Peer Assessment of Writing in Large Classes: Reliability, Validity, and Improving Student Attitudes

Catherine Rawn, Peter Graf, Janel Fergusson, & Bosung Kim

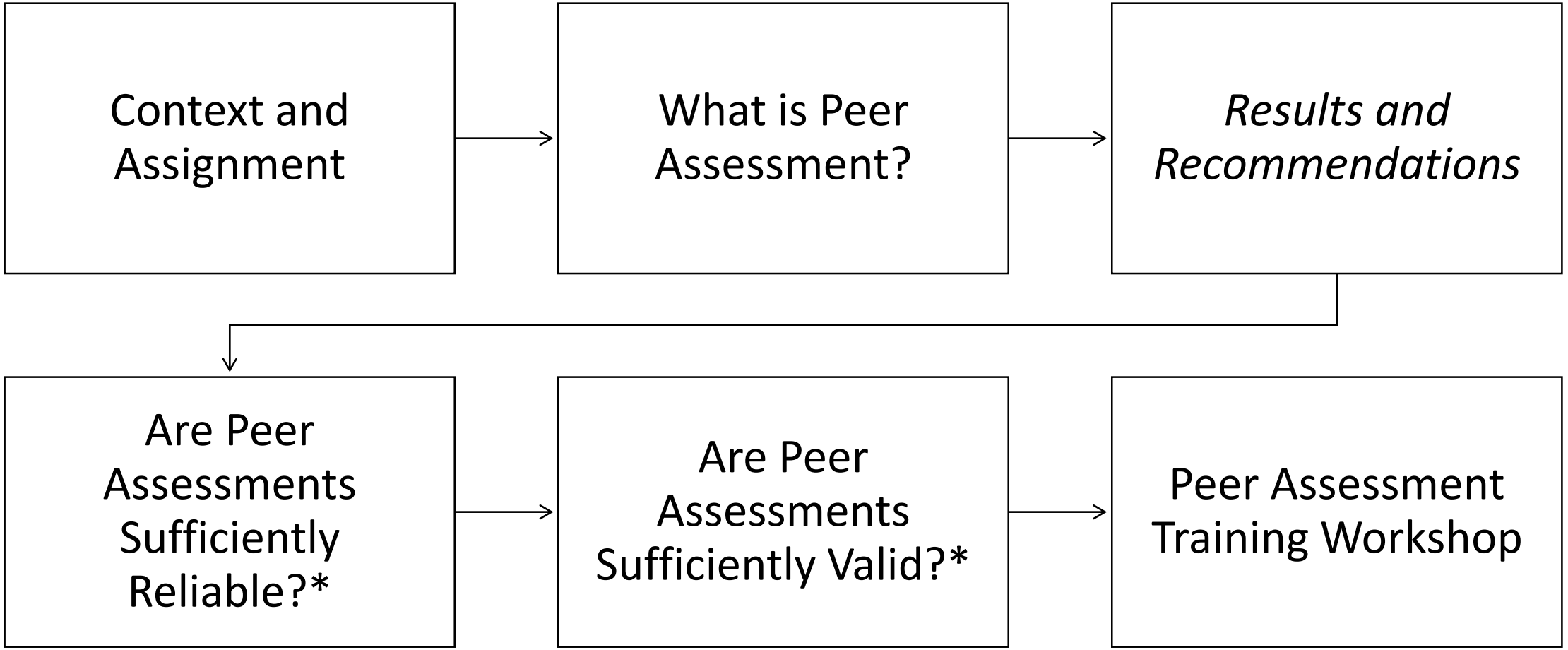
University of British Columbia, Vancouver Canada

Talk presented at the Society for the Teaching of Psychology's Annual Conference on Teaching, 2018

@cdrawn cdrawn@psych.ubc.ca

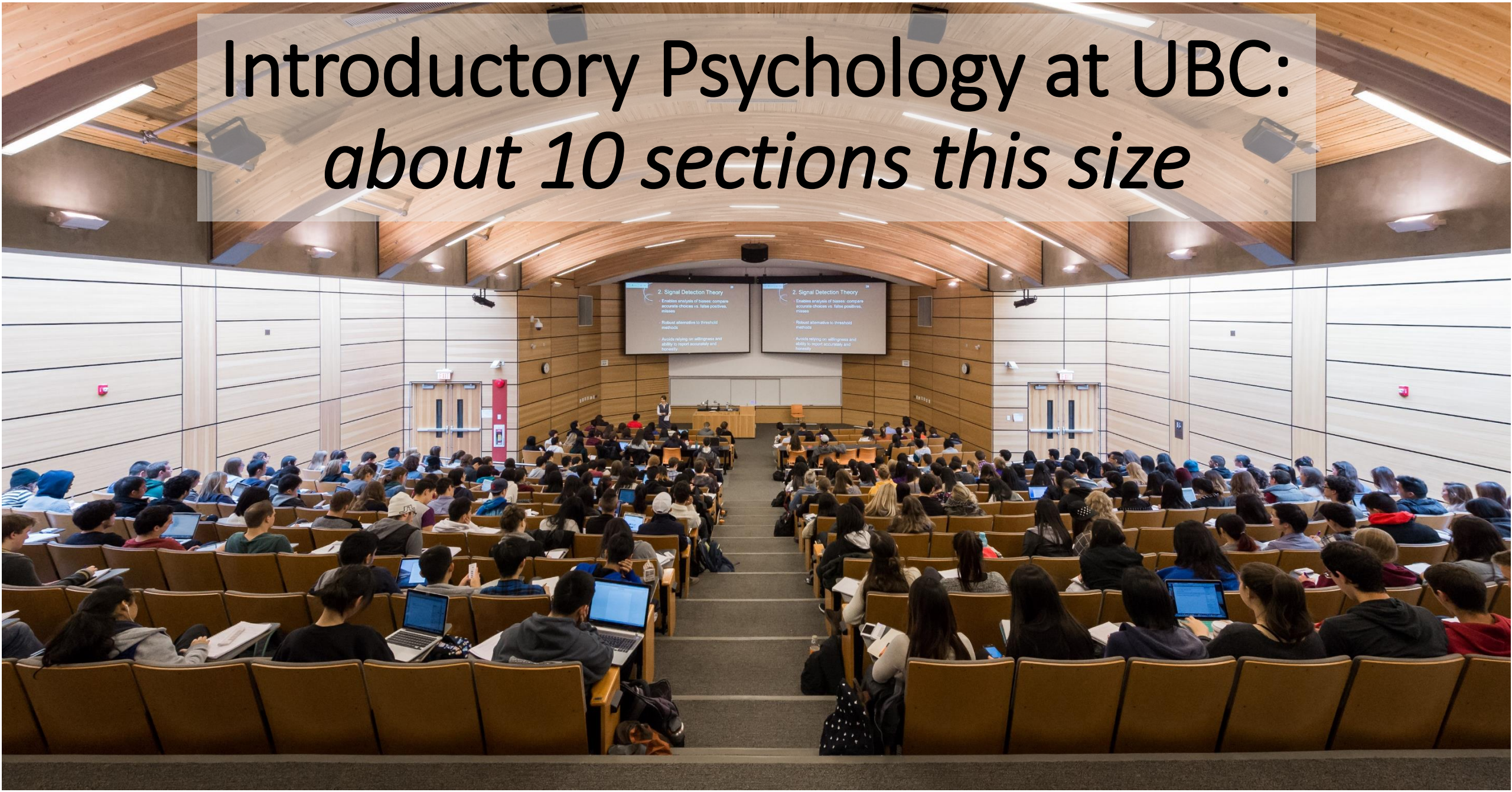
blogs.ubc.ca/catherinerawn/ peerassessment.arts.ubc.ca/

