Aggregating quantitative data to draw meaningful conclusions

AMEGA EGAD Workshop May 2019

Resources at: bit.ly/AMEGA-EGAD-2019

Session focus

Comparing approaches to aggregating data, going:

From: Task-level scores assigned to individual students (which most of us are doing)

To: Demonstrating that a student cohort possesses a graduate attribute

Administration

Slides will be available afterwards egad.engineering.queensu.ca
We will circulate key elements arising from the discussion.

For tasks in this workshop,

Instructions for online participants will be in GREEN ITALICS

Instructions for face-to-face participants will be in ORANGE.
We will ask people in the face-to-face workshop to form small groups shortly.

3.1.5 Assessment results: At least one set of assessment results must be obtained for all twelve attributes over a cycle of six years of less. The results should provide clear evidence that graduates of a program possess the above list of attributes.

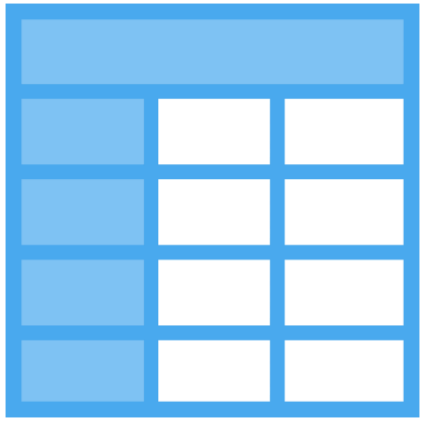
3.2.1 Improvement process: There must be processes in place that demonstrate that program outcomes are being assessed in the context of the graduate attributes, and that the results are validated, analyzed and applied to the further development of the program.

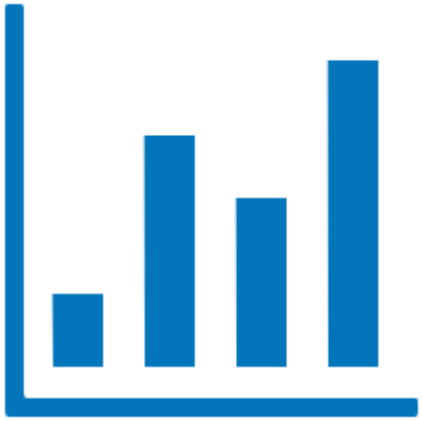
When you **aggregate data**, you replace **groups of observations** with **summary statistics** based on those observations.

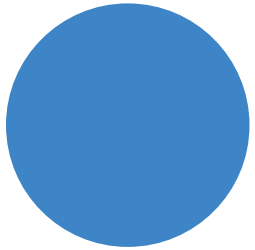
Aggregation level

Low

High



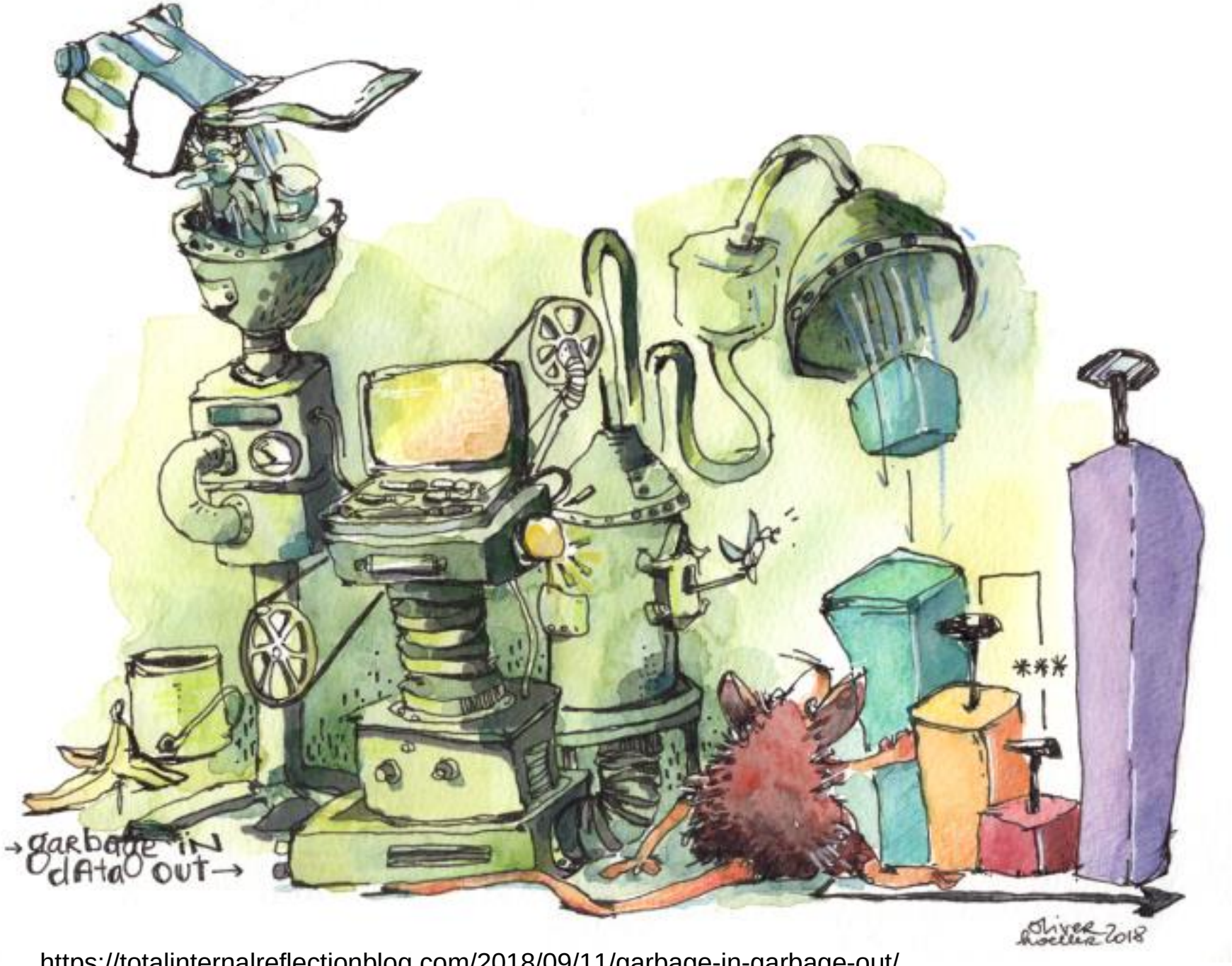




High

Low

Amount of information conveyed



“It is not possible to carry out meaningful statistical analysis of data that is fundamentally inaccurate.”

