OASIS: an online assessment for individual scores Methodology

C.K. Woo

Department of Asian and Policy Studies

The Education University of Hong Kong

oasis

Overview

ONE

Objective: To present a new online process for assessing individual contributions to a team project (e.g., a final year project of an undergraduate program).

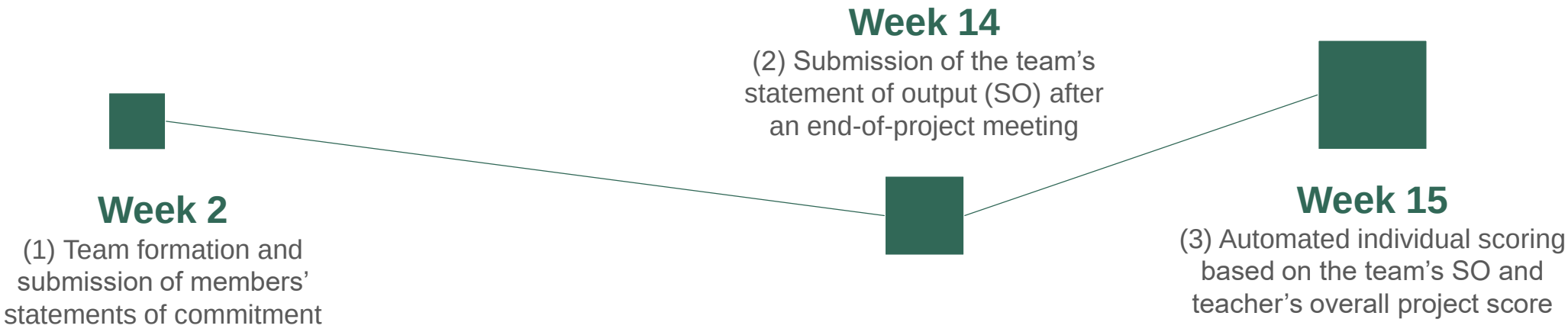
TWO

This process is fair, easy to use, low cost and perfectly general – a think-outside-the-box solution to a common problem faced by teachers worldwide.

THREE

It is based on my business and litigation experience and applied research in electricity ratemaking, performance-based regulation, and contract negotiations.

An online process for a 15-week teaching period



- Three current users at HK Poly U (Engineering), Goa Institute of Management in India (MBA) and UT Austin (Economics)

- Benefits of adoption:
  - General applicability to any course of any discipline
  - Time-efficient and low-cost implementation
  - Promotion of a team project's learning goals
  - Performance-based assessment with strong disincentives for free riding behavior
  - Furthering the overarching policy of fair assessment

oasis

Examples of current practice at HK Poly U: *shortcomings*

ONE

Use member-specific intermediate output (e.g., drafts of a power point presentation): ***time consuming sans a clear link to the final output that drives a project's overall grade.***

TWO

Make an individual member solely responsible for a specific portion of the team project (e.g., the project's literature review): ***what if the project's overall quality sucks?***

THREE

Use a combination of individual assessments based on (1) and (2) and an overall assessment of the team project: ***opaque, subjective and hence arbitrary.***

FOUR

Adjust the team's overall grade in light of evidence of individual contributions via a declaration of contribution, peer assessment and self-evaluation: ***no individual scores.***