Today's plan



**for what? Always consider the context.*

Introductory Psychology at UBC: *about 10 sections this size*

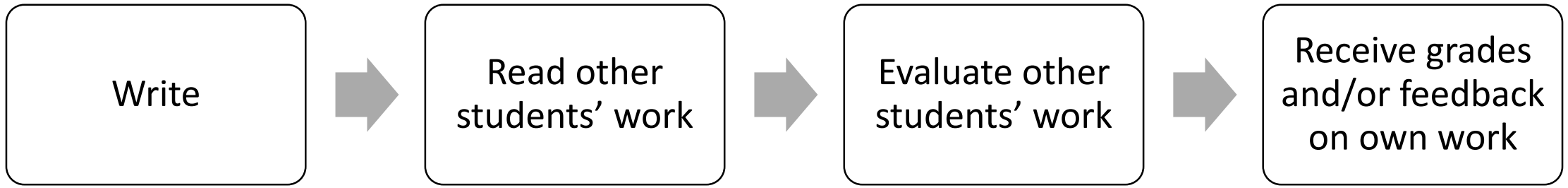


Assigning Writing takes Creativity

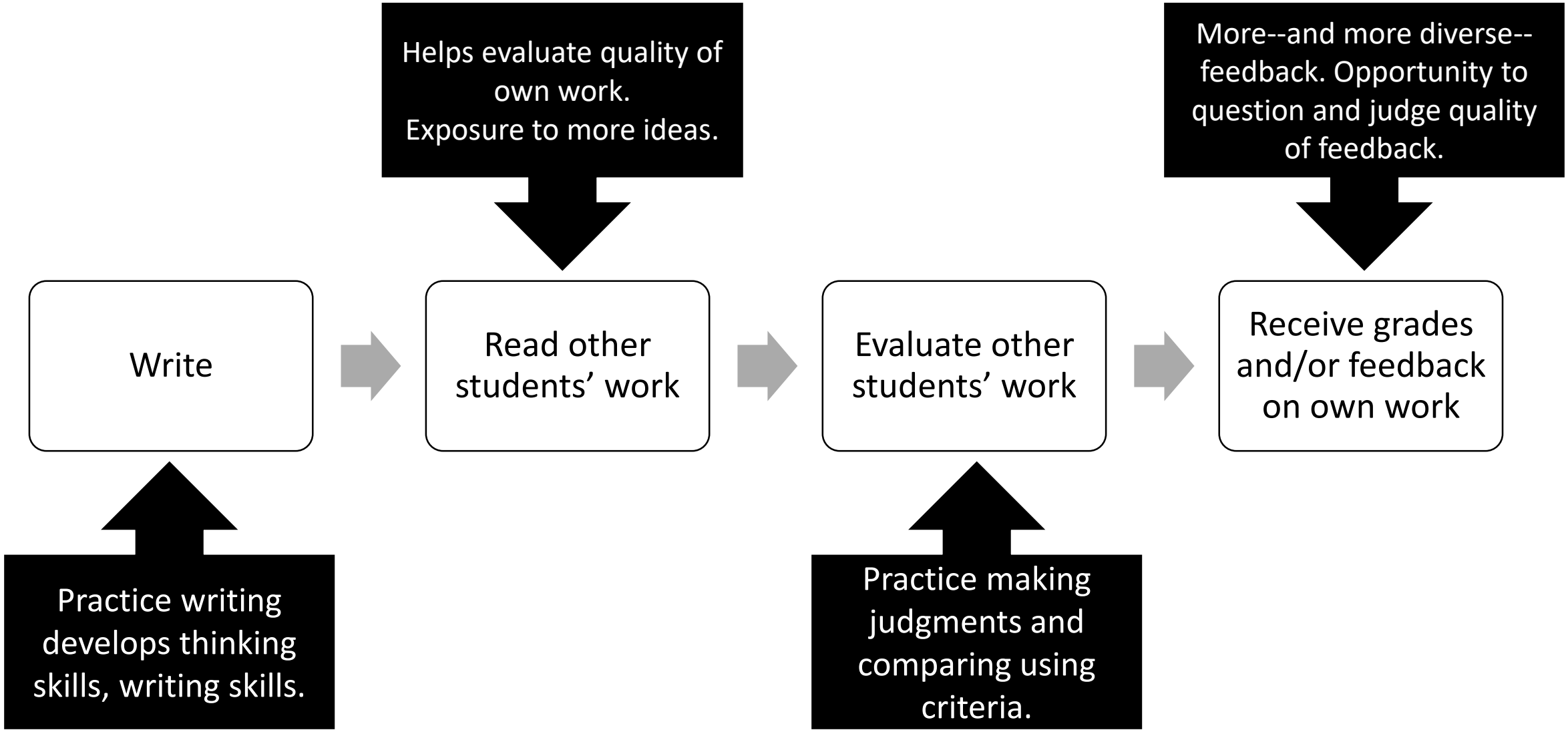
Mini-assignments helps students learn (Gingerich et al., 2014; Nevid et al., 2012; Nevid et al., 2017)

Assessing peers' work helps students learn (Dochy et al, 1999; Nicol et al., 2014)