https://totalinternalreflectionblog.com/2018/09/11/garbage-in-garbage-out/

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

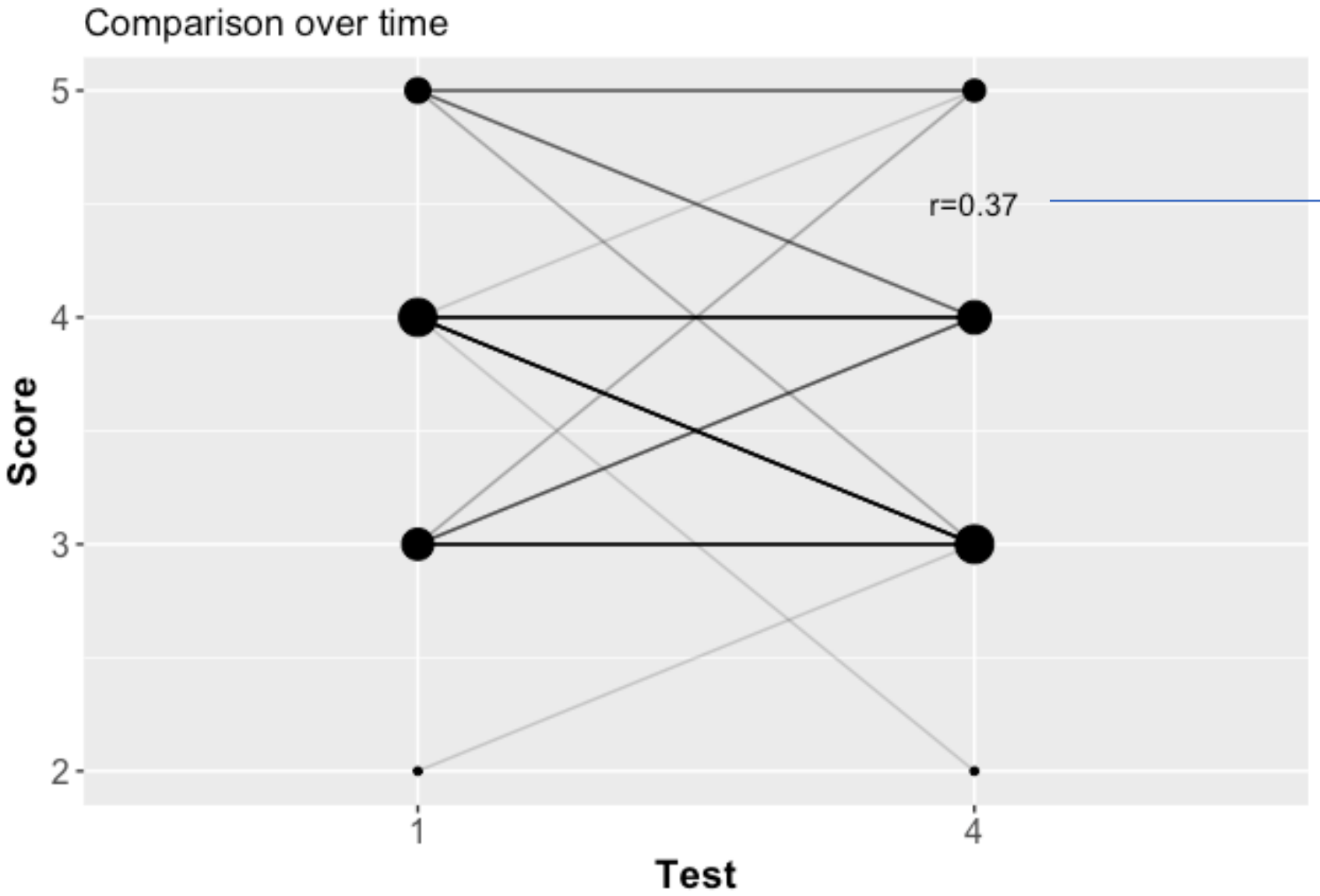
Validity of conclusions depends on:

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

OTHER WORKSHOPS

- Reliability

Characterizing reliability could involve comparing over time (e.g. multiple tests):



Correlation coefficient suggests poor consistency

Characterizing reliability could involve comparing grader agreement:

Task 1: As a group, identify what you are doing, or could do, in your GA process to make your data consistent, considering two possible risks:

1. Consistency over time
- E.g. correlation of scores remeasured over multiple years of delivering a course
- E.g. within different tasks within the same year
2. Consistency between graders (inter-rater reliability)

ONLINE: We will turn microphones on and discuss, using the Google Doc in the shared Google Drive folder (below) to collect our ideas.
FACE-TO-FACE: Form a small group and respond.

bit.ly/AMEGA-EGAD-2019

Task 1 follow-up: What are you are doing, or thinking of doing, in your GA process to make your data consistent?

Considerations

1. Consistency over time
- E.g. correlation of scores remeasured over multiple years of delivering a course
- E.g. within different tasks within the same year
2. Consistency between graders (inter-rater reliability)

bit.ly/AMEGA-EGAD-2019

`group_by(Student ID, Course, Attribute, Indicator)`

Student ID	Program	Year of Study	Course	Attribute	Indicator	Assessment	Score
A	ENGR	1	ENGR 101	KB	ENGR-KB-1	Midterm	3
A	ENGR	1	ENGR 101	KB	ENGR-KB-1	Final	5
B	ENGR	1	ENGR 101	KB	ENGR-KB-1	Midterm	4
B	ENGR	1	ENGR 101	KB	ENGR-KB-1	Final	4
C	ENGR	1	ENGR 101	KB	ENGR-KB-1	Midterm	4
C	ENGR	1	ENGR 101	KB	ENGR-KB-1	Final	1

Raw Data

`summarize(Score = mean(Score)`

Student ID	Course	Attribute	Indicator	Score
A	ENGR 101	KB	ENGR-KB-1	4
B	ENGR 101	KB	ENGR-KB-1	4
C	ENGR 101	KB	ENGR-KB-1	2.5

Distribution

`group_by(Course, Attribute, Indicator)`

Student ID	Course	Attribute	Indicator	Score
A	ENGR 101	KB	ENGR-KB-1	4
B	ENGR 101	KB	ENGR-KB-1	4
C	ENGR 101	KB	ENGR-KB-1	2.5

