allowing for refinement and adjustment before broader implementation.

C. Continuous Improvement Framework

The implementation plan includes mechanisms for ongoing assessment and improvement of the evaluation system based on evidence and stakeholder feedback.

IX. Recommendations

Recommendation # 1: Suspend High-Stakes Use of Current FCQs Pending Comprehensive Reform

The evidence presented in this report demonstrates that current Faculty Course Questionnaires systematically disadvantage specific faculty groups while failing to serve their stated purposes of improving instruction or fairly evaluating teaching effectiveness. The quantitative analysis of 29,564 courses taught by 3,881 instructors reveals that BIPOC faculty consistently score 0.12-0.13 points lower on critical omnibus questions used for promotion and merit decisions, first-year faculty face systematic disadvantage with scores 0.13-0.15 points lower than established faculty, and larger enrollment courses receive consistently lower ratings regardless of teaching quality. These patterns represent systematic bias

that "consistently skews results in a particular direction" and "is not reduced by sample size," creating fundamentally unfair evaluation conditions.

Given the documented harm to faculty careers and mental health, particularly for new and underrepresented faculty members, we recommend the immediate suspension of high-stakes use of current FCQs for promotion, tenure, and merit decisions until comprehensive reforms are implemented. This suspension should remain in effect until alternative evaluation methods are developed and validated through pilot programs. During this transitional period, departments should rely on peer evaluations, teaching portfolios, self-reflection mechanisms, and direct assessment of student learning outcomes for personnel decisions. To support faculty development during this transition, we recommend establishing optional student feedback mechanisms such as mid-term course evaluations that provide timely, actionable feedback for current instruction. Additionally, the Center for Excellence in Teaching and Learning should develop and recommend various forms of course evaluation tools that individual faculty can voluntarily utilize for improving their teaching practice, separate from any high-stakes personnel decisions. The institution must acknowledge that continuing to use demonstrably biased evaluation tools for high-stakes decisions violates principles of equity and fairness that are central to our institutional mission.

Recommendation # 2: Implement Comprehensive Bias Awareness Training and Standardized Interpretation Guidelines

The administrative survey results reveal significant awareness of FCQ limitations while simultaneously showing continued reliance on these flawed measures. Seven administrators acknowledged seeing "challenges and biases mainly among minorities, women, challenging graders, and course format," yet there exists no standardized process for how FCQs are used across campus. This institutional dysfunction creates arbitrary and inequitable evaluation conditions where identical scores may be interpreted differently across departments, schools, and colleges.

We recommend implementing ways for administrators to be involved in faculty evaluation processes in terms of mitigating bias, including deans, associate deans, department chairs, and promotion and tenure committee members. This training should address the documented patterns of bias in student evaluations, appropriate statistical interpretation of FCQ data, and recognition of contextual factors that influence evaluation scores. Additionally, the institution must establish clear, standardized guidelines for FCQ interpretation that account for course size, discipline-specific factors, instructor demographics, and other variables identified in our analysis. These guidelines should explicitly address how bias patterns affect different faculty groups and require written justification when FCQ scores are used in personnel decisions. The training program should be developed collaboratively with the Office of Equity and the Faculty Assembly, with annual refresher sessions required for all personnel involved in faculty evaluation.

Recommendation # 3: Develop Context-Specific Evaluation Tools and Flexible Assessment Methods

The qualitative findings reveal that FCQs fail to account for crucial contextual differences that fundamentally affect the learning experience. Faculty and students identified four critical areas where standardized evaluation creates unfair comparisons: course size (faculty teaching 500 students versus 15

students cannot provide identical experiences), course content (required foundational courses or diversity-focused content face systematic disadvantage), course modality (online versus in-person instruction), and disciplinary context (STEM courses show systematic disadvantage across most evaluation items). The current one-size-fits-all approach ignores these fundamental differences and creates systematic bias against faculty teaching in challenging contexts.