FIVE

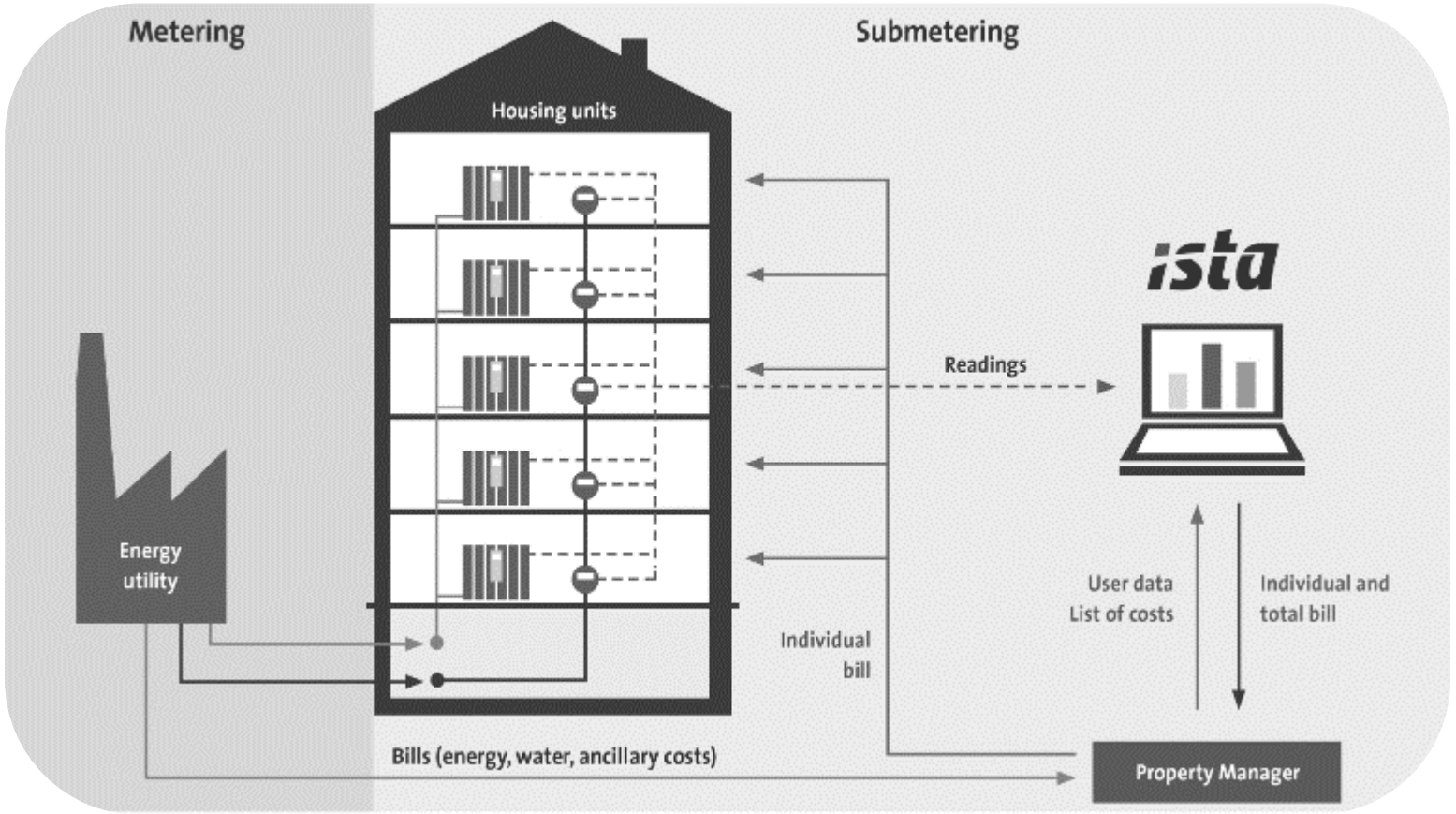
Give the entire team the same grade but prevent free riding by encouraging good team work and close monitoring: ***time consuming and no individual scores.***

oasis

Material and methods

- The Eureka moment
- Electricity ratemaking vs. individual scoring
- Criteria for an acceptable process
- Key components of the proposed process
- Calculation of a team's member-specific scores
- Preliminary comparison of methods

The Eureka moment: scoring individual contributions resembles submetering individual consumption



A teacher's scoring of individual contributions to a team project resembles a property manager's allocation of a building's total electricity bill among the individual units. The need for and usefulness of an individual scoring system greatly diminish under the assumption that all team members are responsible, hard-working and fair-minded individuals with similar abilities. But this assumption's validity is seldom known *a priori*, especially for a course with heterogeneous students randomly assigned to different teams.