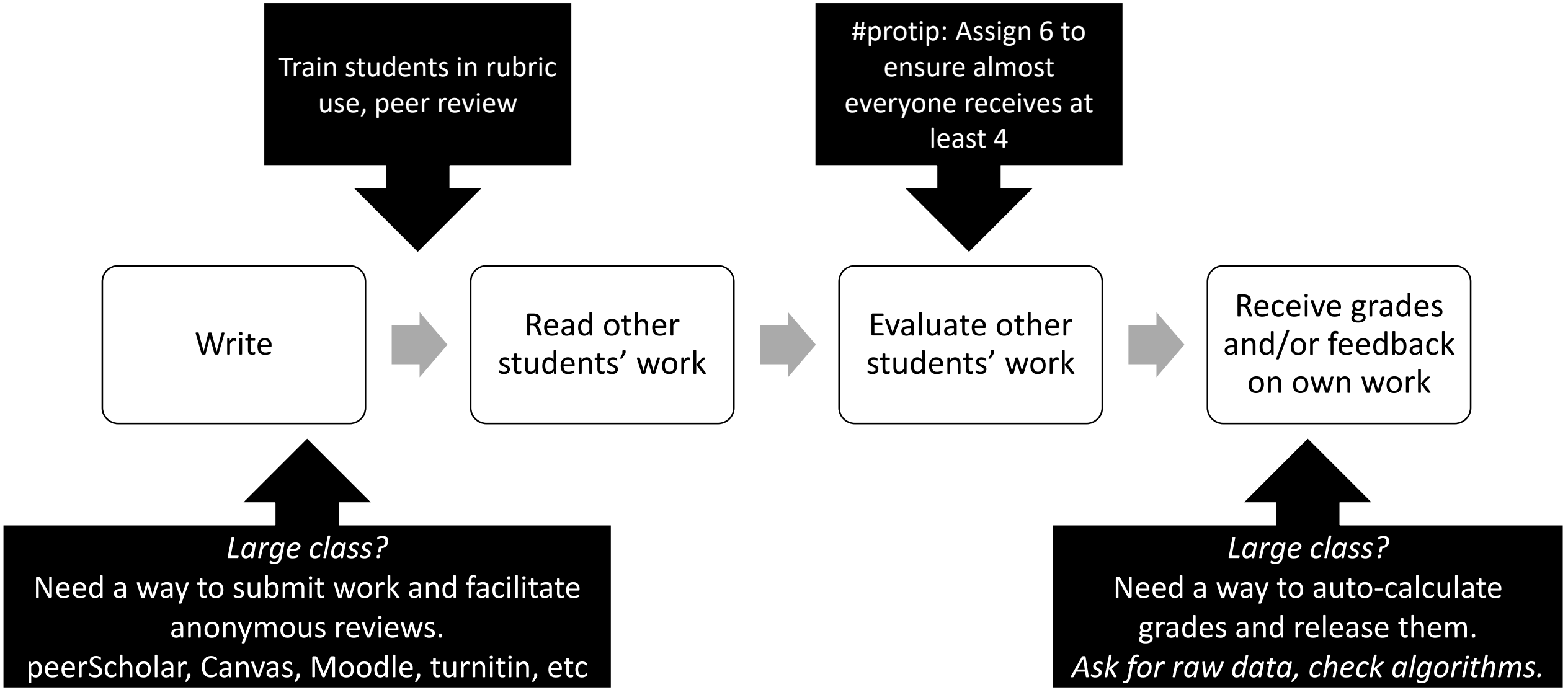
Peer Assessment Overview



Learning Opportunities



Implementation Tips



On Choosing a Platform for Peer Assessment

https://isit.arts.ubc.ca/peer-evaluation-and-review/

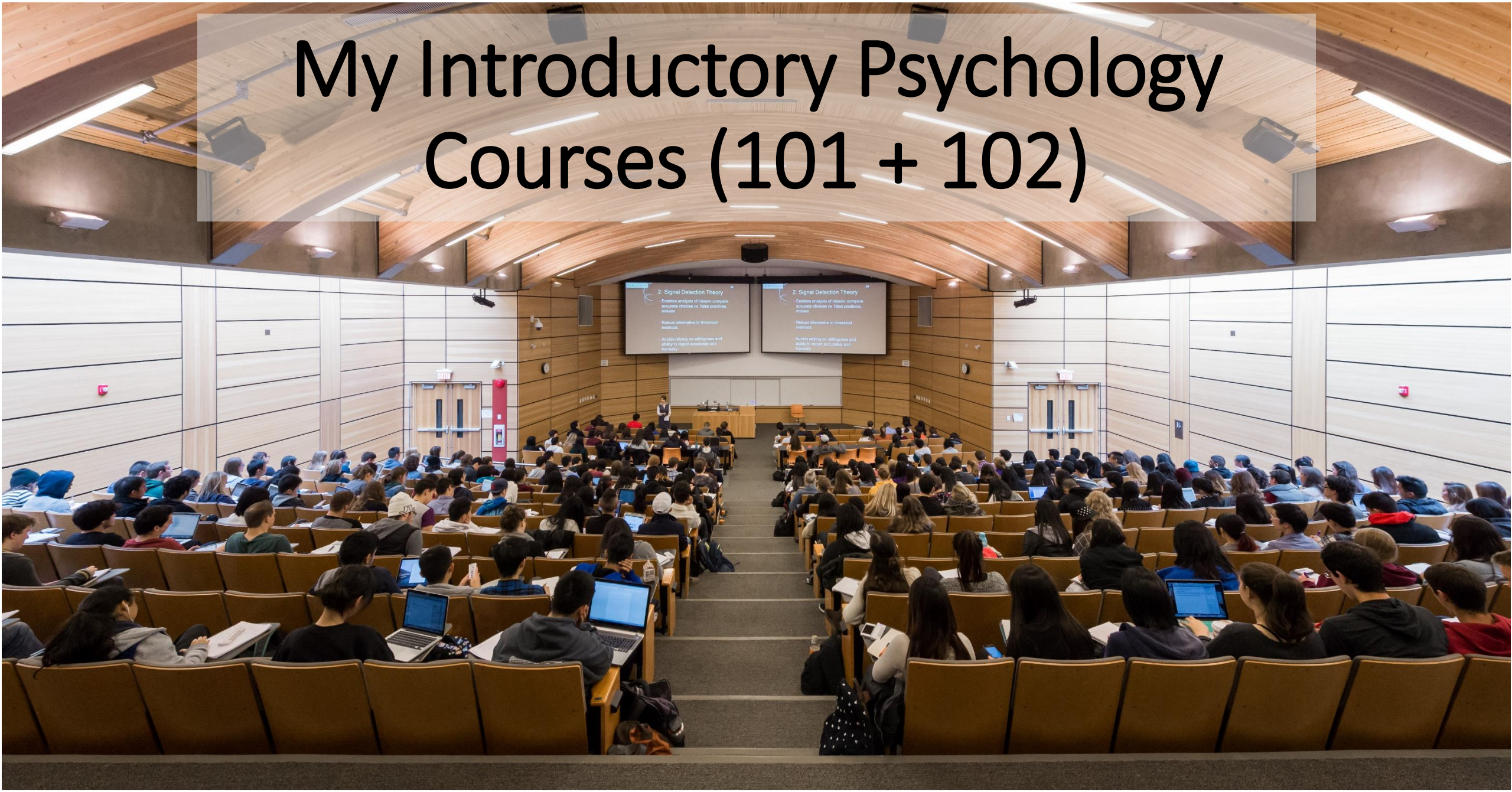
peerScholar v Canvas for Peer Assessment

http://blogs.ubc.ca/catherinerawn/2018/06/01/peerscholar-v-canvas-peer-review/

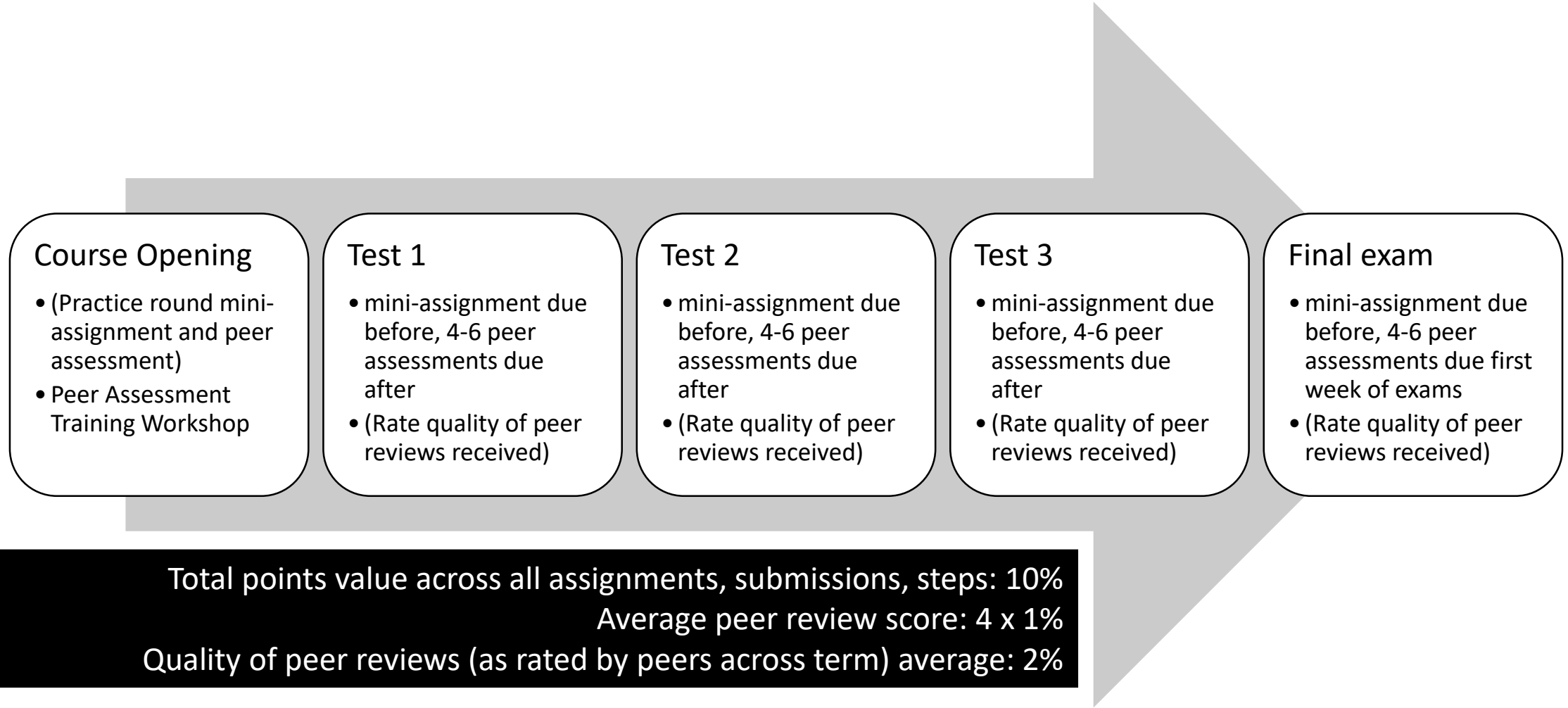
http://blogs.ubc.ca/catherinerawn/files/2018/06/Finding-a-Tool-to-Facilitate-Peer-Review-in-Large-Classes.pdf

#protip: **Ensure you will be able to export the data so you can calculate per student, across reviewers** (i.e., one row per student including a single score from each reviewer).

My Introductory Psychology Courses (101 + 102)



General Overview (my) PSYC 101 & PSYC 102 *Introductory Psychology* course assessments



Student Rubric Overview

Your written work is evaluated based on the following criteria:	
1. Selecting a concept from the appropriate key terms list	5
2. Describing the concept thoroughly and accurately	5
3. Drawing an interesting and useful connection between the concept and an experience or example beyond the course	5
4. Communicating ideas so they are easy to understand	5
	20 points