We recommend developing context-specific evaluation tools that recognize the diversity of teaching situations across the institution. Large enrollment courses should be evaluated differently than small seminars, with questions focused on appropriate pedagogical approaches for each context. Required courses should be assessed using different criteria than elective courses, acknowledging that student satisfaction may be lower in mandatory classes regardless of teaching quality. STEM courses should incorporate discipline-specific evaluation criteria that recognize the unique challenges of teaching quantitative content. Online and hybrid courses require evaluation tools specifically designed for these modalities rather than assuming traditional in-person classroom experiences. The development of these tools should be faculty-led within each college and department, with input from students and validation through pilot testing before implementation.

Recommendation # 4: Establish Multi-Source Evaluation Systems Emphasizing Faculty Development

The focus group findings reveal that FCQs function as high-stakes assessments for faculty while providing minimal useful feedback for improvement. Faculty consistently reported that the system is "not set up to help faculty at all" and that innovative teaching practices result in lower FCQ scores, creating a risk-averse teaching environment. Students expressed frustration with the tool's limitations, noting that questions were "pretty vague and generalized" and expressing uncertainty about "what exactly the questions were measuring." This fundamental misalignment between stated purpose and actual function represents a system failure that requires comprehensive reform.

We recommend establishing multi-source evaluation systems that prioritize faculty development and pedagogical growth over punitive assessment. These systems should combine structured peer observations conducted by disciplinary colleagues, comprehensive teaching portfolios that allow faculty to present evidence of their pedagogical approach and student learning outcomes, regular self-reflection mechanisms that enable continuous improvement, and direct assessment of student learning rather than satisfaction measures. The peer observation component should follow standardized protocols while allowing for discipline-specific considerations, with observers trained in effective feedback practices. Teaching portfolios should include evidence of pedagogical innovation, student learning outcomes, curricular development, and reflective analysis of teaching practices. Self-reflection mechanisms should be structured to promote continuous improvement while providing valuable insights into teaching philosophy and development goals. This multi-source approach will provide a more comprehensive and fair assessment of teaching effectiveness while supporting faculty professional development.

Recommendation # 5: Implement Strategic Timing Reforms and Meaningful Feedback Mechanisms

The current FCQ system provides feedback too late to benefit either instructors or students in the course being evaluated. Faculty focus groups revealed that they can "only look at FCQ results after [they've] steeled themselves and given [themselves] some emotional preparation," indicating that the current system creates psychological harm rather than constructive feedback. Students noted using FCQs primarily for extreme reactions and expressed frustration that their feedback does not lead to meaningful improvements in their educational experience.

We recommend implementing strategic timing reforms that provide multiple opportunities for feedback throughout the academic term. Mid-semester feedback should be collected using brief, course-specific questionnaires that allow faculty to make adjustments during the current term. End-of-term evaluations should focus on broader pedagogical effectiveness and learning outcomes rather than satisfaction measures. Faculty should receive training on how to interpret and respond to student feedback constructively, with support from the Center for Excellence in Teaching and Learning. The feedback process should be redesigned to benefit both faculty development and student learning, with clear communication to students about how their input will be used to improve educational experiences. This approach will create a more collaborative and constructive evaluation environment that serves both faculty and student needs.

Recommendation # 6: Create Pilot Programs for Alternative Evaluation Methods

The administrative survey revealed that all but one administrator believed alternatives should be explored, with one responding enthusiastically: "PLEASE PLEASE PLEASE PLEASE PLEASE PLEASE (yes. 200%)." This overwhelming desire for alternatives, combined with the documented problems of the current system, creates an opportunity for systematic experimentation with improved evaluation methods. However, any changes must be carefully tested and validated before full implementation to ensure they achieve intended goals without creating new forms of bias or institutional dysfunction.