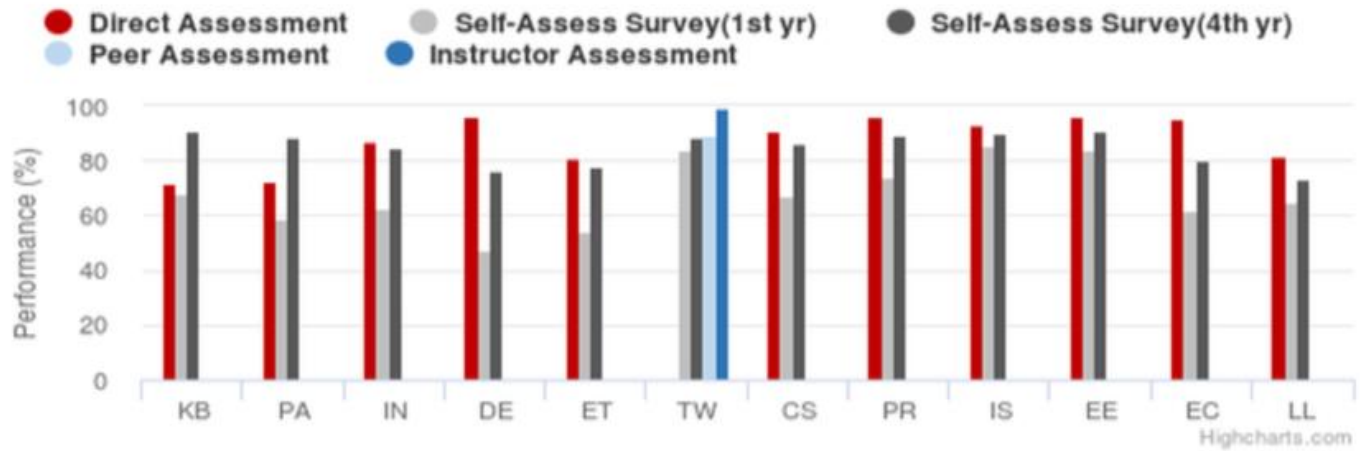
`summarize(Score = mean(Score)`

Course	Attribute	Indicator	Score
ENGR 101	KB	ENGR-KB-1	3.5

Aggregate

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

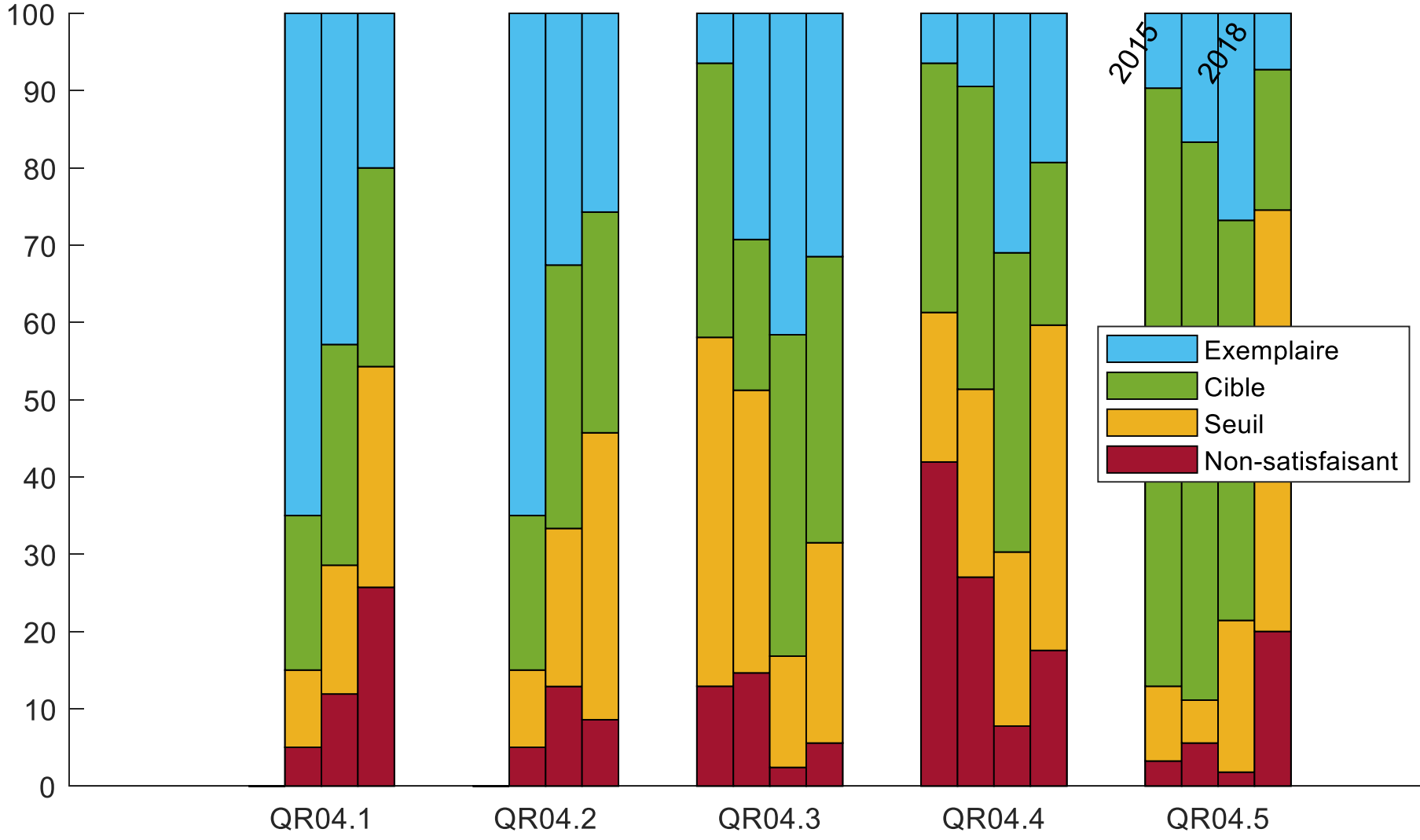


	KB		PA		IN		DE		ET		TW		CS		PR		IS		EE		EC		LL	
#	Val	Targ	Val	Targ	Val	Targ	Val	Targ	Val	Targ	Val	Targ	Val	Targ	Val	Targ	Val	Targ	Val	Targ	Val	Targ	Val	Targ
1	87.3	80	82.6	80	86.7	80	100.0	80	88.2	80	97.2	-	91.6	80	95.7	80	89.9	80	95.0	80	92.8	80	73.7	-
2	75.9	-	67.0	-	90.3	80	87.2	80	74.8	80	90.4	80	98.5	80	96.7	80	90.8	80	84.7	-	57.5	-	76.1	80
3	93.9	80	93.5	80	82.1	80	97.4	80	68.0	80	70.8	-	82.8	80	95.5	80	89.2	80	93.9	80	76.5	80	70.0	-
4	82.9	80	66.8	80	93.1	80	97.4	80	75.9	80	78.8	80	84.7	80	74.2	-	100.0	80	93.7	80	94.4	80	75.5	80
5	80.5	80	52.6	-	82.1	80	100.0	80	97.1	80	88.7	80	87.4	80	88.0	80	86.6	80		80	100.0	80	84.4	80
6	73.7	-	85.1	80	62.5	-	96.9	80	54.3	-	99.1	80	100.0	80			85.5	-	99.4	80	66.8	-	43.3	-
7	87.5	80	64.8	80	84.7	80	100.0	80	79.2	80	94.3	80	66.8	-			90.4	80	83.1	-	83.5	80	76.4	80
8	80.7	80	60.2	-			88.8	80			99.1	80	86.4	80					88.2	80			78.0	80
9	73.9	-	89.6	80			100.0	80			82.0	80											89.3	80
10	86.4	80	72.0	80			47.6	-			99.1	80											73.9	-
11	48.0	-	54.1	-			76.4	80			91.6	80											88.1	80
12	86.7	80	87.0	80							93.5	-												
13	84.4	80									96.8	80												
14	58.3	80									75.5	-												
15	74.0	80									82.5	80												
16	55.8	80																						
17	80.7	80																						
18	58.3	80																						

Target met or exceeded
Performance below target
Missing data
No target

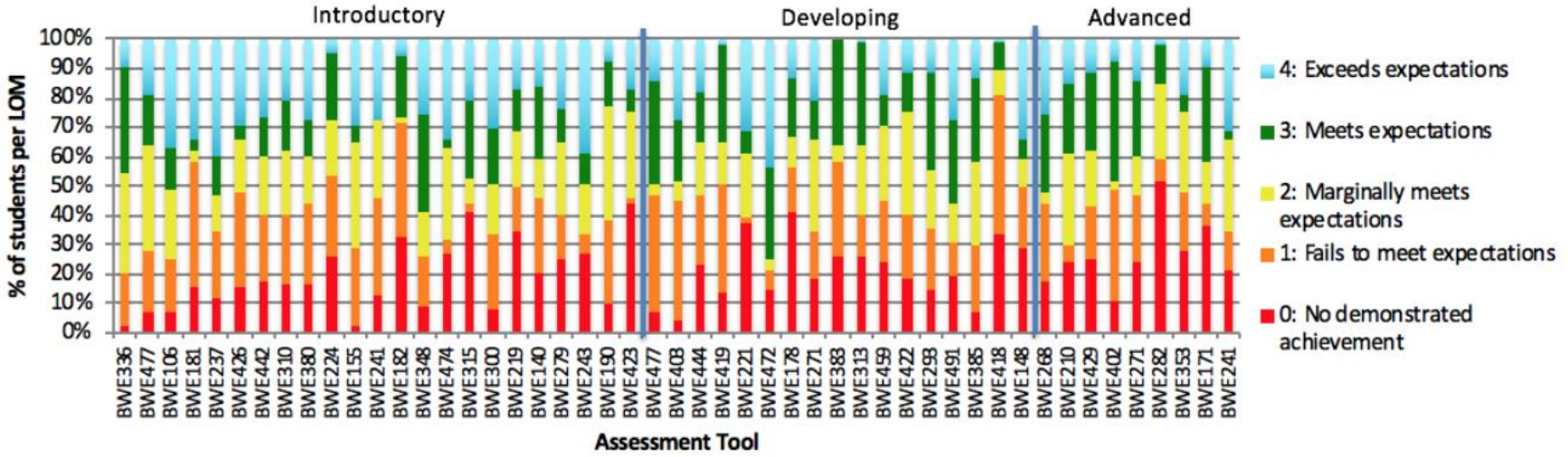
Factor	Approach
Aggregation target	Single values: Mean and % meeting target
Aggregation level	Attribute, but only for CEAB reporting, not internal use
Differentiation factors	Direct/peer/self assessment
Reliability measure	Factor analysis