Electricity ratemaking vs. individual scoring

Electricity ratemaking

- Under master metering, a renter of a 5-unit apartment building has a muted incentive to conserve because the renter pays a fixed *pro rata* share (20%) of the building's total bill.
- Solution: individual metering that enables an accurate billing of the renter's own kWh consumption at the applicable tariff.
- A sound rate design obeys the Bonbright principles - economic efficiency, fairness: cost-causation and no cross subsidization, low-cost implementation, easy understanding, objectivity, and transparency.

Individual scoring

- Under the single-score-for-all assessment method, a member has the perverse incentive of no penalty for an ill-gotten gain.
- Solution: individual scoring, which triggers the questions of how to estimate an individual's contribution and how to score the estimated contribution.
- These two questions highlights the conceptual connection of electricity ratemaking to academic assessment of individual contributions.

oasis

Criteria for an acceptable process

01 **It should advance a team project's learning goals (e.g., team work), thus precluding an assessment method solely based on individual responsibilities and deliverables.**

02 **It should discourage free riding behavior, a direct consequence of the single-score-for-all assessment method's perverse incentive of no penalty for an ill-gotten gain.**

- Unintended free ridership: a capable member does most of the work because of the member's concerns of a low grade and other members' slow progress or inability.

03 **It should be easy to implement sans costly monitoring and evaluation, thus excluding assessment methods based on each member's log of time spent and intermediate outputs.**

oasis

Criteria for an acceptable process

04

It should be fair:

- horizontal equity – similar performances imply similar scores;
- vertical equity – different performances imply different scores; and
- anonymous equity - individual scores do not depend on student demographics and course attributes, thus imparting fairness among students and across courses.

05

It should be transparent and objective as required by sound regulatory governance because an academic assessment method is defined by a set of rules and regulations.

By satisfying the above criteria, the proposed process is a superior alternative to an online peer assessment system like CATME, which is difficult to use or WebPA, which is vulnerable to gaming

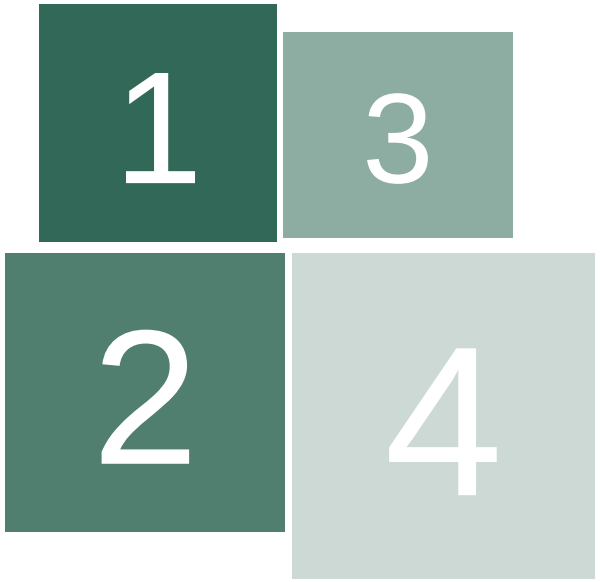
Key components of the proposed process

Statement of commitment (SC)

Each member **must**: (a) sign a SC, reflecting the course's expectation of fair contribution and high cooperation; and (b) submit the SCs of all members as part of the team's mandatory declaration of membership (no undue burden).

Statement of output (SO)

A team's final report **must** include a pro forma SO of each member's output contribution to the project's completion and quality as part of the mandatory requirement of team members' reflections (no undue burden).



Individual contribution

Based on the SO, the individual contribution of a given member (e.g., A) is a **median** estimate based on **other** members' assessments of A's contribution.

Individual scoring

Using the results from (3) and a team project's overall score as input, a simple spreadsheet automatically produces a member's individual score.