Sample Criterion from Student Rubric

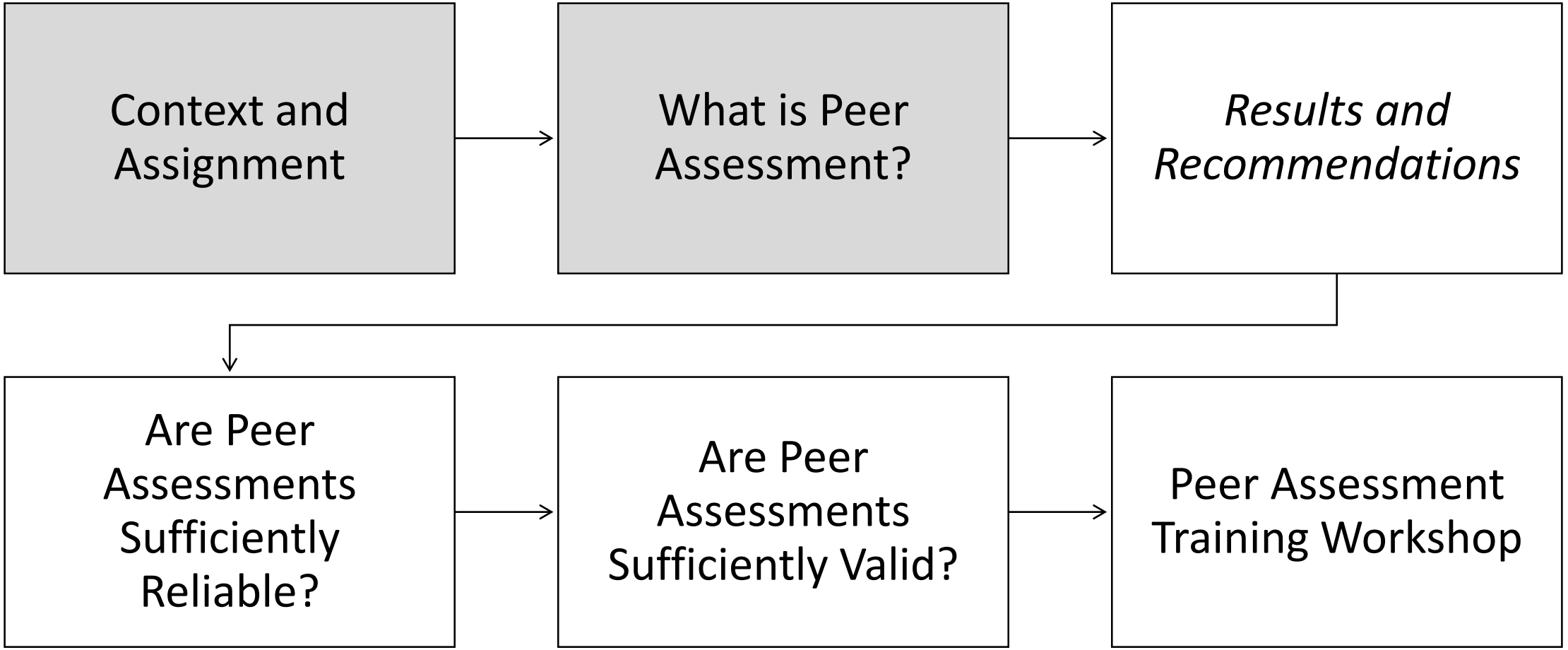
Criterion 2. **Describing the concept.**

What have you learned about this concept? Describe the concept thoroughly, including its nuances, *in your own words*. In your response, you might explore how this concept compares and contrasts with another concept to show its nuances. Or you might compare your current understanding of this concept with what you used to think was true, and how your thinking has changed.

To be evaluated on a 0-5 scale, with these anchors:

5 points	4 points	3 points	2 points	1 point	0 points
• Perfect mastery of the concept. Accurate (check the course materials), complete, detailed, and thorough.	• Accurate, no key aspects about the concept are missing.	• Expected value for most work. Some detail, no major errors.	• Minimal detail, multiple small errors or one major one, key aspects about the concept are missing.	• Minimal description taken word-for-word from the text (if word-for-word and not quoted, provide feedback that it should be quoted or paraphrased).	• Missing.

Today's plan


Context and Assignment
What is Peer Assessment?
Results and Recommendations
Are Peer Assessments Sufficiently Reliable?
Are Peer Assessments Sufficiently Valid?
Peer Assessment Training Workshop

Sample Information

Year 1 2015/2016 (Exploratory Sample N = 647)		
	PSYC 101 (Term 1)	PSYC 102 (Term 2)
Total n	**366**	**281**
Reported Gender		
Females	?	193
Males	?	88
Total n gender data	*0*	*281*
Year in Program		
1	226	188
2	76	59
3	46	23
4	17	11
5	1	0
Total n year data	*366*	*281*

Year 2 2017/2018 (Confirmatory Sample N = 633)		
	PSYC 101 (Term 1)	PSYC 102 (Term 2)
Total n	**330**	**303**
Reported Gender		
Females	213	184
Males	115	119
Total n gender data	*330*	*303*
Year in Program		
1	167	202
2	117	70
3	35	18
4	8	10
5	1	1
Total n year data	*328*	*301*

8 students received course grades < 10% and were excluded from analyses due to failure to complete a substantial amount of the course.
Across the whole remaining sample (N=1280), 98 students have taken both 101 and 102 with me. They appear twice in the dataset.

Preregistered analysis strategy: **2015/16 exploratory,** 2017/18 confirmatory

Reliability

- Per cohort
  - Up to 4 mini-assignments
  - Up to 3-6 peer assessment scores *per assignment per student*
  - ~640x4x5 = 12000+ observations
- Strategy
  - SD, alpha across reviews
  - Compare calculation methods

(Predictive) Validity

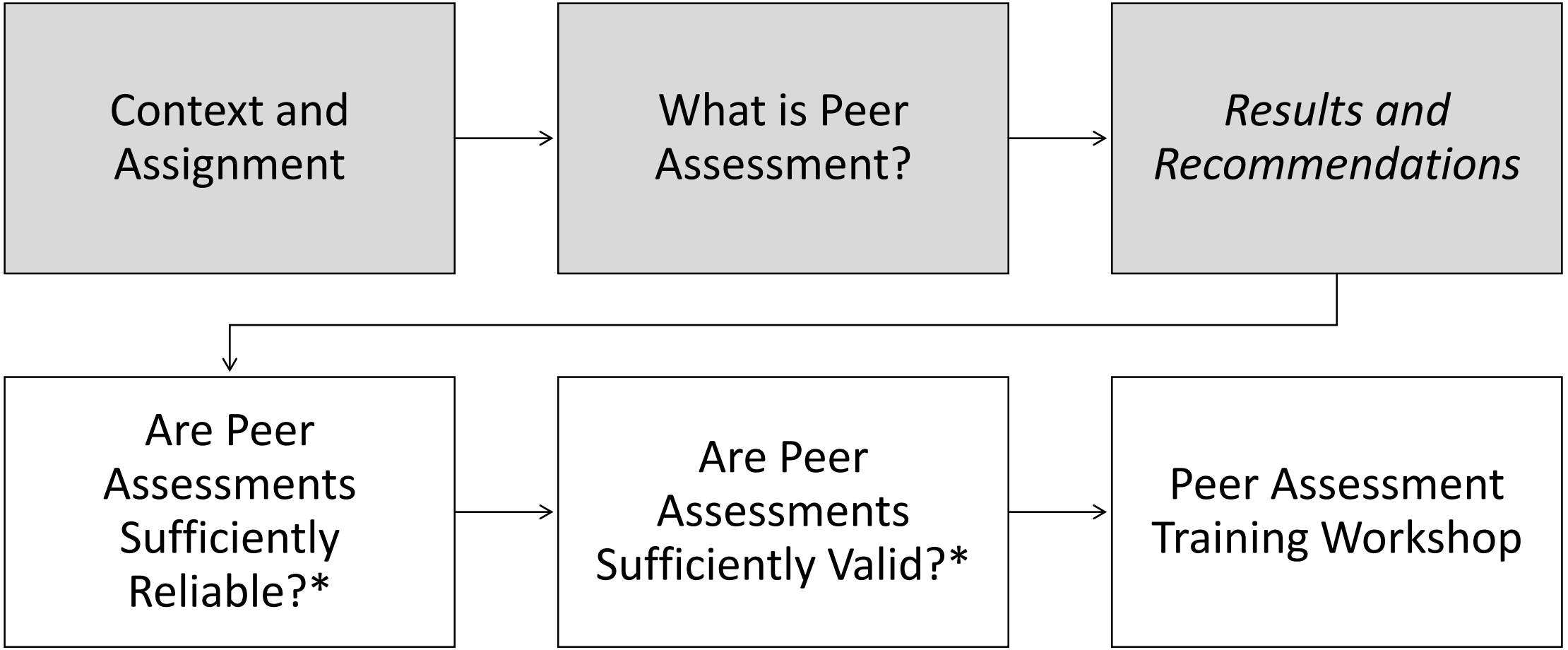
- Correlate, compare w final exam:
  - Multiple choice
  - Fill-in-the-blanks
  - Written (reproduce 2 of your best mini-assignments), graded by TA, similar rubric
- *Are peers picking up on something more than just participation?*