We recommend establishing pilot programs in select departments and colleges to test alternative evaluation methods before broader implementation. These pilots should include pre-test/post-test approaches to measure student learning outcomes, course-specific surveys rather than standardized instruments, narrative feedback systems that provide more detailed insights into teaching effectiveness, and holistic assessment methods that combine multiple evaluation sources. Each pilot program should be designed collaboratively with faculty participants and include clear metrics for success, regular assessment of outcomes, and mechanisms for refinement based on evidence and feedback. The pilot programs should run for a minimum of two academic years to allow for comprehensive evaluation and adjustment. Results from these pilots will inform the development of permanent alternative evaluation systems that better serve faculty development and student learning outcomes.

Recommendation # 7: Establish Ongoing Research and Continuous Improvement Framework

The patterns of bias and institutional dysfunction identified in this report represent systemic problems that require ongoing attention and monitoring. The intersection of faculty demographics, course characteristics, and institutional context creates complex evaluation challenges that cannot be addressed through one-time reforms. Additionally, the rapidly evolving landscape of higher education, including

changes in student populations, pedagogical approaches, and institutional priorities, requires evaluation systems that can adapt to new challenges and opportunities.

We recommend establishing a permanent Faculty Evaluation Research Committee charged with ongoing monitoring of evaluation system effectiveness, bias patterns, and institutional impact. This committee should include faculty representatives from each college, student representatives, and administrative support from the Office of Institutional Research and Effectiveness. The committee should conduct annual analysis of evaluation data to identify emerging bias patterns, assess the effectiveness of implemented reforms, and recommend adjustments based on evidence and stakeholder feedback. The committee should also stay current with research on faculty evaluation best practices and pilot new approaches as they emerge in the literature. This ongoing research and improvement framework will ensure that evaluation systems continue to evolve in response to evidence and institutional needs rather than remaining static and potentially problematic.

Recommendation # 8: Align Faculty Evaluation Reform with Institutional Equity Goals and Strategic Planning

The documented patterns of bias in FCQ systems directly contradict the institution's stated goals of becoming an equity-serving institution. The systematic disadvantage faced by BIPOC faculty, first-year faculty, adjunct instructors, and those teaching in challenging contexts represents a fundamental misalignment between evaluation practices and institutional values. The mental health impact on faculty, particularly new and underrepresented faculty members, undermines efforts to create an inclusive and supportive academic environment.

We recommend aligning faculty evaluation reform with broader institutional equity goals and strategic planning initiatives. This alignment should include explicit recognition of evaluation bias as an equity issue requiring institutional attention and resources, integration of evaluation reform into diversity, equity, and inclusion planning processes, and allocation of sufficient resources to support comprehensive reform implementation. The reform process should be coordinated with other institutional initiatives focused on faculty success, student learning outcomes, and organizational culture change. Additionally, the institution should establish clear accountability mechanisms for evaluation reform, including regular reporting on progress toward equity goals and assessment of reform effectiveness. This comprehensive approach will ensure that evaluation system changes contribute to broader institutional transformation rather than operating in isolation.

Draft Implementation Plan - Fall 2025 Start

Phase 1: Report Presentation and Initial Approval (September - October 2025)

September 2025

Tasks:

- Finalize working group report with comprehensive recommendations
- Prepare executive summary for Faculty Assembly and administrative review
- Prepare presentation materials for October 7th Faculty Assembly meeting
- Brief Faculty Assembly Budget Priorities Committee on recommendations
- Coordinate with Office of the Provost for provost briefing preparation

October 2025

Tasks:

- October 7: Present final report to Faculty Assembly for formal review and endorsement
- Conduct provost briefing on FCQ reform recommendations and institutional context
- Begin stakeholder engagement with deans and department chairs
- Establish communication strategy with UComm for campus-wide announcement
- Form FCQ Reform Implementation Committee with faculty, student, and administrative representatives

Phase 2: Immediate Harm Reduction and System Preparation (November 2025 - January 2026)