GA4 (QR4) by year



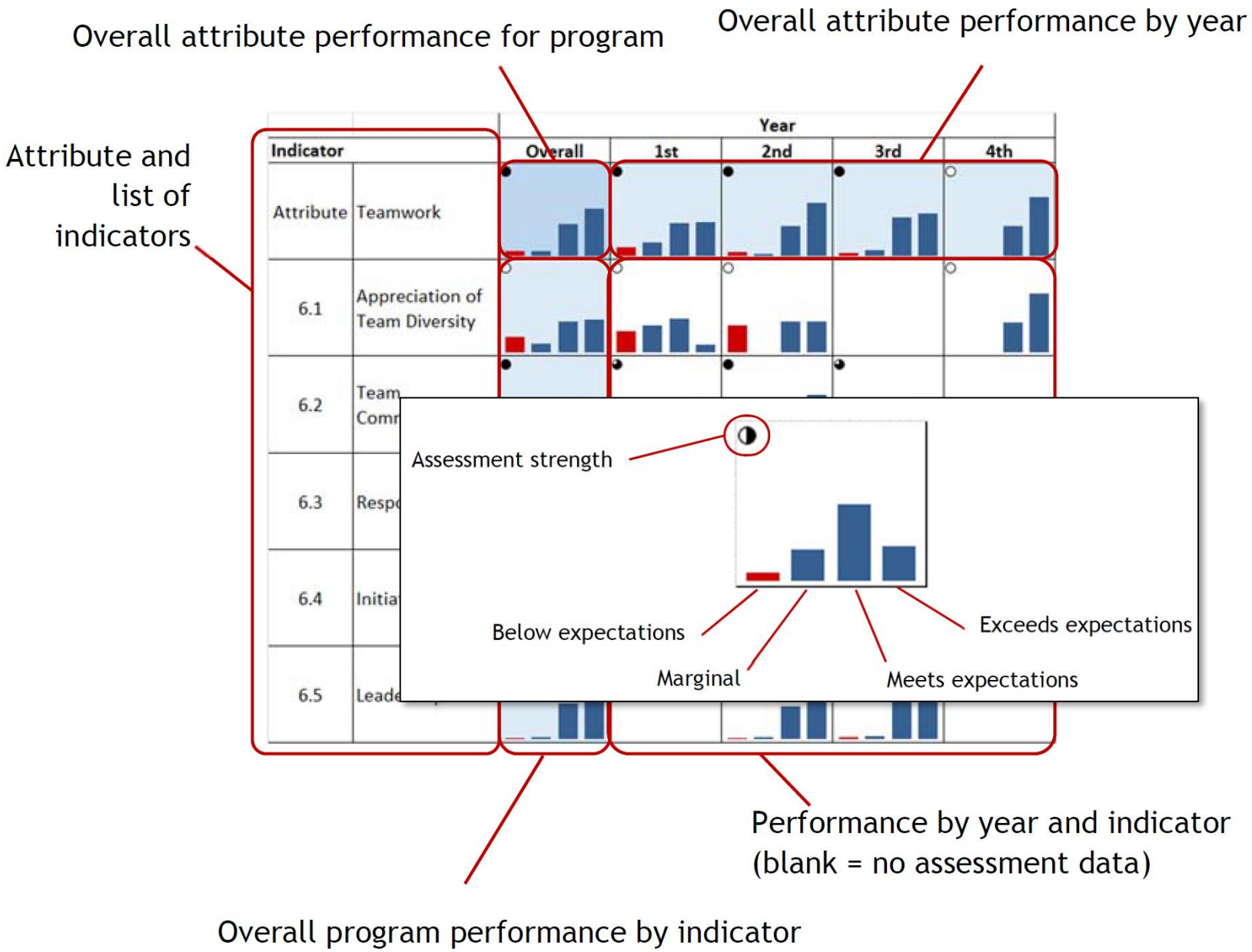
Factor	Approach
Aggregation target	Frequency distribution of performance
Aggregation level	Indicator
Differentiation factors	Year
Reliability measure	

2a) Problem Analysis: Indicator (a)

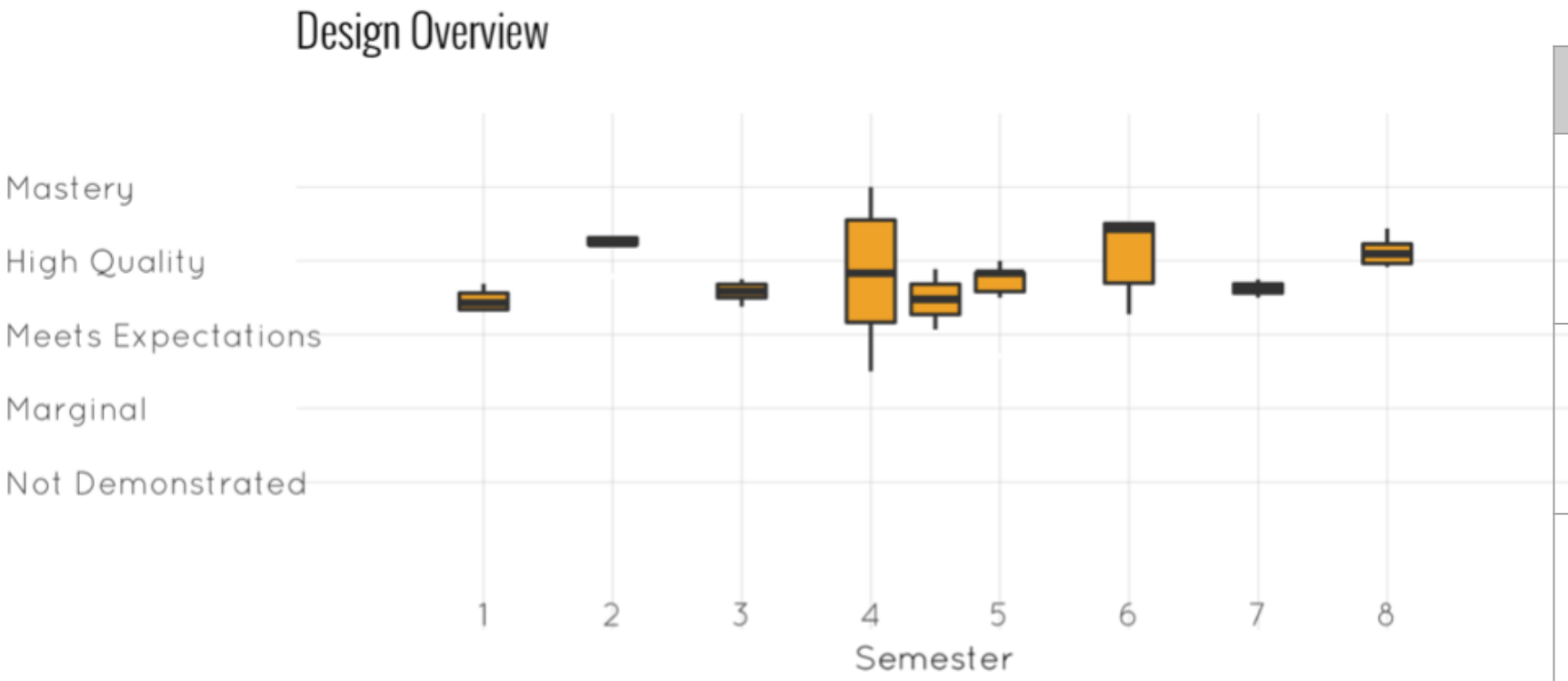


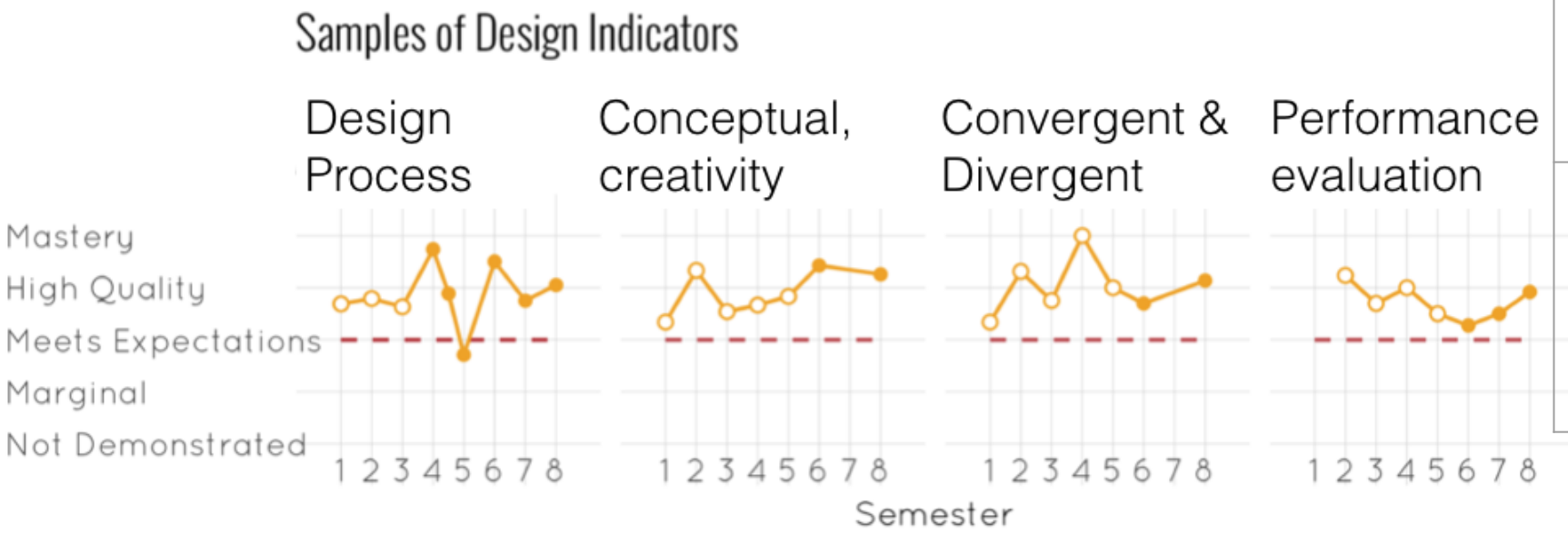
Tool	Ind	Level	Assessor	Question or course learning outcome	# of students at LOM...					% of students over threshold
					0	1	2	3	4	
BWE336	a	I	1. Instructor	CLO #5 (Awesome assessment method #5)	7	47	88	93	25	45%
BWE477	a	I	1. Instructor	CLO #4 (Awesome assessment method #5)	18	54	93	44	50	36%
BWE106	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	