https://aspredicted.org/

7 Overall Results and Recommendations
(based on 2015/16 data)

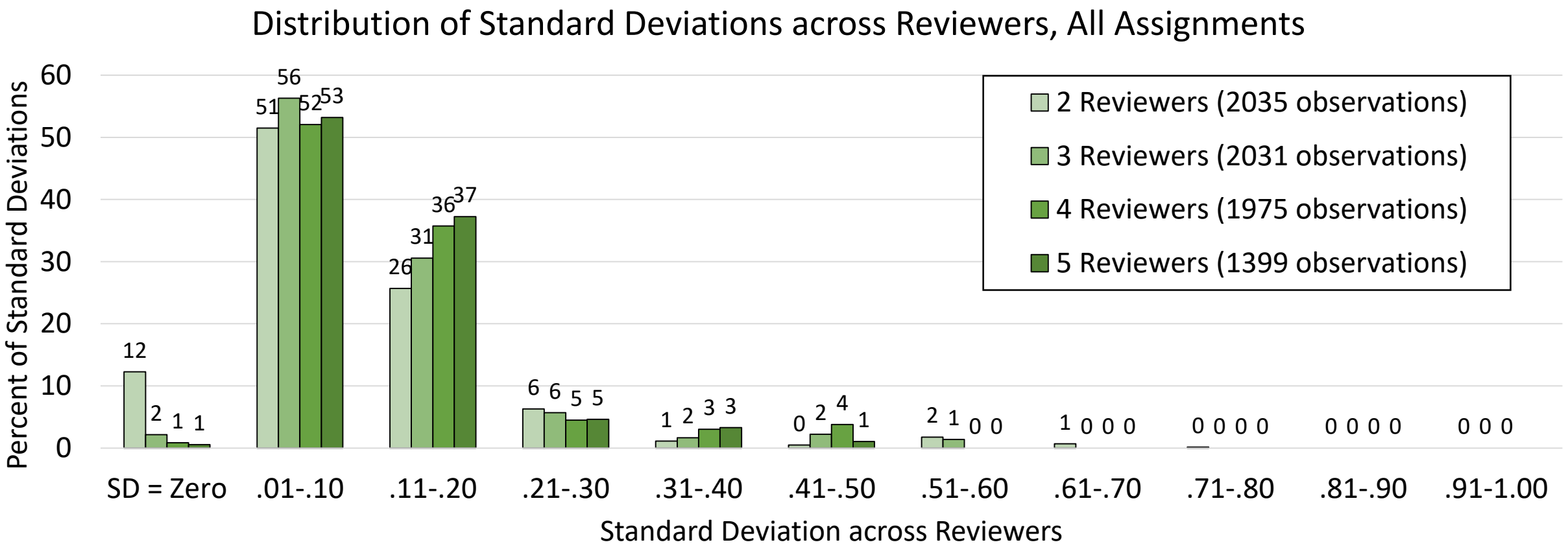
1. Before choosing a program, **ensure you can export data**, 1 row per *student*.
2. **Assign students to be peer reviewers on 6 others' work**, so most receive at least 4 reviews (most should receive 5-6).
3. Do not use a simple arithmetic average of peer reviews. Instead, **drop highest and lowest scores and take average of the middle scores**.
4. **At least 4 reviews, preferably 5, in aggregate, are acceptably reliable** to use for student grades. Any work receiving 3 or fewer peer reviews should be carefully considered.
5. **Aggregate peer reviewed assignment grades predict final exam scores,** above and beyond the effect of various student participation indicators, suggesting validity.
6. Prepare to explain grading decisions and cite references. **Always offer students a chance to challenge a peer reviewed grade**, regardless of number of reviews received or final score.
7. **Before generalizing from these results, consider own context**, including the quantity, length, and value of the assignment(s) for which you are assigning peer review. These analyses are based on 4 x 200-400 word mini papers worth 1% each.

Today's plan


Context and Assignment
What is Peer Assessment?
Results and Recommendations
Are Peer Assessments Sufficiently Reliable?*
Are Peer Assessments Sufficiently Valid?*
Peer Assessment Training Workshop

Are Peer Assessments Sufficiently Reliable?

For these assignments /1, 88% of reviews are within 20% of each other.



Are Peer Assessments Sufficiently Reliable?

Number of Reviews	α range across assignments	n range across assignments
5 reviews	.657-.717	221-482
4 reviews	.571-.698	381-576
3 reviews	.495-.685	436-576
2 reviews	.406-.605	437-576

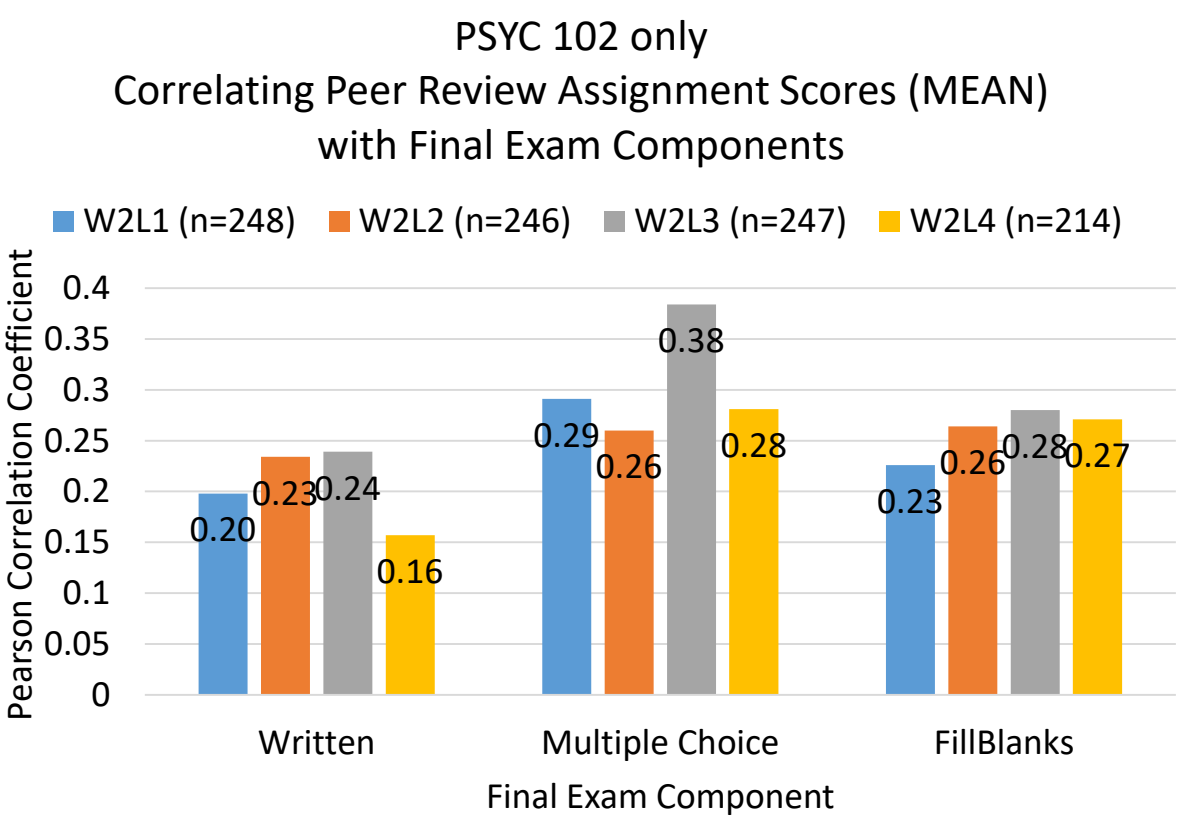
4-5 better than 2-3, especially within course (lower n)

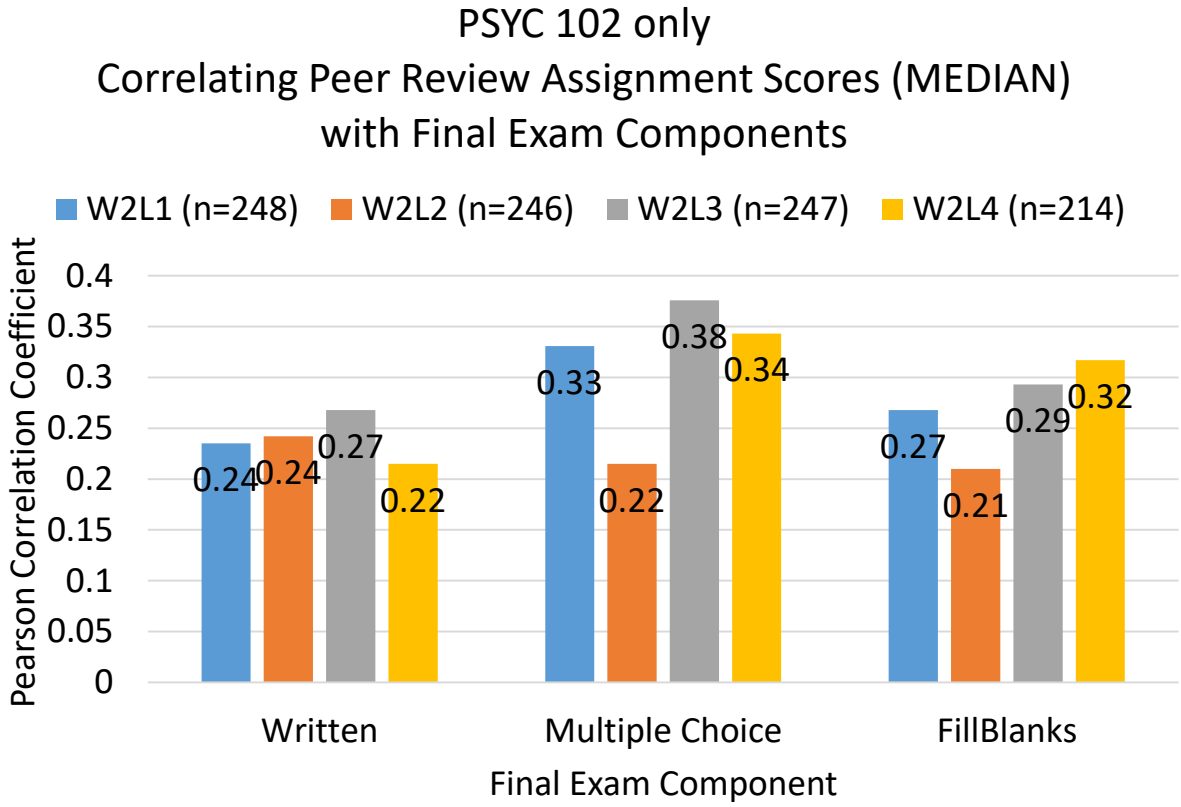
Not super-high, so don't just take arithmetic mean

Are Peer Assessments Sufficiently Valid?

