TQF Suggested Guidance for the use of FCQs* in Teaching Evaluation
[UPDATED 05/31/2022]

Executive Summary

The following includes a set of general principles, departmental considerations, and a series of scenarios for use of student evaluations of teaching (SETs or FCQs* at CU Boulder).

- FCQs are ideally used only in combination with other measures of teaching - combining multiple measures helps triangulate data and de-emphasizes FCQs as an independent measure for departmental decision making around faculty advancement.
- Forms of student feedback that can help to de-emphasize a focus on FCQ ratings include qualitative interpretation of FCQ student comments, classroom interviews, and solicitation of student letters (see some examples here).
- The data from FCQs should be contextualized around similar course types, the longitudinal development of faculty, among other valued contexts in the department.
- Use FCQ scores to look for outliers -- either rather high or rather low -- compared to the appropriate groups, or other courses from the instructor. Use these as a starting point for further inquiry.
- Contextualize the results in terms of response rate - make sure that the data you are looking at is arguably representative of the class(es) they are drawn from.
- A statement from the department and / or college leadership should provide clarification on how FCQs should be used and interpreted for teaching evaluations - such a statement should convey the limitations of FCQs (e.g., since these FCQ data are of limited value due to lack of comparative results, low response rate, or other reasons, we will use it in a limited fashion").

This document is intended to be a living document - we welcome your feedback. Please contact Cynthia Hampton (cynthia.hampton@colorado.edu) or Noah Finkelstein (noah.finkelstein@colorado.edu) if you have questions or suggestions.

Principles for implementation, use, and interpretation of FCQs

Please see "Using FCQs and Multiple Measures of Teaching for Annual Reviews" on the FCQ website by the Center for Teaching & Learning and the Office of Data Analytics (CTL/ODA) for additional information (many of these principles are derived from that resource). And see here for the full set of CU Boulder FCQ items, effective as of Fall 2020.

- FCQs are best used in triangulation with other data sources to identify strengths and weaknesses of teaching practices, rather than as a stand alone tool. Major personnel decisions should not be made based on FCQ data alone - when FCQs are used as part of the evaluation of teaching for personnel decisions, they must be accompanied by other means of evaluation (use multiple measures).
- Use this as an opportunity to deemphasize the role of FCQs for annual merit.
    - Incremental change is OK
    - Adjust as you go as you gather information on what has worked hasn't worked
    - Think of this as longer term change rather than needed to be instantaneously done and set
- FCQs ought to be used as a point of feedback. Focus on providing actionable feedback

*FCQ = Faculty Course Questionnaire, CU's end-of-semester student evaluations. For more information on the Teaching Quality Framework Initiative visit us at https://www.colorado.edu/teaching-quality-framework/.

to faculty that identify strengths and areas for improvement.

- Encourage use of student voice beyond FCQs, e.g.,
    - Student surveys (e.g., SALG, MUSIC model of motivation questionnaire, etc.)
    - Letters of recommendation from students (see here for some example guidelines for getting better student letters)
    - Classroom interviews (see here for example processes)
- Provide students detailed guidance (and training) on submitting FCQs and any other forms of data (example guidance can be found here)
- Rank order lists of faculty performance in a unit are discouraged. Rather, use a fixed number of bins (e.g., 3- 5)
    - Note that small differences in FCQ scores are not significant - FCQs probably should not be used to create more than 3 levels of achievement, with a wide range for the middle ("meets expectations"). Noting, however, that annual merit has 5 levels - consider how you will combine FCQs with additional sources of data to determine where faculty fall into the 5 bin system.
- Only use comparison to others (e.g., to a departmental/college/university-wide average) to flag faculty as falling below an acceptable range and to highlight faculty achieving well above average, otherwise faculty FCQ scores should be compared to their own scores over time, not to those of other faculty
- FCQ scores prior to fall 2020 (at CU) should not be compared directly to the FCQ scores in use currently - the items are not the same and the number of response options is different - rather, consider these independently in situations where you have both.
- Use FCQs primarily to look for "red flags" such as average ratings that are one point or more below the comparative average, and consistently low scores across courses or over a number of semesters.
    - However, be sure to look at the distribution of the ratings across the scale. There may be outliers—one or a few ratings that pull the average down. These outliers should be considered in context and should not be allowed to represent the collective views of students in the course.
- We do not recommend averaging all the FCQ items to get one score for the following reasons:
    - Moves us back in direction of relying on omnibus items (see BFA resolution BFA-M-1-040518 for more information on why we want to avoid this)
    - Suggests that all items are of equal importance / should be weighted equally
    - Assumes that the scale and discrimination of each item is the same.
    - Assumes similar thresholds for what counts as satisfactory across all items (e.g., it may be that some items everyone should score very high on so the threshold for cause for concern is higher, whereas there may be more of a range of acceptable scores for other items)
- Pay attention to course context. It may not be appropriate to combine/compare ratings across different course types as student feedback on, for example, lower division, high enrollment courses are likely to be quite different from those in smaller upper division/graduate level courses (see also the "What counts as a meaningful difference in FCQ scores" section here).