November 2025

Tasks:

- Implement immediate suspension of high-stakes FCQ use for promotion, tenure, and merit decisions
- Develop interim evaluation guidelines for personnel decisions during transition period
- Begin design of bias awareness training curriculum with Office of Equity
- Establish Faculty Evaluation Research Committee with representatives from each college
- Initiate development of context-specific evaluation tools pilot programs
- Create communication materials explaining changes to campus community

December 2025

Tasks:

- Launch bias awareness training pilot with select administrators
- Begin development of standardized interpretation guidelines for FCQ use
- Establish partnerships with Center for Excellence in Teaching and Learning for faculty development support
- Design multi-source evaluation system framework with faculty input
- Create timeline for pilot program implementation in select departments
- Develop feedback collection mechanisms for reform process

January 2026

Tasks:

- Complete bias awareness training for all current administrators involved in faculty evaluation
- Finalize standardized interpretation guidelines for any continued FCQ use
- Establish pilot program departments and colleges for alternative evaluation methods
- Design peer observation protocols and training materials
- Create teaching portfolio guidelines and examples
- Prepare materials for spring 2026 faculty development workshops

Phase 3: Pilot Program Launch and Faculty Development (February - June 2026)

February 2026

Tasks:

- Launch pilot programs for alternative evaluation methods in selected departments
- Begin faculty development workshops on portfolio development and peer observation
- Implement mid-semester feedback mechanisms in pilot programs
- Establish data collection protocols for pilot program assessment
- Create support systems for faculty participating in pilot programs
- Begin development of context-specific evaluation tools for different course types

March 2026

Tasks:

- Continue faculty development workshops across campus
- Monitor pilot program implementation and provide ongoing support
- Collect preliminary feedback from pilot program participants
- Refine peer observation protocols based on initial experience
- Develop training materials for student feedback literacy
- Create resources for faculty interpreting and responding to feedback

April 2026

Tasks:

- Conduct mid-pilot assessment of alternative evaluation methods
- Refine teaching portfolio guidelines based on faculty feedback
- Expand bias awareness training to all faculty involved in peer evaluation
- Develop context-specific evaluation tools for STEM courses
- Create evaluation instruments for large enrollment courses
- Establish protocols for online and hybrid course evaluation

May 2026

Tasks:

- Collect comprehensive feedback from pilot program participants
- Analyze preliminary data on alternative evaluation method effectiveness
- Refine multi-source evaluation system based on pilot experience
- Develop context-specific tools for required and elective courses
- Create evaluation protocols for diversity-focused course content
- Prepare interim report on pilot program outcomes

June 2026

Tasks:

- Complete first-year assessment of pilot programs
- Analyze quantitative and qualitative data from alternative evaluation methods
- Refine implementation protocols based on pilot experience
- Develop recommendations for expanded implementation
- Create faculty development curriculum for ongoing support
- Prepare summer planning for expanded implementation

Phase 4: System Refinement and Expansion Planning (July - September 2026)

July 2026

Tasks:

- Conduct comprehensive evaluation of pilot program outcomes
- Refine alternative evaluation methods based on evidence and feedback
- Develop expansion plan for additional departments and colleges
- Create standardized training materials for broader implementation
- Establish ongoing support systems for faculty and administrators
- Design long-term assessment framework for evaluation system effectiveness

August 2026

Tasks:

- Finalize refined evaluation protocols for expanded implementation
- Develop comprehensive training curriculum for faculty and administrators
- Create technology infrastructure for new evaluation systems
- Establish partnerships with additional campus offices for ongoing support
- Design communication strategy for expanded implementation
- Prepare materials for fall 2026 expanded pilot launch

September 2026

Tasks:

- Complete preparation for expanded pilot implementation
- Train additional faculty and administrators in new evaluation methods
- Establish ongoing research protocols for continuous improvement
- Create accountability mechanisms for evaluation reform progress
- Develop annual assessment schedule for evaluation system effectiveness
- Prepare for transition to broader institutional implementation