Positively correlate with all final exam components

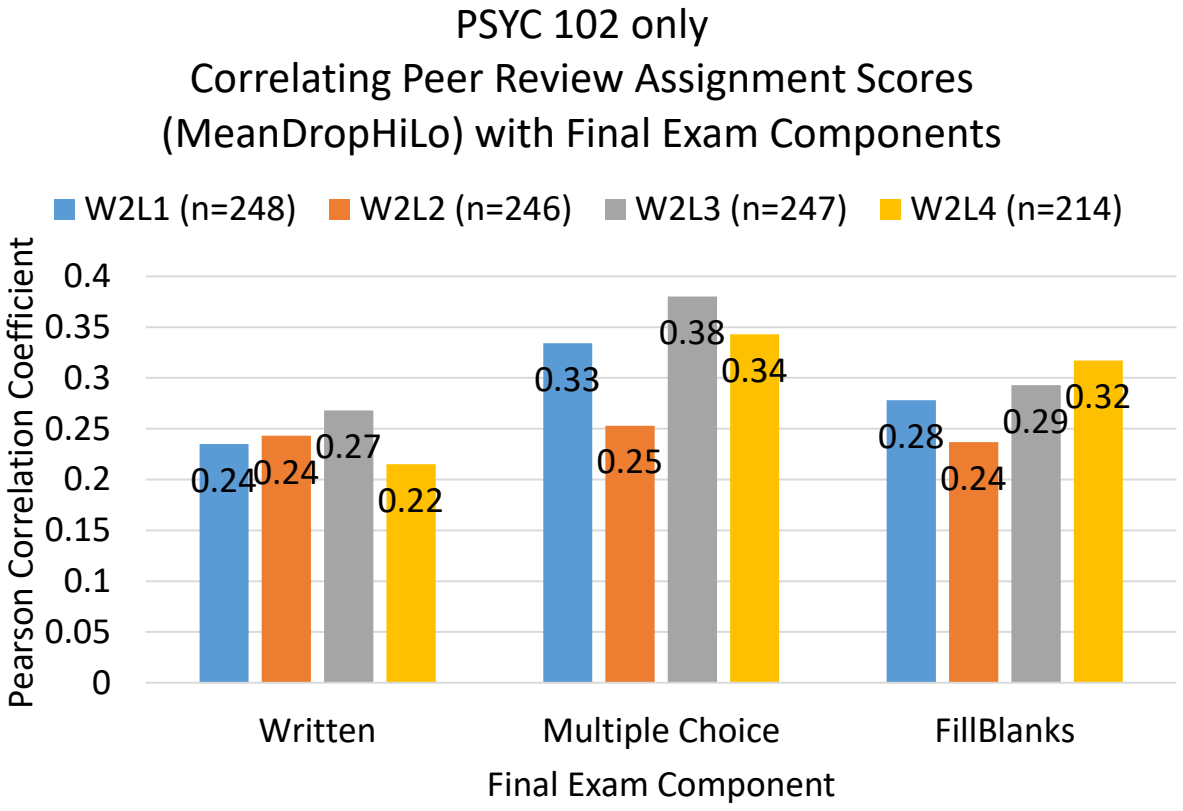
Median correlates a little better than Mean, especially for written component

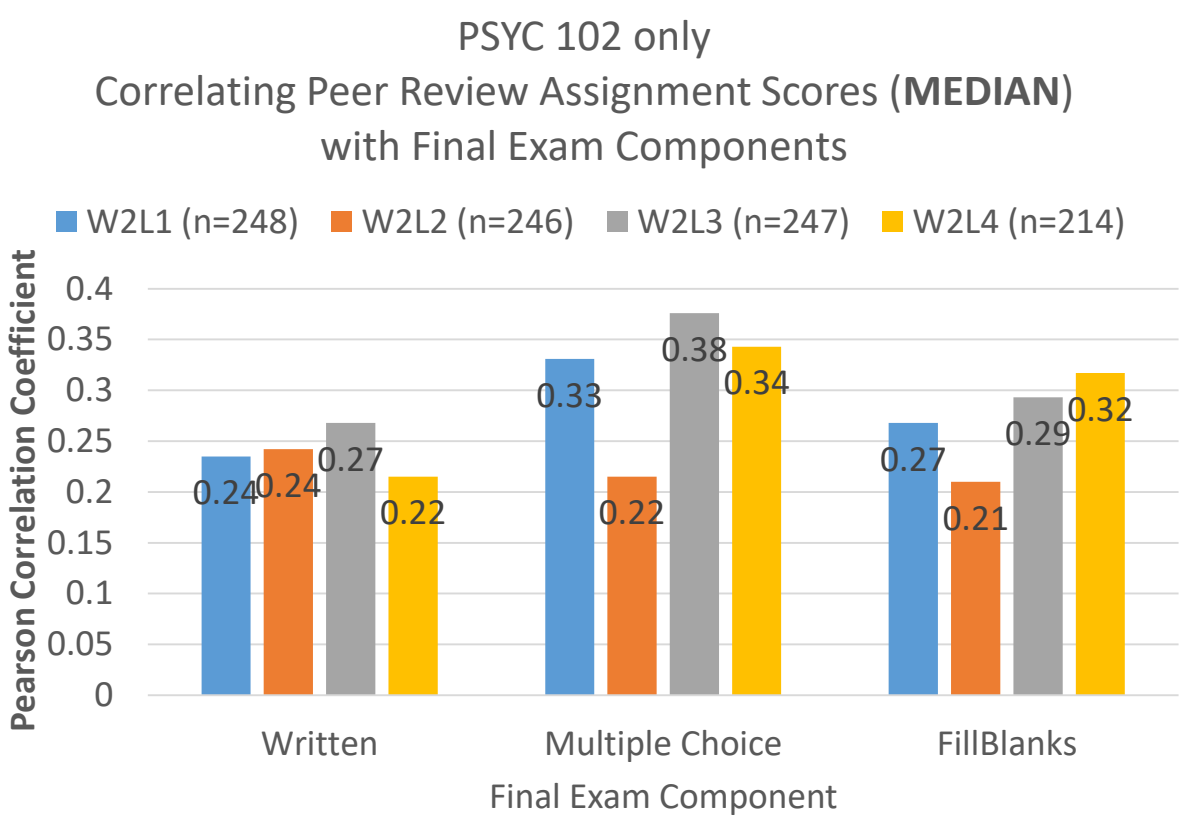




Are Peer Assessments Sufficiently Valid?

MeanDropHiLo predicts about the same as median. Because uses more data when 5-6 peer reviewers, advise this.





Are Peer Assessments Sufficiently Valid?