Example

Example of a statement of commitment

As a valuable member of this 5-person team, I commit to actively cooperate and diligently contribute approximately 20% (= 1/5) of the project's deliverables. To ensure the project's timely completion and high quality, my primary areas of responsibility are marked by "√" below:

- ☐ Topic selection
- ☐ Research plan: what and when to do?
- ☐ Literature review
- ☐ Data collection and analysis
- ☐ Graphics, tables and charts
- ☐ Discussion of results
- ☐ Presentation preparation
- ☐ Final report preparation
- ☐ Editing and proofreading
- ☐ Project management and coordination
- ☐ Other _____

Name (Student ID) _____; Signature _____; Date _____

Example of a statement of output

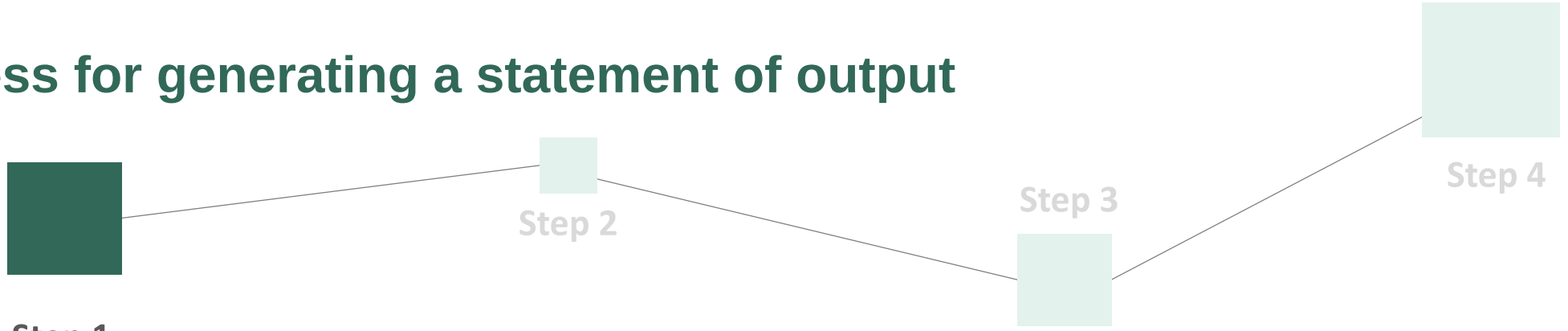
Panel A: Initial assessment

Student ID *k*	Own and others' contributions: C_{km} for *m* = student ID					
	A	B	C	D	E	Reason
A	0.5	0.2	0.1	0.1	0.1	
B	0.3	0.3	0.2	0.1	0.1	
C	0.2	0.2	0.2	0.2	0.2	
D	0.3	0.2	0.2	0.3	0.0	
E	0.3	0.1	0.2	0.2	0.2	

Panel B: Final assessment determined by consensus ____ or majority vote ____, with changes highlighted in red

Student ID *k*	Own and others' contributions: C_{km} for *m* = student ID					
	A	B	C	D	E	Reason
A	0.3	0.2	0.2	0.2	0.1	
B	0.3	0.3	0.2	0.1	0.1	
C	0.2	0.2	0.2	0.2	0.2	
D	0.3	0.2	0.2	0.2	0.1	
E	0.3	0.2	0.2	0.2	0.1	