Phase 5: Ongoing Implementation and Continuous Improvement (Fall 2026 and Beyond)

October 2026 - December 2026

Tasks:

- Launch expanded pilot programs in additional departments and colleges
- Implement ongoing faculty development programming
- Establish regular assessment schedule for evaluation system effectiveness
- Create feedback loops for continuous improvement
- Monitor bias patterns in new evaluation systems
- Develop long-term sustainability plan for evaluation reform

Ongoing Responsibilities

Tasks:

- Conduct annual analysis of evaluation system effectiveness and bias patterns
- Provide ongoing faculty development and support
- Refine evaluation methods based on evidence and feedback
- Maintain alignment with institutional equity goals and strategic planning
- Assess impact on faculty satisfaction, student learning outcomes, and institutional culture
- Adjust implementation based on emerging best practices and institutional needs

Resource Requirements and Support Structure

Key Personnel

- FCQ Reform Implementation Committee (faculty, student, and administrative representatives)
- Faculty Evaluation Research Committee (ongoing monitoring and research)
- Training coordinators for bias awareness and faculty development
- Technology support for new evaluation systems
- Communication support for campus-wide implementation

Institutional Support

- Office of the Provost leadership and coordination
- Office of Equity partnership for bias awareness training
- Center for Excellence in Teaching and Learning faculty development support
- Office of Institutional Research and Effectiveness data analysis and research
- UComm communication and outreach support
- Information Technology infrastructure and system support

Success Metrics

- Reduction in bias patterns across faculty demographic groups
- Increased faculty satisfaction with evaluation processes
- Improved quality and usefulness of feedback for faculty development
- Enhanced student learning outcomes and satisfaction
- Successful implementation of alternative evaluation methods
- Sustained institutional commitment to equity-serving evaluation practices

---

References

Alhija, F. N. A. (2017). Teaching in higher education: teaching through students' lens. *Studies in Educational Evaluation*, 54, 4-12.

Angelo, T. A., & Cross, K. P. (2011). *Classroom Assessment Techniques* (3rd ed.). John Wiley & Sons.

Arreola, R. A. (2004). *Developing a comprehensive faculty evaluation system*. https://www.academia.edu/download/30508048/facultyevalsystem.pdf

Bell, A., & Mladenovic, R. (2008). The benefits of peer observation of teaching for tutor development. *Higher Education*, 55(6), 735–752.

Benton, S. L., & Cashin, W. E. (2012). *Student ratings of teaching: A summary of research and literature*. IDEA Paper.

Beran, T. N., & Rokosh, J. L. (2009). The consequential validity of student ratings: What do instructors really think? *Alberta Journal of Educational Research*, 55(4), 497-511.

Beran, T. N., Violato, C., Kline, D., & Frideres, J. (2005). The utility of student ratings of instruction for students, faculty, and administrators: A "consequential validity" study. *Canadian Journal of Higher Education*, 35, 49–70.

Berk, R. A. (n.d.). Top 20 Strategies to Increase the Online Response Rates of Student Rating Scales. https://www.ronberk.com/articles/2012_top20.pdf

Centra, J. A. (2000). Evaluating the teaching portfolio: A role for colleagues. *New Directions for Teaching and Learning*, 83, 87-93.

CU Faculty Council. (2016, September). *Minutes of the CU Faculty Council meeting*. University of Colorado. Retrieved from https://www.cu.edu/doc/fc-sept-2016-minutespdf?download=true

Dana, H., Morrissette, S., & Nelson, S. (2023). Post-secondary student evaluations of teachers: The debate of usefulness continues. *Journal of Education and Learning*, 12(3), 54-61.