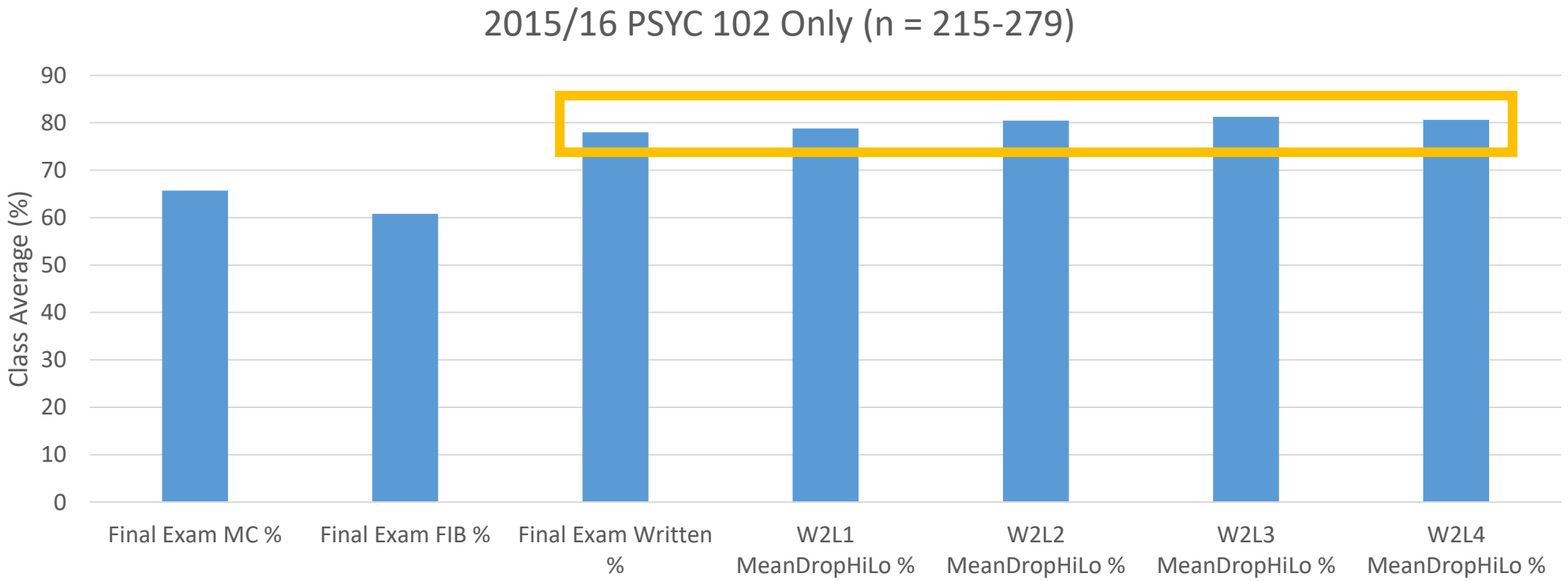
Aggregate peer reviews (MeanDropHiLo) predict final exam performance above and beyond participation, suggesting peers are picking up on learning.

Final Exam Score Component	Written	MC	FIB
*Step 1: 5 x **participation** indicators*	R = .306, **R^2 = .093,** F(5, 185) = 3.82, *p* = .003	R = .422, **R^2 = .178,** F(5, 185) = 8.03, *p* < .001	R = .429, **R^2 = .184,** F(5, 185) = 8.35, *p* < .001
*Step 2: Adding 4 x aggregate MeanDropHiLo **peer review assignment scores***	R = .374, **R^2 = .140,** F(4, 181) = 3.28, *p* = .001 (F change *p* = .047)	R = .533, **R^2 = .284,** F(4, 181) = 8.00, *p* < .001 (F change *p* < .001)	R = .509, **R^2 = .259,** F(4, 181) = 7.02, *p* < .001 (F change *p* = .002)

+ 5-10%

Are Peer Assessments Sufficiently Valid?

Peer Reviewed assignment scores are more similar to written final exam component than they are to MC or FIB



Aggregating across 4-5 peer reviews results in a score that is sufficiently reliable and valid to justify using it for student grades.*

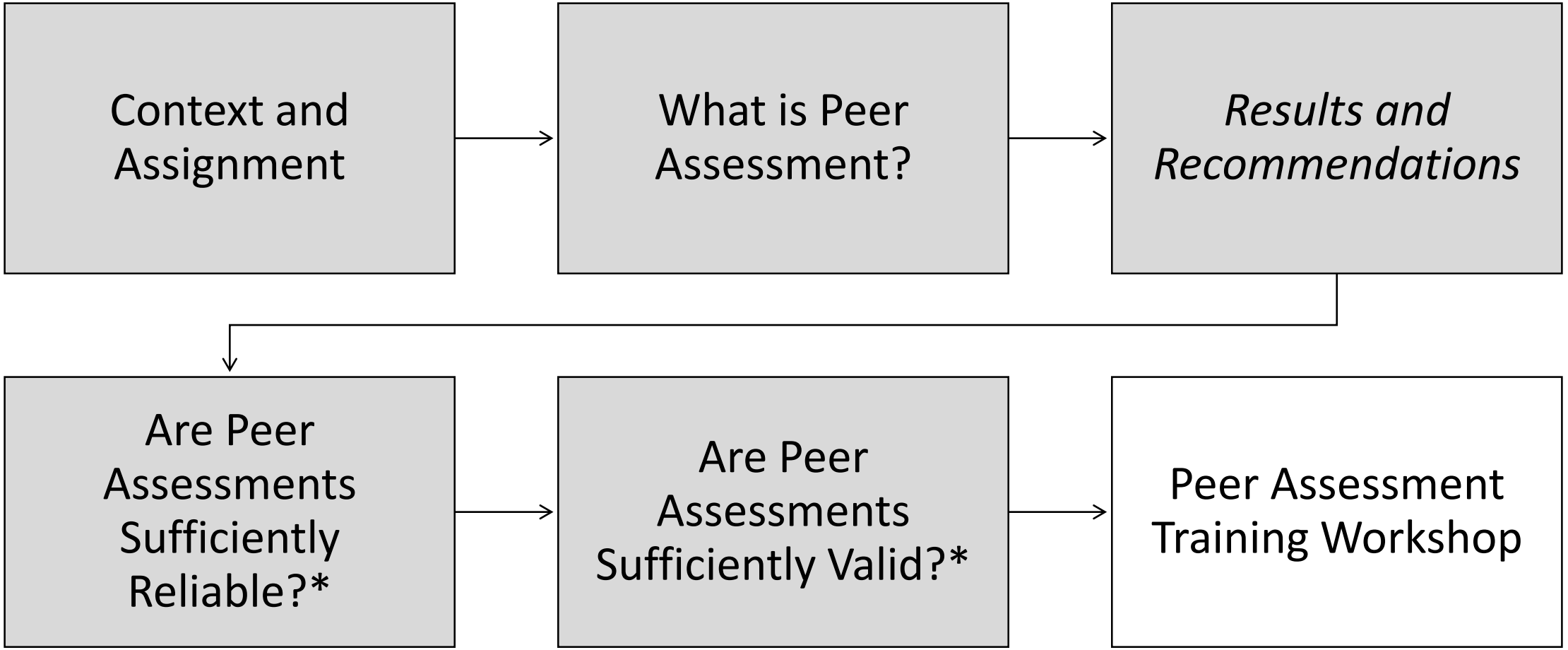
Especially in large classes and low-stakes assignment(s).

Drop highest and lowest score, take average of remainder.

Check any score resulting from 3 or fewer peer reviews.

**Based on 2015/16 exploratory data. Confirmatory analyses in progress.*

Today's plan

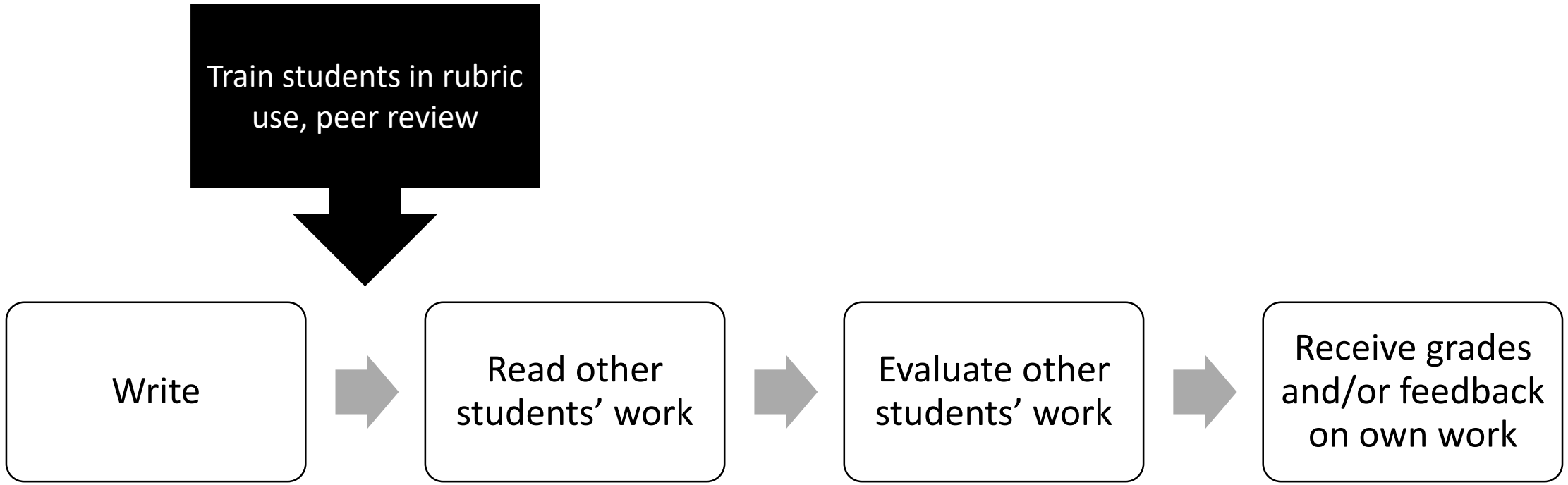

Context and Assignment
What is Peer Assessment?
Results and Recommendations
Are Peer Assessments Sufficiently Reliable?*
Are Peer Assessments Sufficiently Valid?*
Peer Assessment Training Workshop

