Supporting Information:

"Why Chemistry Instructors Are Shifting to Specifications Grading: Perceived Benefits and Challenges"

Ying Wang[a†], Haleigh Machost[a†], Brandon J. Yik[b], & Marilyne Stains[a]*

[†]Authors contributed equally to this work

[a]Department of Chemistry, University of Virginia, Charlottesville, VA, 22904, USA

[b]Department of Chemistry, University of Georgia, Athens, GA, 30602, USA

*Corresponding author email: mstains@virginia.edu

Contents

Section A. Additional Participant Data

Table S1. Number of potential participants identified through each method

Potential participant identification	Potential participants identified
Conference abstracts	26
Journal article publications	23
Snowball sampling	14
Social media posts	10
Online searches	5
Personal communications	4
Book chapters	3
Total	**85**

Table S2. Carnegie classifications of participants' institutions. Classifications include baccalaureate colleges and universities (BAC), master's colleges and universities – larger programs (M1), doctoral/professional universities (D/PU), doctoral universities – high level of research activity (R2), and doctoral universities – very high level of research activity (R1).

	BAC	M1	D/PU	R2	R1	Total
Public	2	1	-	2	6	**11**
Private	7	1	2	2	1	**13**
Total	**9**	**2**	**2**	**4**	**7**	**24**

Table S3. Course characteristics of participants. If instructors taught more than one specifications-graded course, they were asked to describe the course of their first implementation.

	Enrollment	Number of instructors
Lab	Small (<40)	0
	Medium (40-150)	1
	Large (150-1000)	2
	Very large (>1000)	2
	Total	**5**
Lecture	Small (<40)	4
	Medium (40-150)	4
	Large (150-1000)	2
	Very large (>1000)	1
	Total	**11**
Lab-Lecture	Small (<40)	9
	Medium (40-150)	3
	Large (150-1000)	1
	Very large (>1000)	0

	Total	**13**

Section B. Summary of IRR process

Table S4. Summary of changes to the codebook during the interrater reliability process

Iteration	Documents	Summary of changes post-inter-rater analyses
Round 1	178, 179, 180, 184, 186, 188, 189, 192, 195, 197	• Clarified definition of 'reduced student proficiency' • Clarified definition of 'increased instructor workload' • Added to definition of 'unfamiliar system for students' • Verbally clarified that challenges with training TAs falls under 'increased instructor workload' • Changed 'potential negative impact on instructor' to be specific to career affects • Changed 'lack of support' to include damage professional relationships without any affect on career • Added a code 'does not accommodate different groups of students' • Added a code 'maintaining rigor'
Round 2	102, 117, 118, 119, 120, 124, 125, 129, 145, 149	• Clarified the distinction between codes 'unfamiliar system to students' and 'student resistance;' correspondingly clarified the respective definitions • Clarified 'increased tension between instructor and student' to be tension as perceived by the instructor • Removed unused code 'increased student engagement' • Incorporated the code 'multiple opportunities to revise or retake' into the code 'increased flexibility'
Round 3	152, 153, 154, 163, 170, 171, 173, 174, 177	• No further changes
Final Review		• All instances of codes altered, as described above, were revisited to ensure consistency