Davidovitch, N., & Eckhaus, E. (2019). Teaching students to think - faculty recommendations for teaching evaluations employing automated content analysis. *International Journal of Higher Education*. https://doi.org/10.5430/IJHE.V8N3P83

Davidovitch, N., & Eckhaus, E. (2019). The influence of student evaluations of teachers on the approval of tenure. *Athens Journal of Education*, 6(1), 49-62.

Esarey, J., & Valdes, N. (2020). Unbiased, reliable, and valid student evaluations can still be unfair. *Assessment & Evaluation in Higher Education*, 45(8), 1106–1120.

Ferguson, K., Frost, L., & Hall, D. (2020). Predicting teacher anxiety, depression, and job satisfaction. *Journal of Teaching and Learning*, 8(1), 27-42.

Ferguson, S. L., Walpole, M., & Fall, M. S. (2020). Achieving statistics self-actualization: Faculty survey on teaching applied social statistics. *Statistics Education Research Journal*, 19(2), 57–75.

Golding, C., & Adam, L. (2016). Evaluate to improve: Useful approaches to student evaluation. *Assessment & Evaluation in Higher Education*, 41(1), 1-14.

Gonzales, L. D., & Terosky, A. L. (2016). From the faculty perspective: Defining, earning, and maintaining legitimacy across academia. *Teachers College Record*, 118(7), 1–44.

Gonzales, L. D., & Terosky, A. L. (2016). On their own terms: Women academic leaders' professional identity development. *Studies in Higher Education*, 41(3), 571-585.

Hammer, R., Peer, E., & Babad, E. (2018). Faculty attitudes about student evaluations and their relations to self-image as teacher. *Social Psychology of Education: An International Journal*, 21(3), 517–537.

Hu, C. (2020). Understanding college students' perceptions of effective teaching. *International Journal of Teaching and Learning in Higher Education*, 32(2), 318-328.

Kember, D., Leung, D. Y. P., Jones, A., Loke, A. Y., McKay, J., Sinclair, K., Tse, H., Webb, C., Yuet Wong, F. K., Wong, M., & Yeung, E. (2000). Development of a questionnaire to measure the level of reflective thinking. *Assessment and Evaluation in Higher Education*, 25(4), 381–395.

Kreitzer, R. J., & Sweet-Cushman, J. (2022). Evaluating student evaluations of teaching: A review of measurement and equity bias in SETs and recommendations for ethical reform. *Journal of Academic Ethics*, 20(1), 73–84.

Liu, S., Keeley, J., & Buskist, W. (2016). Chinese college students' perceptions of excellent teachers across three disciplines: Psychology, chemical engineering, and education. *Teaching of Psychology*, 43(1), 70-74.

Lohman, L. (2021). Evaluation of university teaching as sound performance appraisal. *Studies in Educational Evaluation*, 70(101008), 101008.

MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4), 291-303.

Marsh, H. W. (2007). Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *Journal of Educational Psychology*, 99, 775-790.

McKeachie, W. J. (1997). Student ratings: The validity of use. *The American Psychologist*, 52(11), 1218–1225.

Mitchell, K. M., & Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science & Politics*, 51(3), 648-652.

Murray, J. P. (2005). Meeting the needs of new faculty at rural community colleges. *Community College Journal of Research and Practice*, 29(3), 215–232.

Penny, A. R., & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: A meta-analysis. *Review of Educational Research*, 74(2), 215–253.

Safadi, S. A., Bakar, K. A., Tarmizi, R. A., & Alwi, N. H. (2013). Faculty perception of improvements to instructional practices in response to student ratings. *Educational Assessment, Evaluation and Accountability*, 25(2), 143-153.

Stark, P., & Freishtat, R. (2014). An evaluation of course evaluations. ScienceOpen. Center for Teaching and Learning, University of California, Berkeley.

Stupnisky, R. H., BrckaLorenz, A., Yuhas, B., & Guay, F. (2018). Faculty members' motivation for teaching and best practices: Testing a model based on self-determination theory across institution types. *Contemporary Educational Psychology*, 53, 15–26.