Process for generating a statement of output

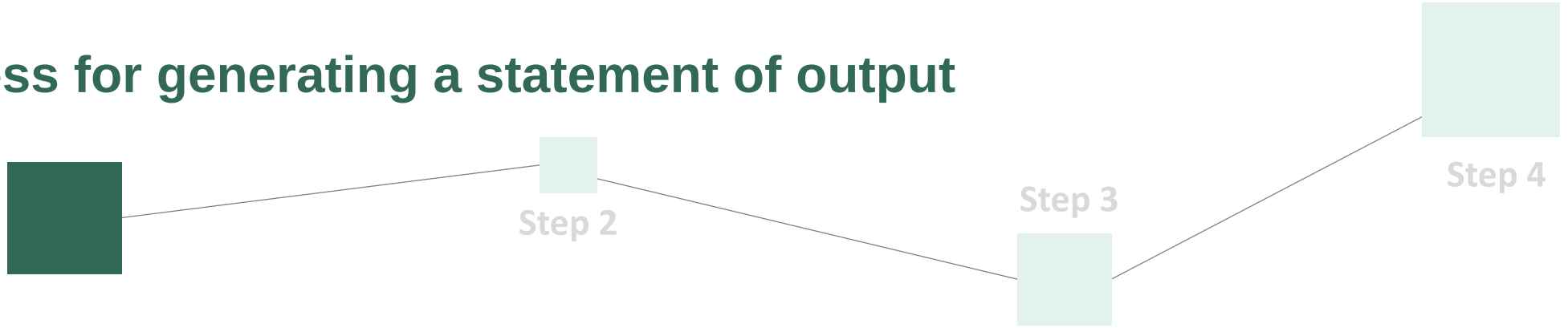


Step 1

After the project's completion, members hold a 1-hour meeting to reflect on their experiences, using the SCs to frame the discussion of member-specific contributions.

- Using documented evidence, each member recaps his/her contributions to the project, thereby discouraging unsubstantiated claims. Each member then announces his/her assessment with reasons of his/her own and other members' contributions.
- In a 5-person example, A announces (C_{AA}, C_{AB}, …, C_{AE}), the contributions of members A, B, C, D and E that must sum to 1.0. If A announces a high own contribution (e.g., C_{AA} = 0.8), his/her announcement of other members' total contribution is correspondingly low (i.e., C_{AB} + … + C_{AE} = 0.2).
- At a team's unlikely request, a teacher attends the meeting as a mediator, akin to the process used in an arbitration. Video recording discourages abusive and collusive behavior. It also generates a complete report of member interactions, useful for resolving a formal complaint and conducting research in teaching and learning.

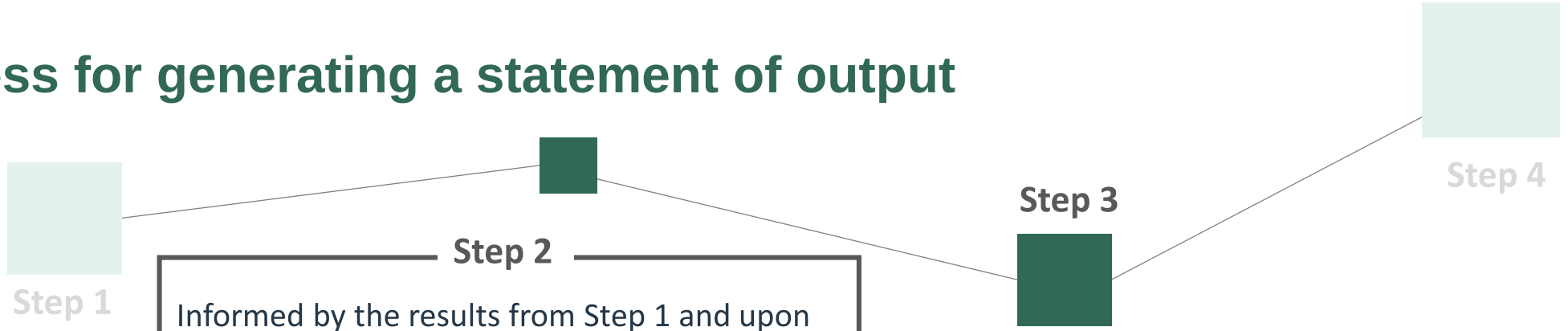
Process for generating a statement of output



Step 1

- The process described in the previous slide may be replaced by an alternative that does not involve open discussion and announcement of member-specific contributions. Thanks to the suggestion of a participant at a HKUST workshop, anonymous submission of self and peer assessment data by each team member directly to a teacher eliminates the distress that some team members may find unnecessary and hurtful.
- If a teacher chooses anonymous submission, Steps 2 to 4 no longer apply, as the self and peer assessment data thus obtained are sufficient for calculating individual contributions to a team project's completion and overall quality, see Slide 17 below.