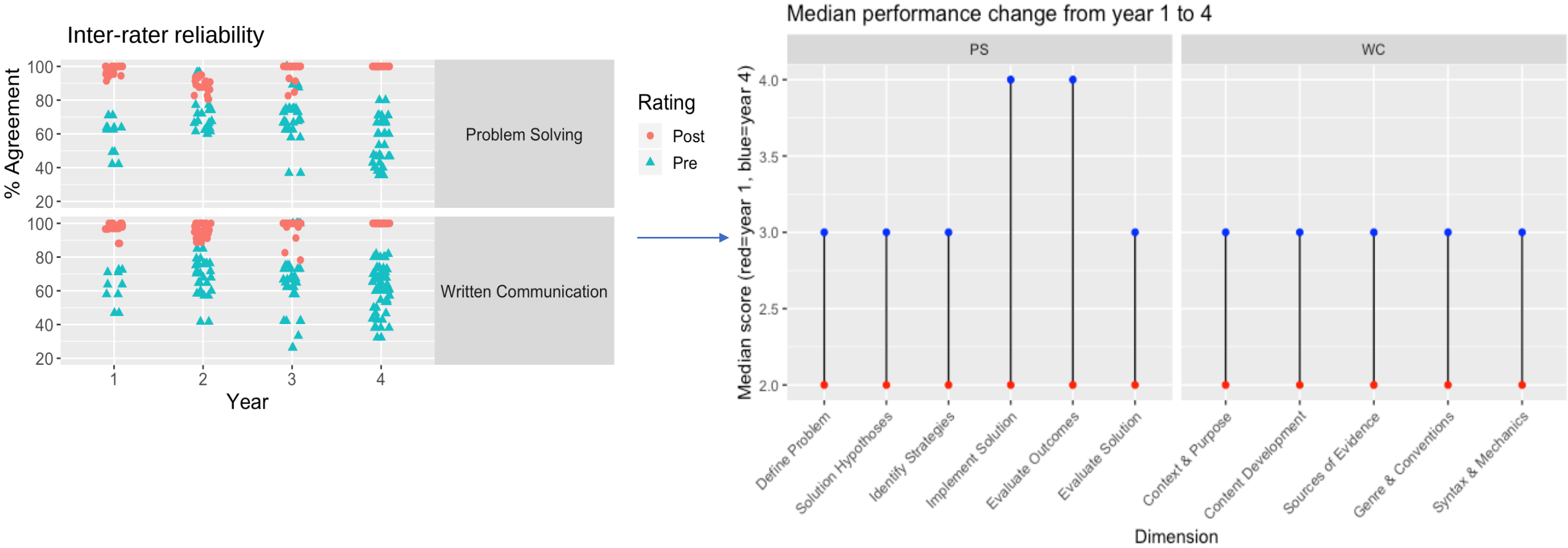


Factor	Approach
Aggregation target	Frequency distribution of performance
Aggregation level	Indicator
Differentiation factors	Year
Reliability measure	Assessment strength rating by instructor