Additional Challenges

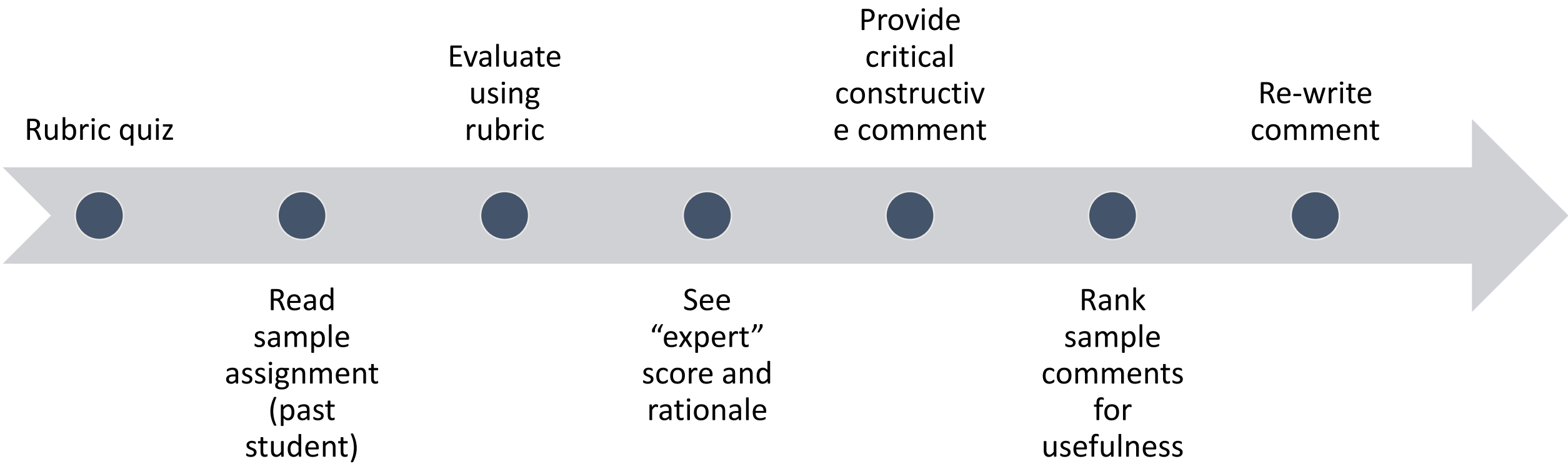
Students don’t trust each other

Comments were poor quality

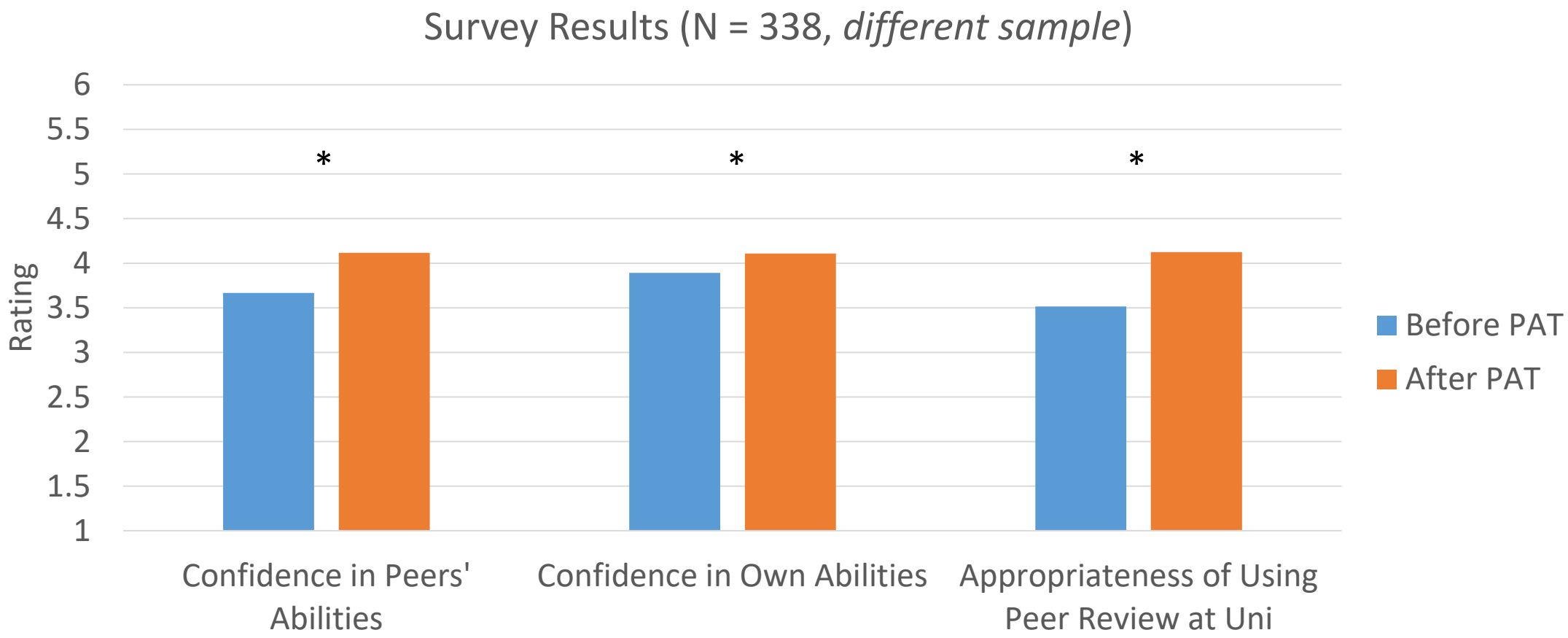
Implementation Tips



Peer Assessment Training Workshop Overview



Completing the Peer Assessment Training Workshop improves students' attitudes



peerassessment.arts.ubc.ca

Setup guides

Other helpful links

Ready to Start?
peerassessment.arts.ubc.ca

Phase 1: Assignment

- Develop Assignment and Rubric
- Choose platform for Peer Reviews (e.g., peerScholar, Moodle, Canvas?), and create assignment there
  - *Get advice!*
- Test the platform *including any automatically-generated scores*

Phase 2: Peer Assessment Training Workshop

- Create brief quiz to test knowledge of assignment & rubric
- Source 2-4 sample assignments, grade them using the rubric and add comments
- Choose Canvas or edX Edge
- Follow instructions in corresponding Workshop Setup Guide
- Option: Assign completion score

Phase 3: Evaluation

- Finalize Peer Review grades
  (Examine auto-generated scores to ensure fairness and accuracy)
- Options: Compare your students' attitudes toward peer assessment before and after doing PAT (PAPAQ)

Is your institution on Canvas? Do you have a Canvas login? Access the template…

Find **Peer Assessment Training Workshop** (Canvas Version) in Canvas Commons, import it into your course

Self-enroll into the **template course to see the workshop from the student's perspective** https://canvas.ubc.ca/enroll/RCW8WR

Selected References and Resources

http://peerscholar.com/research

https://peerassessment.arts.ubc.ca/

Ashenafi, M. M. (2017). Peer-assessment in higher education – twenty-first century practices, challenges and the way forward. *Assessment & Evaluation in Higher Education, 42*, 226-251. DOI: 10.1080/02602938.2015.1100711

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education, 24*, 331-350. https://doi.org/10.1080/0307507991233137993 5

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*, 287-322. https://doi.org/10.3102/00346543070003287

Gingerich, K. J., Bugg, J. M., Doe, S. R., Rowland, C. A., Richards, T. L., Tompkins, S. A.…McDaniel, M. A . (2014). Active processing via write-to-learn assignments learning and retention benefits in introductory psychology. *Teaching of Psychology, 41*, 303–308. doi:10.1177/0098628314549701

Nevid, J. S., Amrose, M. A., & Pyun, Y. S. (2017). Effects of higher and lower level Writing-to-Learn Assignments on higher and lower level examination questions. *Teaching of Psychology, 44*, 324-329. DOI 10.1177/0098628317727645

Nevid, J. S., Pastva, A., & McClelland, N. (2012). Writing-to-Learn assignments in introductory psychology: Is there a learning benefit? *Teaching of Psychology, 39*, 272-275.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education, 39, 102-122.* https://doi.org/10.1080/02602938.2013.795518

Paré, D. E., & Joordens, S. (2008). Peering into large lectures: Examining peer and expert mark agreement using peerScholar, an online peer assessment tool. *Journal of Computer Assisted Learning, 24*, 526-540. https://doi.org/10.1111/j.1365-2729.2008.00290.x

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68,* 249-276. https://doi.org/10.3102/00346543068003249