To draw valid conclusions we need reliable data.

- Reliability of data relies on *consistency*, which can be measured as:
 - Consistency over time
 - Consistency between graders

Validity of conclusions depends on:

- Measuring the right things (e.g. indicators)
- Using appropriate approaches to measure
- Agreement with conclusions drawn from other approaches (students, employers, alumni,...)
- Reliability

Let's use a framework for comparing aggregation approaches in Canada:

Factor	Possible options					
Aggregation target	 single value (e.g. Design = 3.6/5) distribution of performance, (e.g. histogram of student performance) qualitative description (textual based analysis of results) 					
Aggregation level	 up to attribute (e.g. Design) up to indicator within each attribute (e.g. "Problem definition") up to task within indicator within attribute (e.g. "Capstone design report") 					
Differentiation factors	 Year of Program (Year 1 to 4) IDA level (Introduce, Developed, Applied) Program option (e.g. biomechanics vs. materials) Summative vs. Formative Assessment type (e.g. final report, exam, lab simulation, portfolio) Student groups (first in family, racialized, Indigenous) 					
Reliability measure 2019 AMEGA EGA	 Correlation between tasks (e.g. correlation between three measures of "problem definition") Correlation between years (e.g. correlation between scores in 2016, 2017, and 2018) Correlation between multiple ways of measuring an indicator 					

2a) Problem Analysis: Indicator (a)									Factor	Approach		
			Introd	uctory Developing	Adva	nced				_		
100% 90% 80%							F.	4: Exc	eeds e	kpectations	Aggregation target	distribution
ad 60%	1	Ш					<u> i</u> '	3: Me	ets e xp	ectations		of performance
50% 40% 30% 20% 10% 0%	BWE336	BWE106 BWE181 BWE237	BWE426 BWE442 BWE310 BWE380 BWE125 BWE125 BWE125 BWE125	BWE182 BWE348 BWE348 BWE348 BWE348 BWE120 BWE120 BWE120 BWE120 BWE120 BWE120 BWE473 BWE447 BWE473 BWE448 BWE443 BWE448 BWE448 BWE448 BWE448 BWE448 BWE448 BWE231 BWE231 BWE231 BWE231 BWE231 BWE231 BWE232 BWE233 BWE33	BWE 210 BWE 429 BWE 402	BWE282 BWE383	BWE241	2: Ma expect 1: Fail 0: No achiev	rginally tations s to me demon vement	eet expectations estrated	Aggregation level	Multiple (learning outcome within indicator)
	1			Assessment Tool						_	Differentiation factors	IDA
Tool	Ind	Level	Assessor	Question or course learning outcome	# of students at LOM				% of students	Tacion 5		
					0	1	2	3	4	over threshold		
BWE336	а	-	1. Instructor	CLO #5 (Awesome assessment method #5)	7	47	88	93	25	45%		
BWE477	а	1	1. Instructor	CLO #4 (Awesome assessment method #5)	18	54	93	44	50	36%		
BWE106	а	1	1. Instructor	CLO #11 (Awesome assessment method #7)	14	35	45	27	71	51%		
BWE181	а	1	1. Instructor	CLO #11 (Awesome assessment method #6)	27	75	8	6	60	38%		
BWE237	а	1	1. Instructor	CLO #11 (Awesome assessment method #1)	25	46	26	28	83	53%		
BWE426	а		1. Instructor	CLO #7 (Awesome assessment method #8)	43	91	49	15	82	35%	Reliability	
BWE442	а		1. Instructor	CLO #6 (Awesome assessment method #1)	37	48	44	27	57	39%	measure	
BWE310	а		1. Instructor	CLO #3 (Awesome assessment method #2)	61	87	81	65	77	38%		
BWE380	а		1. Instructor	CLO #8 (Awesome assessment method #1)	37	63	35	29	61	40%		
BWE224	а	1	1. Instructor	CLO #2 (Awesome assessment method #6)	95	99	72	80	19	27%		
BWE155	а		1. Instructor	CLO #11 (Awesome assessment method #6)	3	41	54	8	44	35%		
BWE241	а		1. Instructor	CLO #6 (Awesome assessment method #3)	35	93	72	1	75	28%		
BWE182	а		1. Instructor	CLO #10 (Awesome assessment method #3)	77	89	6	47	14	26%		EXAMPLE 1



Overall attribute performance by year



Factor	Approach				
Aggregation target	Frequency distribution: of performance				
Aggregation level	Indicator				
Differentiation factors	Year				
Reliability measure	"Assessment strength" rating by instructor				





Median performance change from year 1 to 4

Dimension

Factor	Approach
Aggregation target	Rubric dimension medians
Aggregation level	Indicator
Differentiation factors	Year level
Reliability measure	% agreement (Inter-rater reliability)