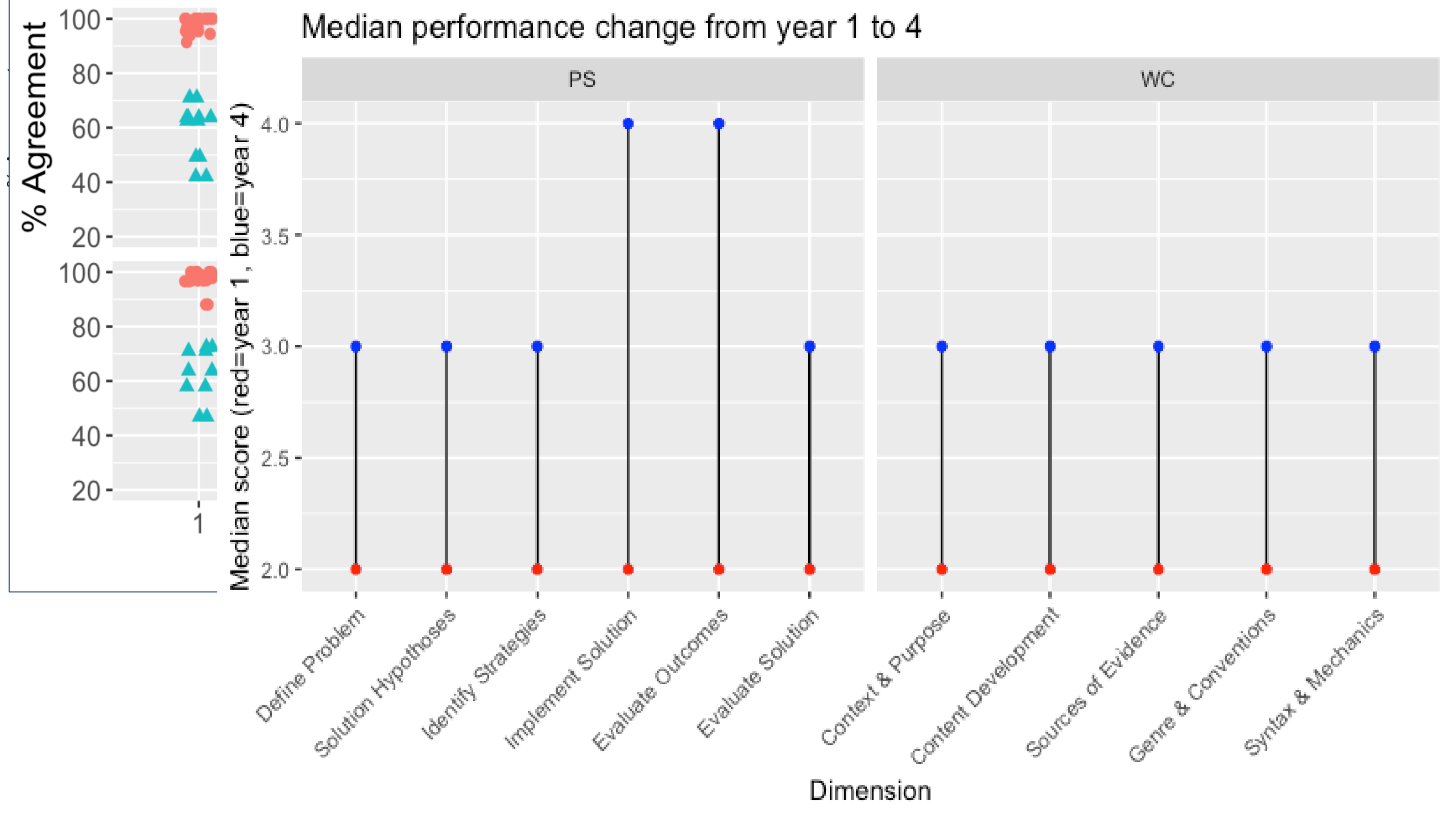




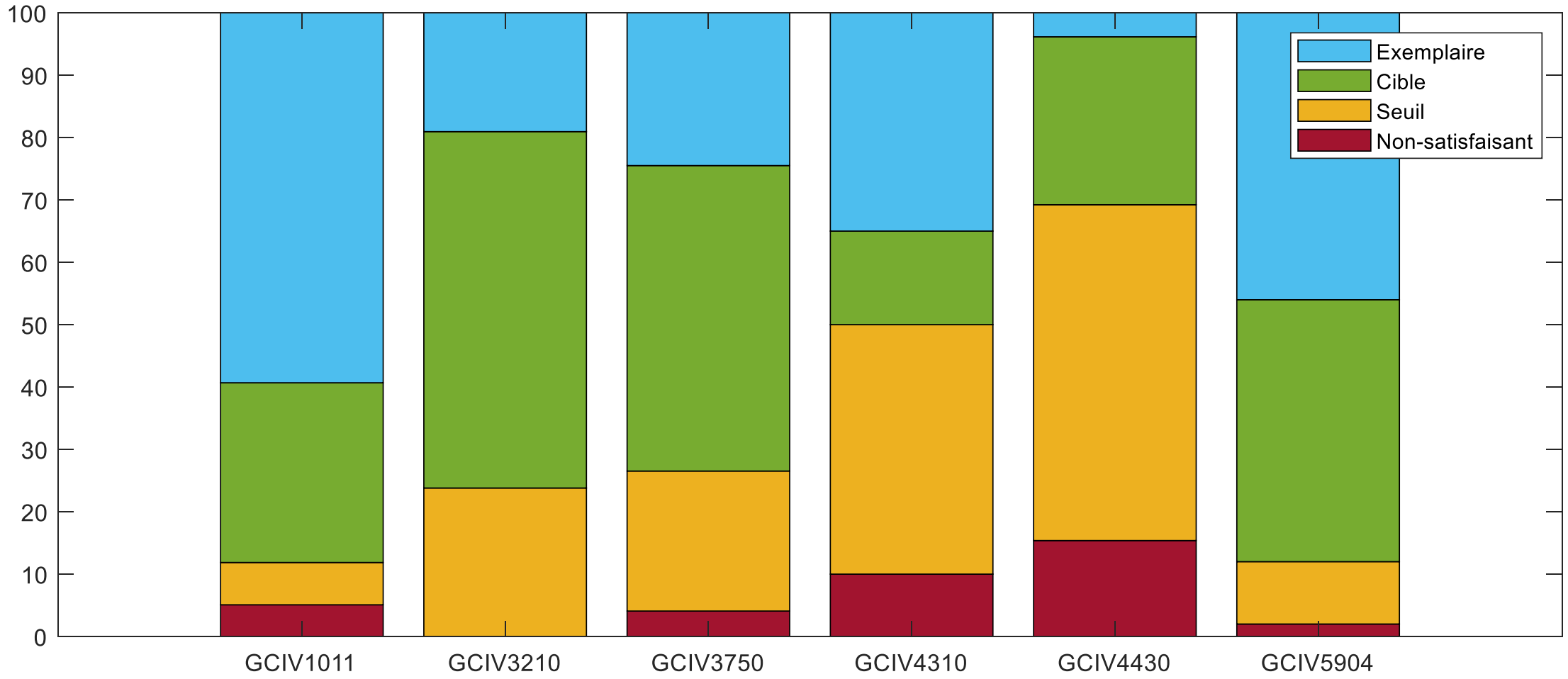
Factor	Approach
Aggregation target	Distribution of performance
Aggregation level	Indicator
Differentiation factors	Semesters
Reliability measure	



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


% Agreement
100
80
60
40
20
1
Median performance change from year 1 to 4
PS
WC
Median score (red=year 1, blue=year 4)
4.0
3.5
3.0
2.5
2.0
Define Problem
Solution Hypotheses
Identify Strategies
Implement Solution
Evaluate Outcomes
Evaluate Solution
Context & Purpose
Content Development
Sources of Evidence
Genre & Conventions
Syntax & Mechanics
Dimension

GA4.3 (QA4.3), same group, civil engineering



Task 2: Connect the factors to current institutional approaches

1. Identify how well the four factors describe the institution's approaches represented within the group. Is there a key factor that is not captured by that list of four? If so add it to your list of factors to consider
2. Briefly describe each institution's approach using the factors

ONLINE: We will turn microphones on and discuss, using the Google Doc in the shared Google Drive folder to collect our ideas.
FACE-TO-FACE: Form a small group and respond.

Task 2: Connect the factors to current institutional approaches

Report out:

1. Any key factors emerge that were not captured by the original list of four?
2. How well are institutional approaches captured by the factors? Are there two extremely different approaches between institutions represented at your table?

Task 3: Consider what aggregation means to key stakeholders

As a table group, identify what key stakeholders are looking for from aggregation:

- Course instructors
- Department administration
- Faculty administration
- CEAB visiting team

Task 3: Consider what aggregation means to key stakeholders

<u>Report out</u>: what are key stakeholders are looking for from aggregation?

- Course instructors
- Department committees, staff, and administration (which may also consult with broader stakeholders)
- Faculty committees, staff, and administration (which may also consult with broader stakeholders)
- CEAB visiting team

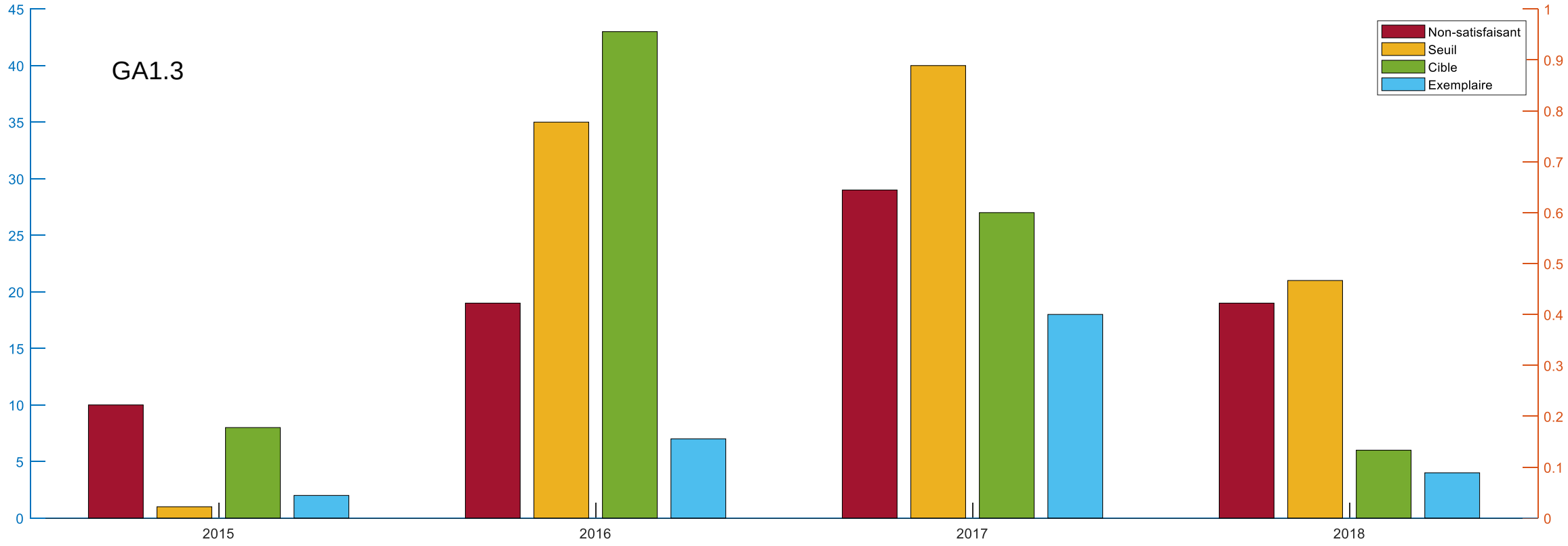
Aggregating data to draw meaningful conclusions