- Choose FCQ questions that are important (e.g., diversity) and meaningful for your discipline, provide actionable feedback to instructors, and tap into students' direct experiences or observations of practice
- Please do not discard items just because the average in your unit is low for a given item - consider this as an excellent opportunity to encourage faculty to improve in these areas. Instead, consider giving a grace period for anything that instructors weren't previously evaluated on, e.g., one year to address in classrooms before those items are used in evaluation.
- Use FCQs as the beginning of a conversation. E.g., where you have FCQ data that suggests red flags, dig deeper and use additional data sources to help triangulate on issue(s). Engage in conversation(s) with that faculty member to document ways to improve, then track improvement

Questions to consider / discuss as a department:

- What FCQ items are most important (e.g., diversity related items), applicable to your department/discipline, and align with a scholarly framework (e.g., TQF rubric, A&S QTI dimensions for defining teaching quality)?
- Will you use all of these items in evaluation or build in flexibility (e.g., could an instructor select the items they want to be reviewed on)? (noting that such flexibility should not allow instructors to avoid selecting items where they could improve)
- Should any FCQ items be weighted more strongly than others?
- Consider response rate - what will you do for FCQs from classes with a low response rate? Could you incentivize faculty to improve FCQ response rates? (additional guidance on improving response rates can be found here)
- How will you use FCQs to foster growth / provide actionable feedback to faculty struggling with teaching?
- What other forms of data (e.g., instructor self-reflection, peer observation/review, other forms of student feedback) do you have access to / might you incorporate into annual teaching evaluations?
- Are there any custom FCQ items that your unit would like to add? (note the deadline for adding custom items to the fall 2021 FCQs is October 29th)
- Are faculty and course instructors aware of visibility and access to open comment sections of the FCQs? (these are not made public like the other FCQ data but they are accessible by department leads, report administrators, and to the individual faculty member. If these comments are to be reviewed as part of the evaluation process, they will need to be downloaded and integrated as primary unit evaluation committees (PUECs) will not automatically have access to the comments)

Example models for using FCQs numerical results in annual merit

Each of the following models is designed to be one component of many elements of FCQ use -- i.e., they can be combined. Additionally, the FCQs ought only be used in combination with a suite of other data sources for determining merit (or making other major personnel decisions). For example, using a rubric that includes self-reflection and any peer review that occurred

along with FCQs to determine level of merit (e.g., see the GSLL merit rubric (noting FCQ items haven't been updated in this version).

Model A: Mean and Outliers. Use questions X, Y, Z (decided at the department level) that should be common across all instructors / courses. Note that faculty within some amount (e.g., 1 standard error on mean above and below or +/- 1 point) are in a proficient bucket. Take time to review cases that are outside the mean, and use these to suggest that practice is below or above expectations.

Model B: Instructor Focused Use. At the beginning of each term, faculty, in consultation with a mentor and/or evaluation team, identify which 3-8 FCQ questions are most relevant (e.g., questions where faculty would like to improve on or those aligned with course objectives and/or program learning outcomes (PLOs)). These agreed-upon questions are then used for tracking outcomes over time. Note instructors should not be allowed to avoid selecting items where their scores are low.

Model C: Course / Departmentally Determined Goals. Given either departmental-wide or faculty specific sets of FCQ questions, look for change over time of faculty practice in a given course, or level of courses.

Model D: For Faculty Reflection. Have faculty reflect on and discuss their FCQ performance and student comments in annual self-reflection (and consider providing specific guidance to all in the department -- which questions to review, to note changes over time, identify outliers, etc.)

Model E: Consider Response Rate for Courses. Provide faculty incentives for increasing response rate. This can include suggesting that too low a response rate may not meet expectations whereas very high response rates are exceeding expectations.

See also the A&S QTI Student Voice Resources for additional suggestions on the use of FCQs.