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22-42.

Wheeler, D. W., Seagren, A. T., Becker, L. W., Kinley, E. R., Mlinek, D. D., Robson, K. J., Seldin, P., & Miller, J. E. (2009). *The academic portfolio: A practical guide to documenting teaching, research, and service* (2nd ed.). Jossey-Bass. https://www.academia.edu/download/97804183/j.0022-0337.2009.73.9.tb04802.x20230125-1-q52i45.pdf

Young, S., Rush, L., & Shaw, D. (2009). Evaluating gender bias in ratings of university instructors' teaching effectiveness. *International Journal for the Scholarship of Teaching and Learning*, 3(2), 1-14.

Appendix A
CU Denver Faculty Course Questionnaire (FCQ) Questions

The following are the standard FCQ questions used at CU Denver to assess instructor effectiveness and course quality. Students respond using Likert scales (1–5) or provide narrative feedback.

Core Instructor Items (Required)

Students rate the instructor on a scale from 1 (Hardly Ever) to 5 (Almost Always), with an option for N/A.

1. Demonstrated respect for diverse students and diverse points of view.
2. Clearly explained learning goals throughout the semester.
3. Gave projects, tests, or assignments that required original or creative thinking.
4. Provided feedback on my work that helped me improve my performance.
5. Effectively fostered student engagement.
6. Explained the grading criteria for assignments.
7. Was available to answer questions or provide assistance when needed.
8. Cared about my learning.

Core Course Items (Required)

Students rate their experiences on a scale from 1 (Strongly Disagree) to 5 (Strongly Agree), with an option for N/A.

9. Interact with other students in a respectful way.
10. Reflect on what I was learning.

11. Connect my learning to “real world” issues or life experiences.

12. Work and learn collaboratively with my classmates.

13. Contribute my ideas and thoughts.

14. Evaluate arguments, evidence, assumptions, and conclusions about key issues (be a critical thinker).

15. Connect, synthesize, and/or transform ideas into a new form (be a creative thinker).

16. Consider diverse perspectives (gender, political, ethnic, racial, etc.) during class or in assignments.

17. Revise my work based on instructor feedback.

18. Use numbers, graphs, and/or statistics in course assignments.

Core Narrative Response Items (Optional)

Students provide open-ended feedback.

19. Please offer constructive comments to your instructor on the most effective aspects of this course.

20. Please offer constructive comments to your instructor on the least effective aspects of this course.

Source: University of Colorado Denver. (2017). *Faculty Course Questionnaire: Fall 2017 Questions.* Retrieved from https://www.colorado.edu/fcq/sites/default/files/attached-files/fall17denverquestionsweb_0.pdf

Appendix B: FCQ Resolution

CU Denver FCQ Resolution

May 2023

Whereas, ample research evidences student evaluations of teaching (SET) like our FCQ are highly biased tools with respect to women and other marginalized groups and have little to no correlation with the quality of teaching1;

Whereas, despite this research, not only do FCQ's continue to play a central role in faculty evaluation, but also results are published on a public site (a practice far outside national norms) thus inviting a level of public scrutiny which our administrators do not have to endure;

Whereas, in 2019 the FCQ Office began mining our FCQ's for key words and related stem words including: "age, of color, creed, disability, gender expression, gender identity, national origin, political affiliation, political philosophy, pregnancy, race, religion, sex, sexual orientation, veteran status," reviewing and then reporting possible equity violations2, a process to which written evaluations of administrators are not subject;

Whereas, per Regent Law and Policy, the evaluation of faculty is the primary responsibility of faculty and yet our recognized shared governance groups were not apprised of this use of our FCQ's or consulted on the appropriateness of this process or asked to participate in this process;

Whereas, the use of this highly biased tool for a purpose for which it is not designed places faculty in additional jeopardy;

Whereas, Offices of Equity are now treating the FCQ as if students filed an equity report, when

they did not3;

Whereas, Offices of Equity are denying faculty the opportunity to call for an investigation, claiming an investigation cannot be done because of the anonymity of FCQ's, when in fact students can be identified and an investigation launched, thus denying faculty due process;

Whereas, students have effective avenues to file actual equity reports upon finding the need to do so;

Whereas, there is no system-level requirement that we use the current FCQ or that all faculty on the same campus use the same FCQ or that we collect written comments;

Therefore, be it resolved that we reject the current FCQ process.