AMEGA EGAD Workshop May 2019

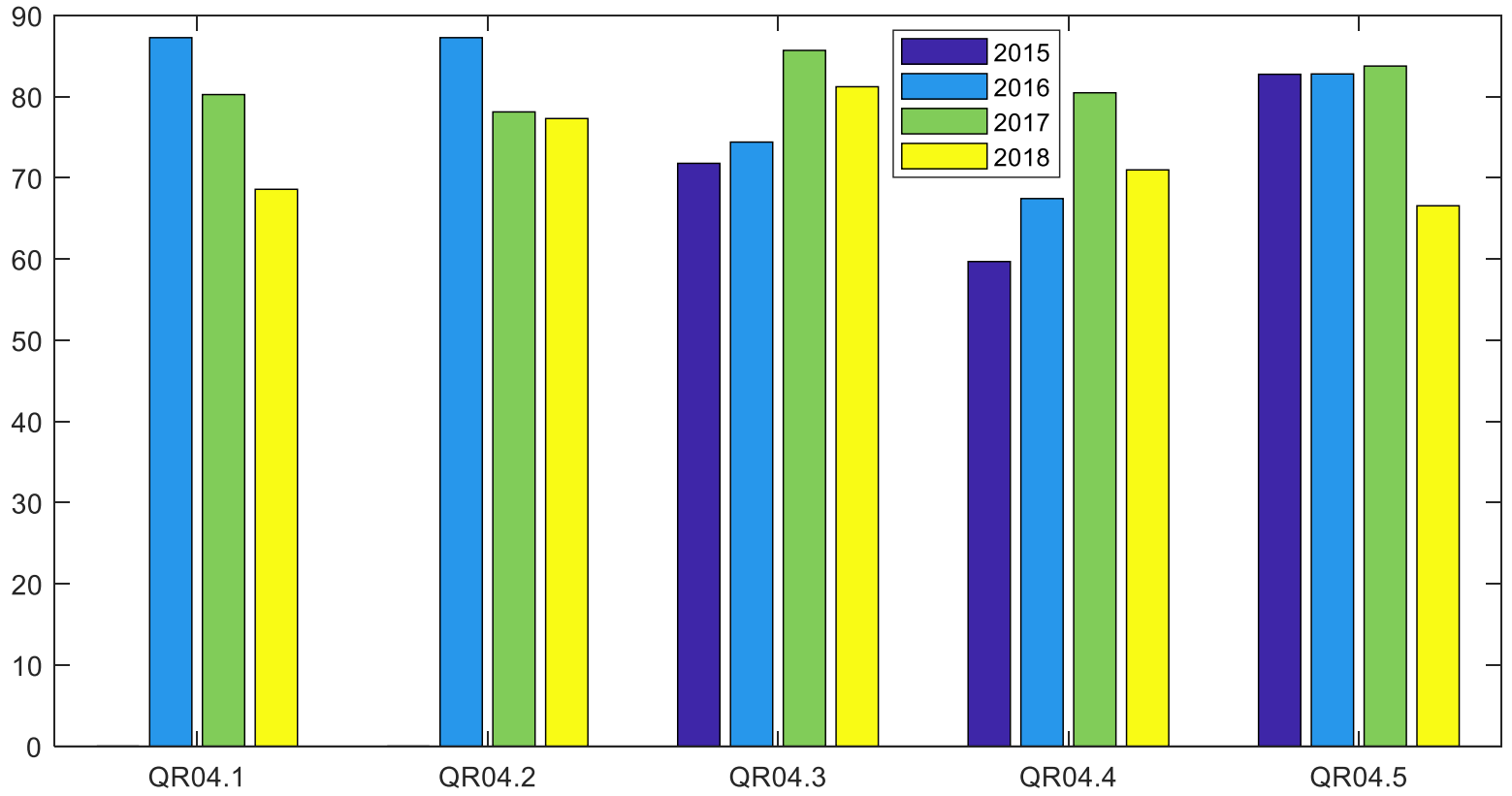
Resources at: bit.ly/AMEGA-EGAD-2019

Extra slides

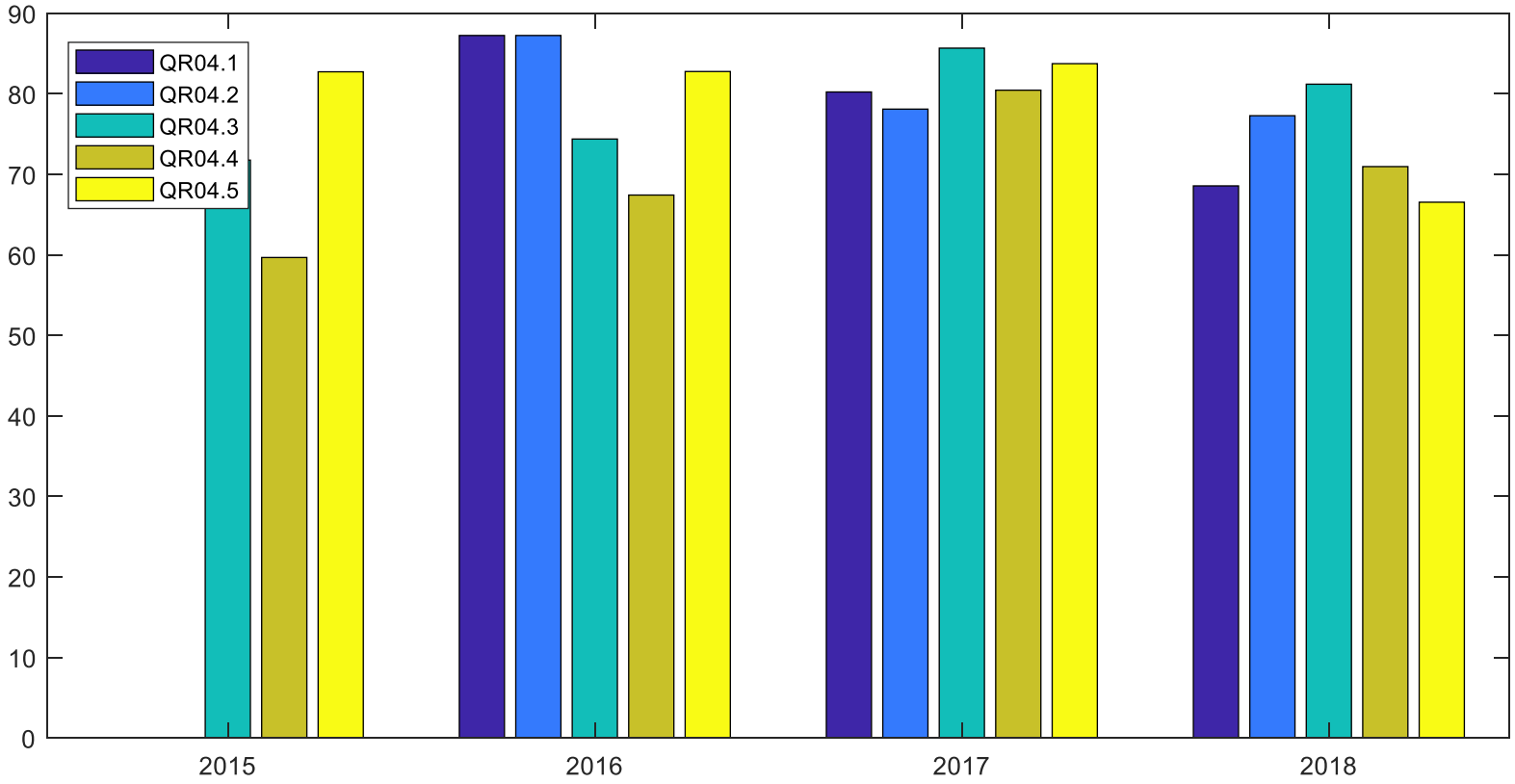
GA1.3 (QR1.3), aggregated data, by year



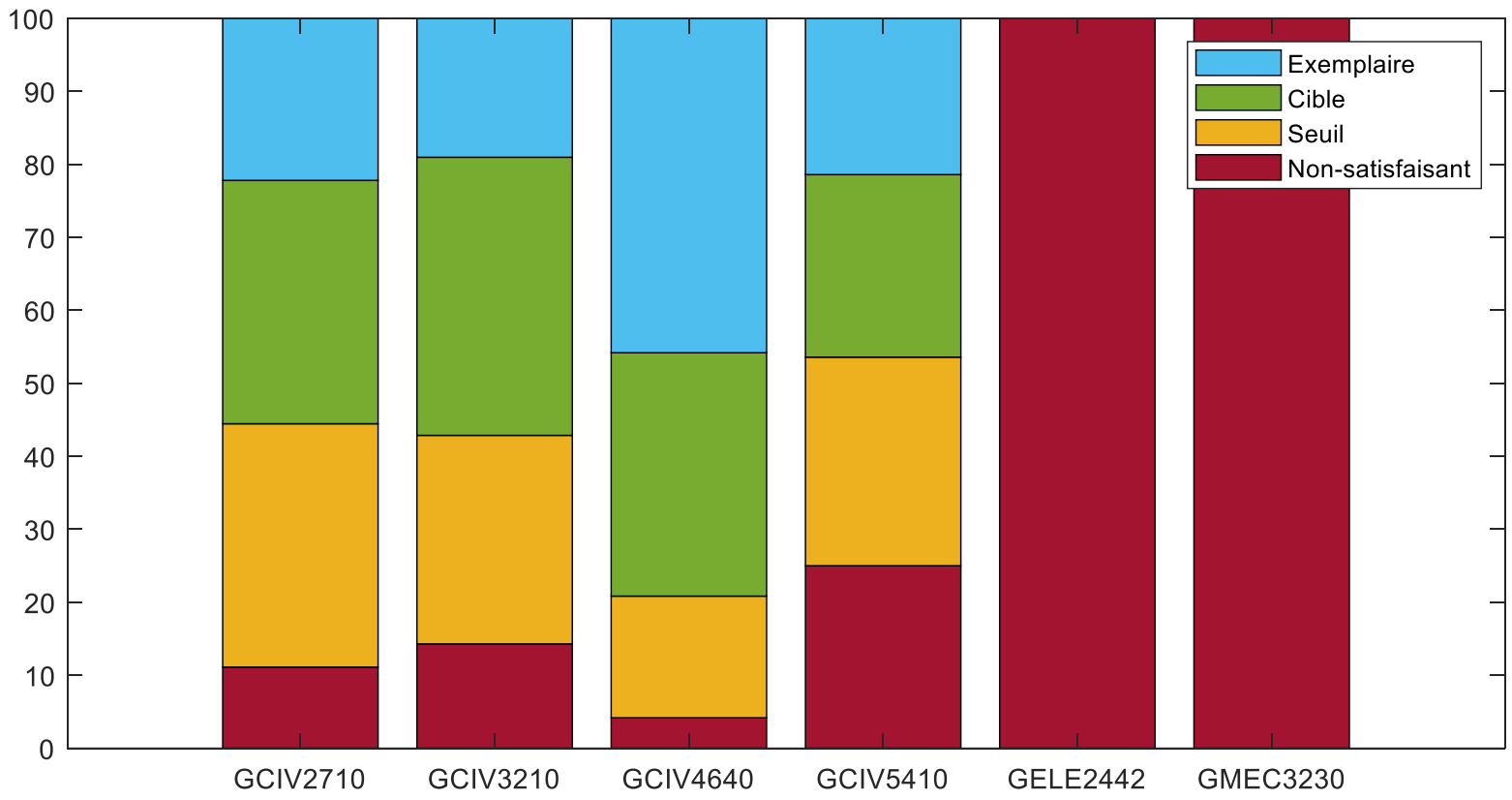
GA4(QR4), by indicator, 2015-2018



GA4(QR4), by year



GA1.4 (QR1.4), same group, civil engineering



GELE2442 and GMEC3230 are not in civil engineering; these are data from students that switched programs.

Task 3: What mix of aggregation approaches would meet the collective needs of stakeholders?

Consider needs of key stakeholders:

- Course instructors
- Department committees, staff, and administration (which may also consult with broader stakeholders)
- Faculty committees, staff, and administration (which may also consult with broader stakeholders)
- CEAB visiting team

Consider factors in aggregation:

1. Aggregation target: single value, distribution of performance, or qualitative description
2. Aggregated level: up to attribute, up to indicator within each attribute, up to task within indicator within attribute
3. Differentiation factor: differentiate by year, IDA level, program, student sub-group, student
4. Reliability measure: qualitative or quantitative instructor rating, correlation between tasks or years

Task 3: Report Out: What mix of aggregation approaches would meet the collective needs of stakeholders?