Process for generating a statement of output



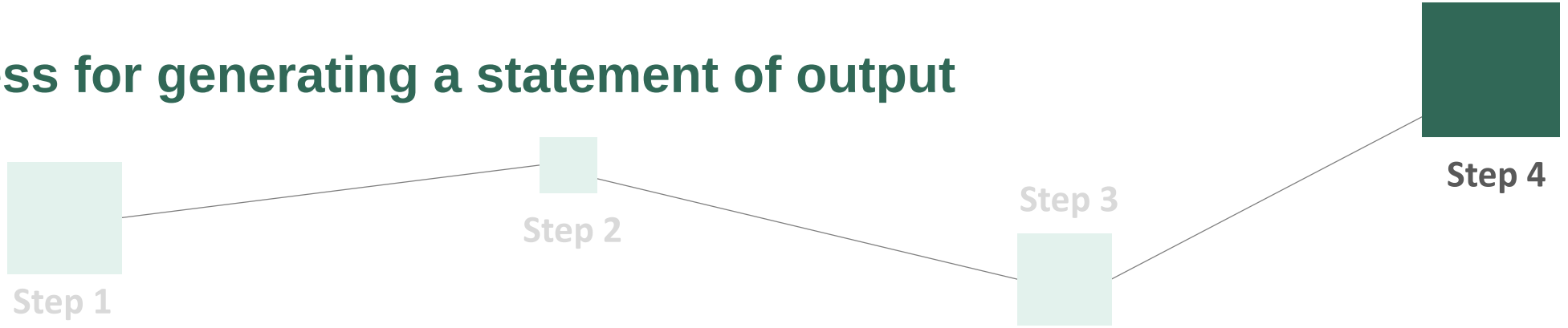
Step 2

Informed by the results from Step 1 and upon further discussions, each member revises his/her assessments. The members may input their revised assessments into an online spreadsheet to calculate individual scores based on an assumed overall project score (e.g., 80 on a 100-point scale). This informs each member of the revised assessments' impact on individual scores, thus facilitating the bargaining process.

Step 3

Each member announces his/her final assessments, after seeing other members' revised announcements and the ensuing individual scores. This encourages settlement via peer pressure, desire for consensus, and if necessary, teacher mediation.

Process for generating a statement of output



Step 4

If the team fails to reach a settlement despite teacher mediation and video recording, the members' final assessments are to be based on majority voting, which may invite collusion by some members. Hence, a member allegedly injured by the voting outcome can request a formal investigation that requires all members to submit affidavits with evidentiary details, a daunting task that preemptively deters collusive behavior.

oasis

What are the benefits of the process?

ONE

The process is perfectly general, applicable to any course of any discipline, thus imparting fairness among students and across courses.

TWO

The process is time-efficient because it encourages settlement, thus preempting subsequent complaints of unfair grading that are hard to handle absent its adoption

THREE

Students appreciate the consequences of commitment, diligence, cooperation, interpersonal skills, leadership, resource planning and management, …, etc.

FOUR

Students learn the art of effective negotiation that is supported by sound reasoning and convincing evidence, an important soft skill that prepares students to enter the workplace.

FIVE

A teacher can use the process' outcome to fairly grade individual contributions in compliance with the overarching principle of horizontal, vertical and anonymous equities.

Individual contribution calculation for a 5-person team

- A's contribution is S_A = median of (C_{BA}, C_{CA}, C_{DA} , C_{EA}) = other members' assessments of A. It is less vulnerable to gaming. Suppose A shamelessly gives himself/herself a high assessment. A's dishonest self-claim (e.g., 30% instead of 10%) cannot benefit A because it does not enter into A's calculated contribution.
- Discouragement of gaming may also come from the tit-for-tat behavior of other members: if A gives B a low assessment, B may retaliate with a low assessment of A.
- The median-based calculation is less affected by a member's severely biased assessments of own and others' performances than WebPA's mean-based calculation, thanks to a median being less sensitive to outliers than a mean. A useful analogy is the scoring system for diving in the Olympics Games where the two highest and two lowest scores of seven judges are not used to determine a diver's performance.