Be it further resolved, that starting with the Summer 2023 administration of FCQ's and holding as long as FCQ's continue to be mined for keywords, all Canvas shells will have a banner with the following text to inform students (and faculty) of the actual FCQ processes:

The primary purpose of the FCQ is to strengthen teaching and learning on our campus. Therefore, constructive feedback is welcome. Know that while your name is withheld from faculty, you may be identified in certain instances. For example, all FCQ's are scanned for specific keywords and phrases indicating attacks, slurs or other prohibited behavior. This behavior may result in "having your FCQ response removed, your comment being referred to the Office of Institutional Equity and Compliance (OIEC) and possible disciplinary action" (FCQ Office Terms of Service). If you have concerns about possible equity violations, you should report them directly to the campus Office of Equity.

Be it further resolved, that in Fall 2023 Faculty Assembly will work in concert with the Provost's

Office to research and propose alternatives to the current FCQ process that are in line with Regent Law and Policy while protecting faculty and student rights, improving the faculty teaching evaluation process, and more effectively mitigating sources of bias.

Notes

1. Evaluating Student Evaluations of Teaching: a Review of Measurement and Equity Bias in SETs and Recommendations for Ethical Reform (https://philpapers.org/rec/KREESE)

In our analysis, two problematic and consistent findings predominate the literature. First, we find that scholars across disciplines and in numerous country contexts consistently reveal that SETs do not measure teaching effectiveness (Uttl et al., 2017; Benton & Cashin, 2012). That is, SETs are prone to measurement bias. Second, most of the literature indicates that men receive higher evaluative scores compared to women (see for instance, Basow & Silberg, 1987; MacNell et al., 2015; Mengel et al., 2018; Sidanius & Crane, 1989; Wigington et al., 1989). There is some evidence of discrimination towards other group as well, though it is less-well documented in the scholarship (as we will discuss below). In other words, SETs are also prone to equity bias.

The authors recommend the restriction or elimination of written comments (in our case we would need to eliminate as the FCQ Office has made clear it will not stop mining our comments):

Across all the studies in our sample, the clearest evidence of gender bias is in qualitative comments. Scholars employing content analysis of qualitative comments finds clear evidence of bias with women faculty and faculty of color are more likely to receive

negative comments about personality traits, appearance, mannerisms, competence, and professionalism compared to white men (Wallace et al., 2019). Furthermore, many faculty report particularly mean-spirited and cruel comments (Lindahl & Unger, 2010). Instead of asking for general "comments," assessments should direct students to provide feedback on certain experiences with the course, as this may reduce irrelevant and mean comments.

There are additional problems with qualitative comments beyond issues of bias. They are difficult to aggregate and have a low sample size (Himelein, 2018). Furthermore, they are not reliable—in fact, they frequently have contradictory feedback (Linse, 2017). Finally, even well-intentioned reviewers of qualitative comments may be susceptible to novelty bias (we are more likely to remember unexpected or uncommon findings) and negativity bias (the tendency to be influenced by negative information more than positive information) (Himelein, 2018). Comments that are anomalous or do not correlate with class averages on quantitative items should be disregarded.

2. 3. Per email communication with the FCQ Office

The standard letter on the CU Denver campus states, "Due to the anonymous nature of the comment in the FCQ, I am unable to follow up with the reporting party [student] to ask for clarification."