Consider needs of key stakeholders:

- Course instructors
- Department committees, staff, and administration (which may also consult with broader stakeholders)
- Faculty committees, staff, and administration (which may also consult with broader stakeholders)
- CEAB visiting team

Consider factors in aggregation

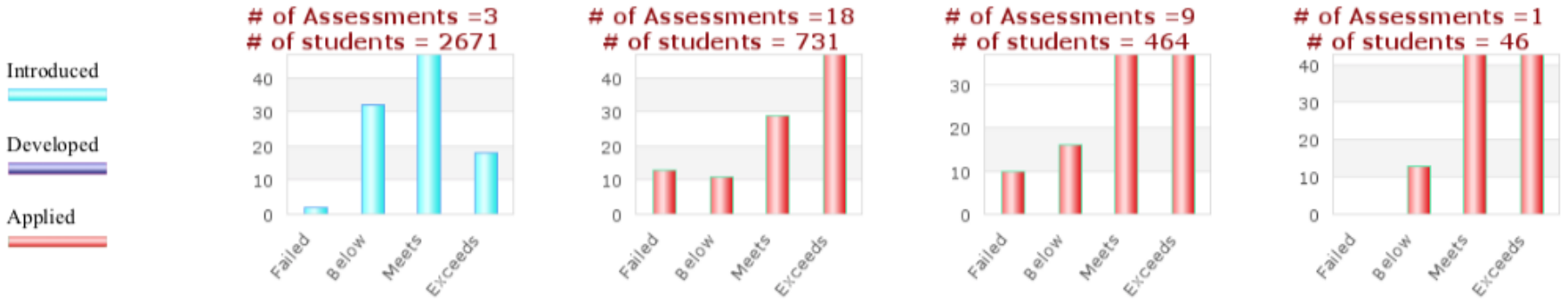
1. Aggregation target: single value, distribution of performance, or qualitative description
2. Aggregated level: up to attribute, up to indicator within each attribute, up to task within indicator within attribute
3. Differentiation factor: differentiate by year, IDA level, program, student sub-group, student
4. Reliability measure: qualitative or quantitative instructor rating, correlation between tasks or years

Graduate Attribute Distribution

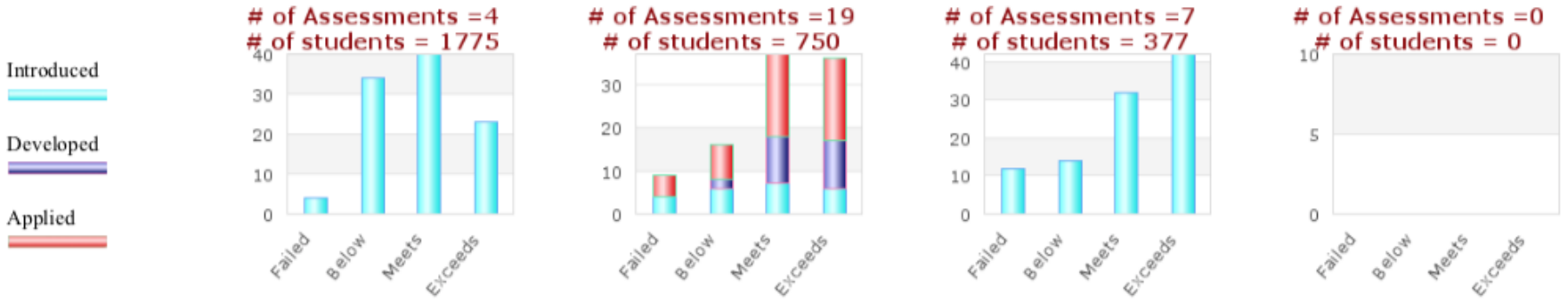
Graduate Attribute 1: 1st Year 2nd Year 3rd Year 4th Year

Knowledge base for engineering

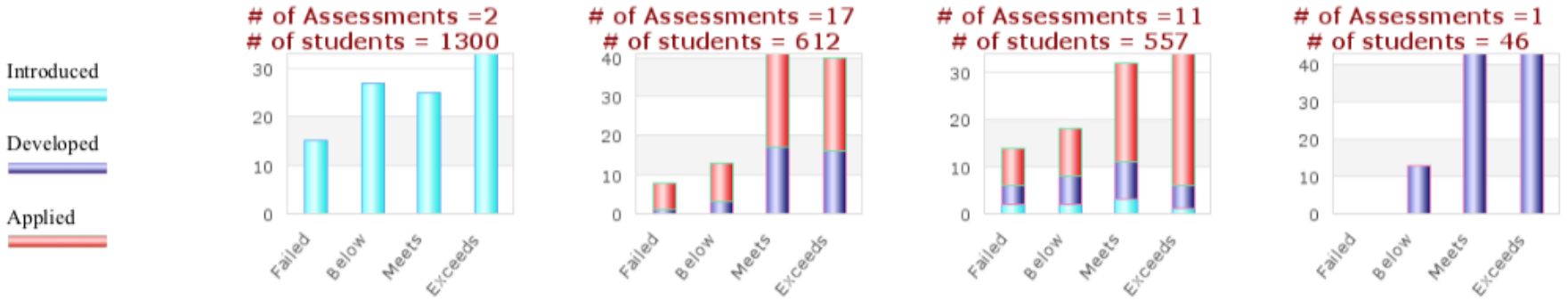
1: A Demonstrate competence in mathematics and modeling



1: B Demonstrate competence in natural sciences



1: C Demonstrate competence in specialized engineering knowledge appropriate to the program



Factor	Approach
Aggregation target	Frequency distribution: of performance
Aggregation level	Attribute
Differentiation factors	IDA, Year
Reliability measure	# students, # assessments,