Individual score calculation

A's score is G_A = min[(D_A / F) G, αG, 100], where $D_A = S_A / (S_A + \ldots + S_E)$ = A's adjusted contribution share because S_k may not sum to 1.0; F = equal share = 0.2 for the 5-person team; α = preset scalar > 1; and G = overall project score. Thus, (D_A / F) is an estimated extent of A's contribution to G relative to the equal share F.

G_A is capped at αG or 100 to remedy the odd outcomes in rare but possible scenarios:

- **Scenario 1**: The team project's overall score is G = 20 and A contributes 100% (i.e., D_A = 1 and D_B = … = D_E = 0). Absent the αG cap, A's final score is 100 [= (1.0/0.2) 20], a silly outcome that ignores the project's poor overall quality.
- **Scenario 2**: A team's overall score is G = 90 and A contributes 100%. Absent the 100 cap, A's project score would be 450 [= (1.0/0.2) 90].

How to determine α

The determination of α may be based on (a) a teacher's view on the maximum individual score as a multiple of *G*; or (b) an algorithm that makes the distribution of the teams' averages of individual scores to closely match that of the teams' overall project scores given by the teacher.

- Example of (b): Define $G(j)$ = team j's overall score and $A(j, a)$ = team j's equally-weighted average of individual scores for a given α. The variance of $A(j, \alpha)$ around $G(j)$ is $V(\alpha) = S_j [A(j, \alpha) - G(j)]^2$. After the process' initial implementation, one can use the course's recorded data to compute $V(\alpha)$ for $\alpha \in \{1.1, \ldots, M = \text{maximum multiple of } G\}$. The optimal α is α^* so that $V(\alpha^*) = \min[V(a = 1.1), \ldots, V(a = M)]$, an Excel calculation that resembles a bill impact analysis for identifying a rate design *sans* extreme distributional effects.

A completely solved numerical example

Individual contributions

Member k	Own and others' contribution assessments by member k				
	A	B	C	D	E
A	0.3	0.2	0.2	0.2	0.1
B	0.3	0.3	0.2	0.1	0.1
C	0.2	0.2	0.2	0.2	0.2
D	0.3	0.2	0.2	0.2	0.1
E	0.3	0.2	0.2	0.2	0.1
S_k	0.3	0.2	0.2	0.2	0.1
D_k / F	1.5	1.0	1.0	1.0	0.5

A completely solved numerical example

Individual scores

Overall score G	Individual scores at $\alpha = 1.5$				
	A	B	C	D	E
10	15	10	10	10	5
20	30	20	20	20	10
30	45	30	30	30	15
40	60	40	40	40	20
50	75	50	50	50	25
60	90	60	60	60	30
70	100	70	70	70	35
80	100	80	80	80	40
90	100	90	90	90	45
100	100	100	100	100	50

Does the individual score calculation make sense? Yes

Typical cases: members making contributions of varying amounts

- Case 1: $D_A = 0$ and $G_A = 0$, thus punishing A, a free rider with zero contribution.
- Case 2: $D_A = F$ and $G_A = G$ = overall score, reflecting that when A makes a contribution that matches F, A receives G as expected.
- Case 3: $D_A > F$, $G_A > G$, thus reflecting that when A makes an above-F contribution, A receives a higher score.

Extreme cases: one member doing the entire project

- Case 4: $D_A = 1$ and $G = 20$, yielding $G_A = 30$ at $\alpha = 1.5$ and $G_B = \ldots G_E = 0$.
- Case 5: $D_A = 1$ and $G = 50$, yielding $G_A = 75$ at $\alpha = 1.5$ and $G_B = \ldots G_E = 0$.
- Case 6: $D_A = 1$ and $G = 70$, yielding $G_A = 100$ at $\alpha = 1.5$ and $G_B = \ldots G_E = 0$.

oasis

What are the benefits of the scoring rule?

ONE

A teacher can use the rule to fairly assess individual contributions, thereby obeying the overarching principle of horizontal, vertical and anonymous equities.

