To draw **valid** conclusions we need **reliable** data.

Reliability of data relies on *consistency*, which can be measured as:

- Consistency over time
  - i.e. test-retest reliability
- Consistency between graders
  - i.e. inter-rater reliability
- Internal consistency
  - i.e. inter-item reliability

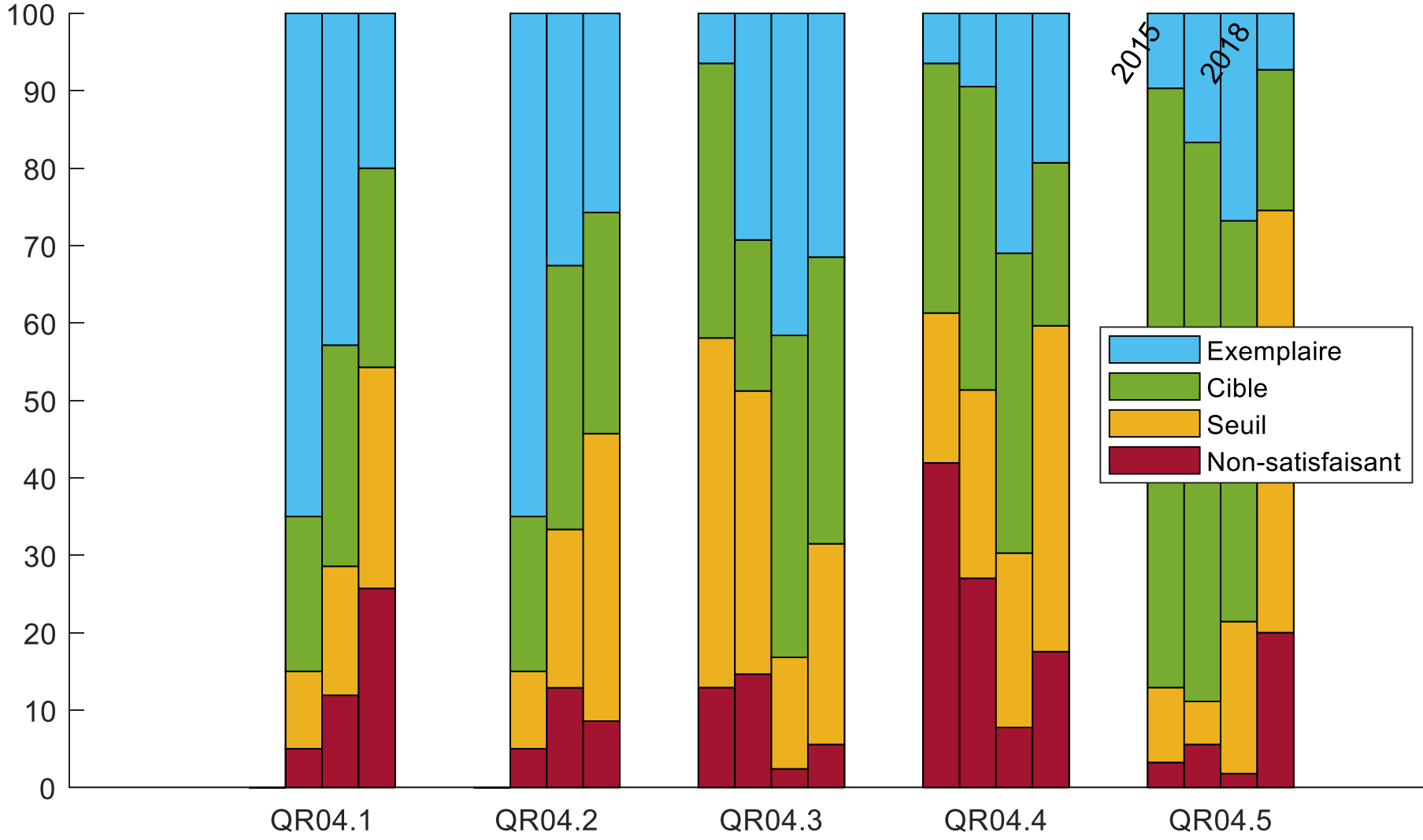
Validity of conclusions depends on:

- 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:

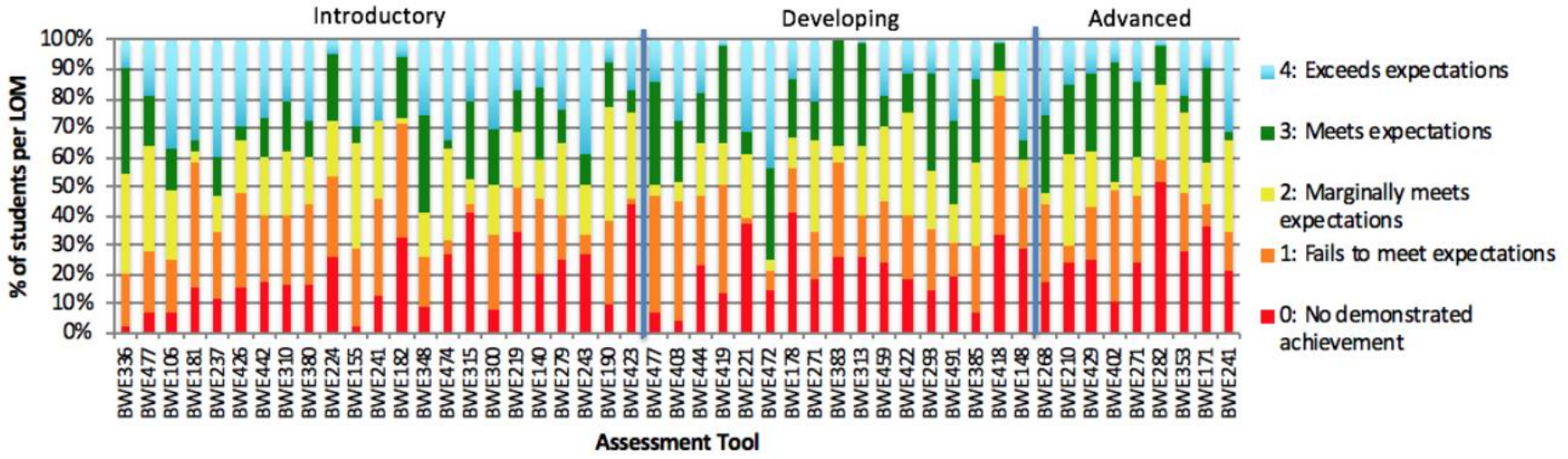
Factor	Possible options
Aggregation target	● single value (e.g. Design = 3.6/5) ● distribution of performance, (e.g. histogram of student performance) ● qualitative description (textual based analysis of results)
Aggregation level	● up to attribute (e.g. Design) ● up to indicator within each attribute (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	● 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

GA4 (QR4) by year



Factor	Approach
Aggregation target	Frequency distribution of performance
Aggregation level	Indicator
Differentiation factors	Year
Reliability measure	

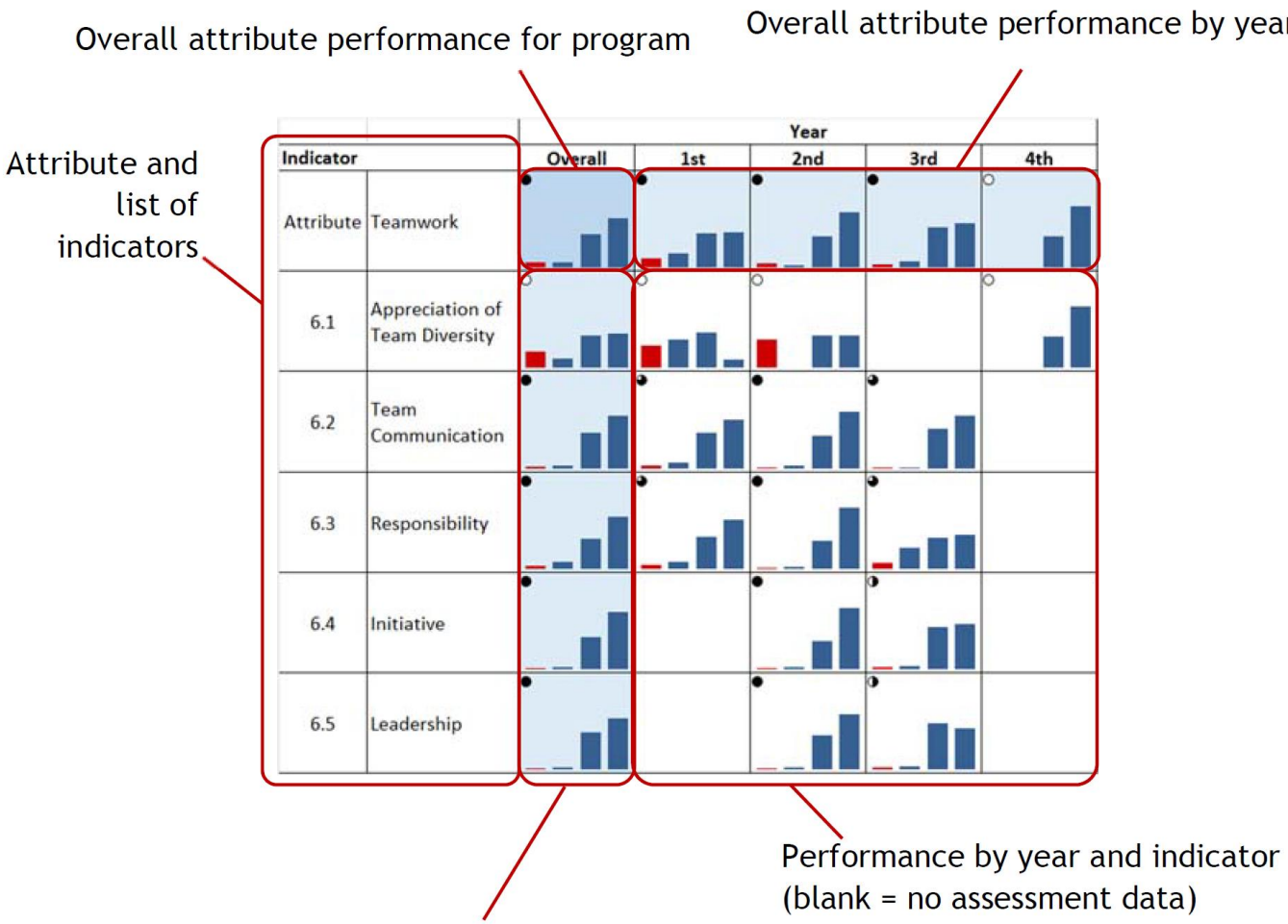
2a) Problem Analysis: Indicator (a)