TWO

It addresses concerns of free riding and unfair grading.

THREE

It is perfectly general, applicable to any course of any discipline.

FOUR

Students appreciate the essence of a well designed incentive scheme that clearly awards a high (low) score for high (low) performance.

FIVE

Students can use the rule to learn effective negotiation in the end-of-project meeting.

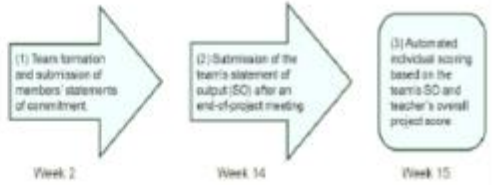
Preliminary comparison of methods

Criteria	CATME	iPeer	WebPA	OASIS
Conceptual validity	Excellent	Excellent	Excellent	Excellent
Ease of implementation by teachers	Good	Very good	Very good	Very good
Ease of use by students	Good	Very good	Very good	Very good
Performance-based grading	Very good	Very good	Very good	Very good
Disincentive to free ride	Good	Good	Good	Average
Discouragement of strategic behavior	Average	Average	Average	Very good
Fairness in grading	Good	Good	Good	Very good
Overall	Good	Very good	Very good	Excellent

Excellent Very good Good Average Poor

- This comparison represents our initial thoughts, prepared solely for eliciting your views on OASIS.
- Please make your own comparison when considering OASIS for your course's assessment.
- CATME, iPeer and WebPA do not have strong measures against free riding and strategic behavior, leading to relatively low ratings of their fairness in grading.

High-level comparison of attributes

Attribute	iPeer	CATME	WebPA	OASIS
Platform	Web-based	Web-based	Web-based	Web-based and Excel worksheet
Basis for determining individual scores	Peer assessment	Survey data analysis	Peer assessment	Peer assessment
Calculation	Point distribution by the users	Statistical method	Mean-based scoring algorithm	Excel-based median estimation
Cost of implementation	• Subscription fee: N.A. • Teachers need to design and setup evaluation exercises	• A minimum license fee of 25 unique students or $50.00 to single instructors; $2.00 per year per unique student who used CATME in the previous academic year (http://info.catme.org/licensing/) • Teachers need to collect survey data and perform the statistical analysis	• No charge for download, install and use. • Cost of ownership and license conditions to be adhered to.	• Subscription fee: None • Students only need to fill in the statement of commitment (SC) and statement of output (SO) • Survey data collection is needed only if a teacher wants to know the effectiveness of OASIS
Ease of use	A **five**-step procedure: 	An **eight**-step procedure (http://catme.org): 		A **three**-step procedure: 

Both WebPA and OASIS use peer assessment data to compute individual scores. The main difference is that unlike WebPA, OASIS uses (a) a negotiation process to generate the statement of output; and (b) a median estimation of member-specific relative contributions, thus mitigating the problems of free riding and strategic behavior (gaming).

oasis

C.K. Woo

PhD (Economics, UC Davis)

- He is Professor of Asian and Policy Studies and Associate Dean (Undergraduate Studies and Student Learning) of Faculty of Liberal Arts and Social Sciences at Education University of Hong Kong (https://www.eduhk.hk/aps/people/prof-woo-chi-keung/) and Senior Partner (now on leave) of Energy and Environmental Economics, Inc. (www.ethree.com), a 40-person consulting firm headquartered in San Francisco.
- With 30 years of industry experience, he has successfully participated in regulatory proceedings, contract negotiations, and dispute arbitrations in California, Missouri, Texas, British Columbia, Alberta, Ontario, Quebec, and Hong Kong.
- He has published over 140 papers in such scholarly journals as *Energy Policy, The Energy Journal, Energy Economics, IEEE Transactions, Journal of Regulatory Economics, Energy Law Journal, Journal of Public Economics, Quarterly Journal of Economics*, and *OMEGA.*
- Recognized by *Who's Who in America*, he is a senior fellow of the United States Association for Energy Economics and an editorial board member of *Energy*, *The Energy Journal* and *Energy Policy.*