Tool	Ind	Level	Assessor	Question or course learning outcome	# of students at LOM...					% of students over threshold
					0	1	2	3	4	
BWE336	a	I	1. Instructor	CLO #5 (Awesome assessment method #5)	7	47	88	93	25	45%
BWE477	a	I	1. Instructor	CLO #4 (Awesome assessment method #5)	18	54	93	44	50	36%
BWE106	a	I	1. Instructor	CLO #11 (Awesome assessment method #7)	14	35	45	27	71	51%
BWE181	a	I	1. Instructor	CLO #11 (Awesome assessment method #6)	27	75	8	6	60	38%
BWE237	a	I	1. Instructor	CLO #11 (Awesome assessment method #1)	25	46	26	28	83	53%
BWE426	a	I	1. Instructor	CLO #7 (Awesome assessment method #8)	43	91	49	15	82	35%
BWE442	a	I	1. Instructor	CLO #6 (Awesome assessment method #1)	37	48	44	27	57	39%
BWE310	a	I	1. Instructor	CLO #3 (Awesome assessment method #2)	61	87	81	65	77	38%
BWE380	a	I	1. Instructor	CLO #8 (Awesome assessment method #1)	37	63	35	29	61	40%
BWE224	a	I	1. Instructor	CLO #2 (Awesome assessment method #6)	95	99	72	80	19	27%
BWE155	a	I	1. Instructor	CLO #11 (Awesome assessment method #6)	3	41	54	8	44	35%
BWE241	a	I	1. Instructor	CLO #6 (Awesome assessment method #3)	35	93	72	1	75	28%
BWE182	a	I	1. Instructor	CLO #10 (Awesome assessment method #3)	77	89	6	47	14	26%

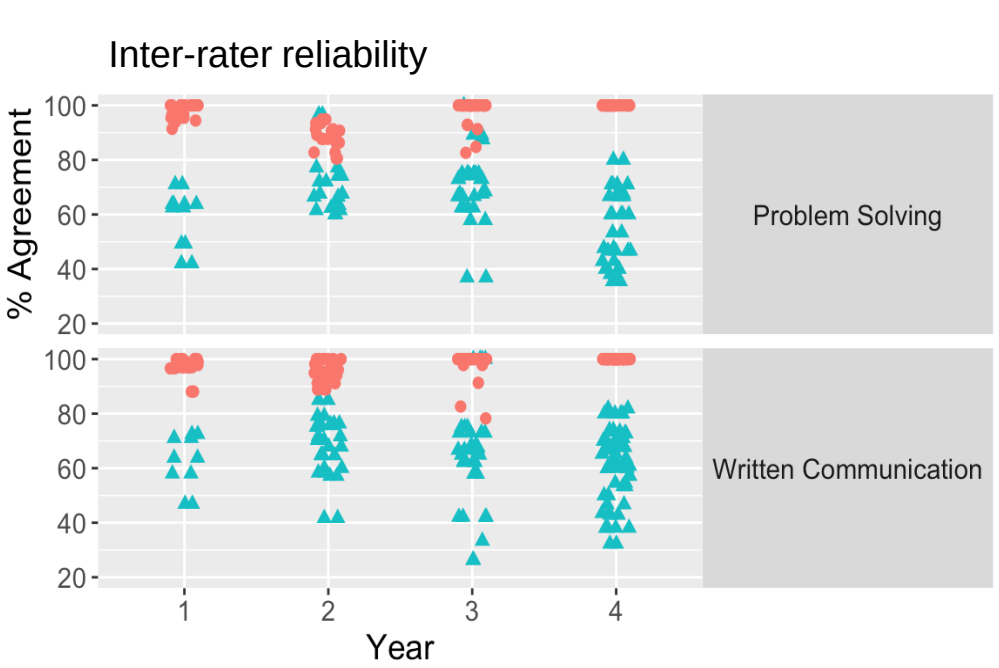
Factor	Approach
Aggregation target	Frequency distribution of performance
Aggregation level	Multiple (learning outcome within indicator)
Differentiation factors	IDA
Reliability measure	

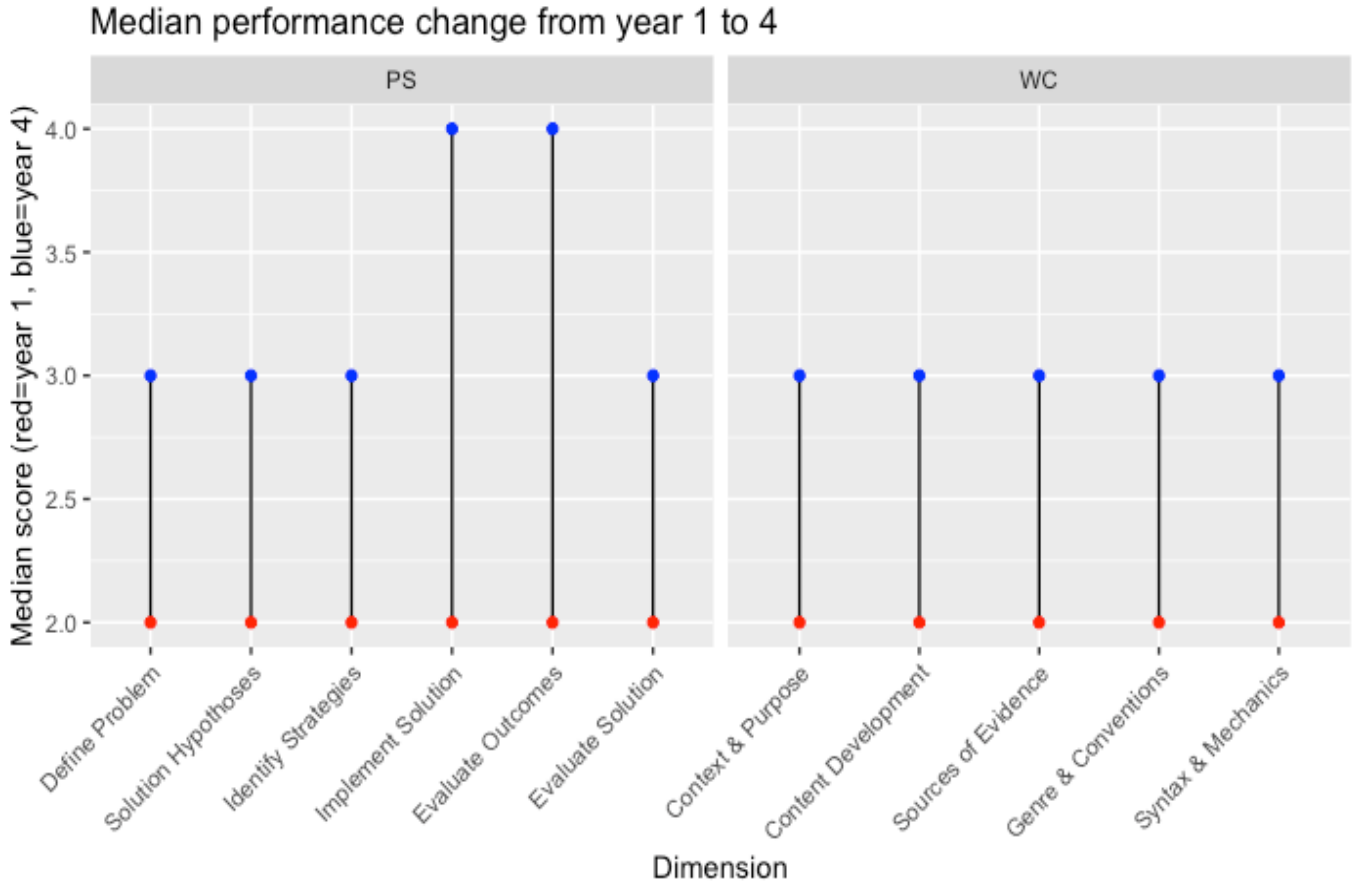
EXAMPLE 1



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

EXAMPLE 2





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