Adaptive Comparative Judgement In Open-ended Design Scenarios

Eva Hartell[1,2], Greg Strimel[3], Scott Bartholomew[3] & Emily Yoshikawa[3]
[1]*KTH Royal Institute of Technology, Stockholm, Sweden*
[2] *Haninge municipality, Sweden*
[3]*Purdue University, USA*

The 44th International Association for Educational Assessment (IAEA) Annual Conference
Assessment & Big Data
9–14th September 2018. Oxford University, UK

Abstract

Adaptive comparative judgment (ACJ) has proven to be a valid, reliable, and feasible method for assessing student performance in open-ended design scenarios. In addition to the use of ACJ for purely assessment and evaluation, research has demonstrated an opportunity to identify the design values of judges involved with the ACJ process and feed that into classroom practice and possible curriculum design. The potential for ACJ, as a tool for understanding cultural design values, and potentially facilitating international collaboration, is intriguing. Therefore, this study established three panels of judges from USA, UK and Sweden, with the purpose of unpacking teachers' assessment practices. These three panels assessed a body of 760 American student works, in technology/ engineering education, using the ACJ method. The similarities, differences, and quantitative and qualitative data findings from these assessment results were analyzed, revealing distinct design values, preferences, and differences for each group of judges from the different locations. This paper will show possible use of ACJ on larger scale to find out and explicate criteria for success in open-ended design tasks to inform formative assessment practices. The paper will tie literature together and provide an overview of possible use of ACJ to inform future work within the field of assessment.

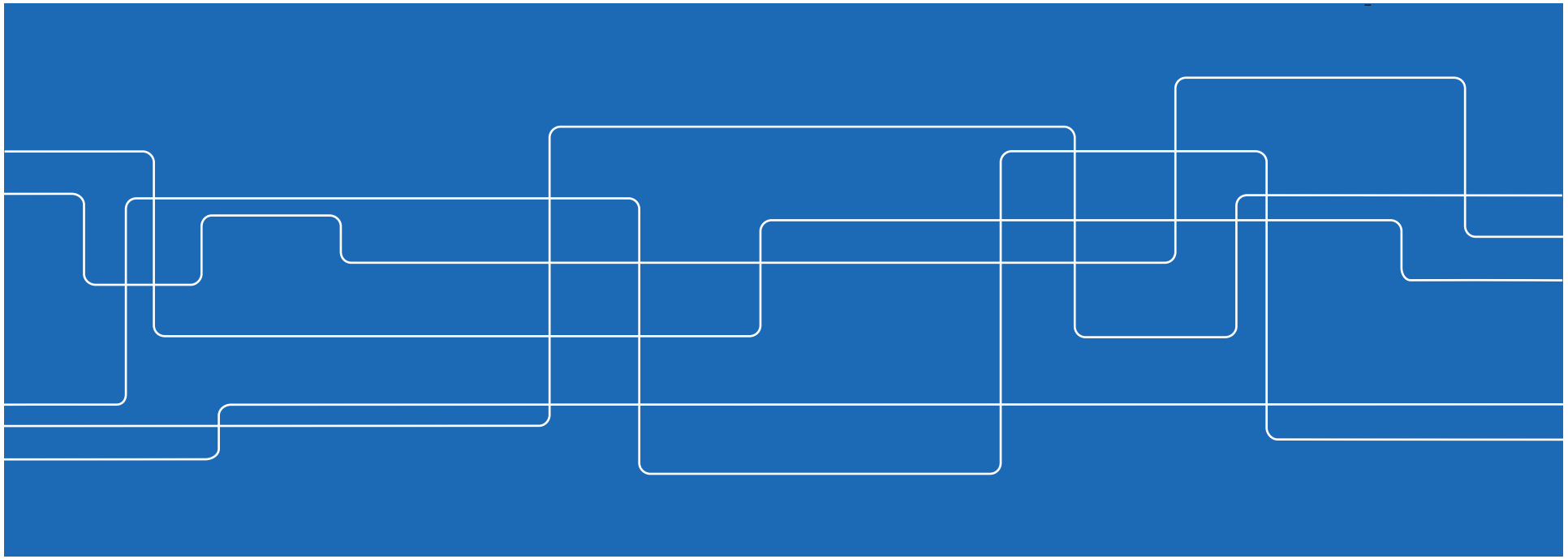
Adaptive Comparative Judgment in Open-Ended Design Scenarios

Eva Hartell[1,2], Greg Strimel[3], Scott Bartholomew[3] & Emily Yoshikawa[3]

[1]*KTH Royal Institute of Technology, Stockholm, Sweden*
[2] *Haninge municipality, Sweden*
[3]*Purdue University, USA*

The 44th International Association for Educational Assessment (IAEA) Annual Conference
9–14th September, 2018. Oxford University, UK

@EvaHartell
@strimelg
@IAEA_2018



IAEA 2018 Oxford

Adaptive Comparative Judgment (ACJ) has proven to be:

- Valid
- Reliable
- Feasible

method for assessing student performance in open-ended design scenarios.

Beyond purely assessment and evaluation, research has demonstrated an opportunity to inform classroom practice and curriculum design by using the ACJ process to identify different design values.

IAEA 2018 Oxford

This study established three panels of judges from the:

- USA
- UK
- Sweden

with the purpose of unpacking teachers' assessment practices.

These panels assessed **760 American student works** using the ACJ method

Similarities and differences from these assessment results were analyzed, revealing distinct **design values**, **preferences**, and **differences** for each group of judges.

IAEA 2018 Oxford

Purpose:

1. Explore the possible use of ACJ to investigate and explicate criteria for success in open-ended design tasks in an effort to inform formative assessment practices.

2. Tie literature together and provide an overview of possible use of ACJ to inform future work within the field of assessment.

Curriculum Differences
Technology / Engineering Education

Manual Training
Manual Arts
Industrial Arts
Technology Education

Technology & Engineering Education

Sloyd (Slöjd)
Science (natural/ social)

Teknik

Teknik Technology

Design & Technology

Different assessment practices!

Assessment in sTEm is Difficult

Context! Teknik? D&T? Engineering? Technology education?

Purposes and Content

Crowded and Broad Curriculum

Open-ended design challenges - difficult to assess with e.g. rubrics

Design *and* History of Technology
- not taught the same way-not assessed the same way

Reasonable level of knowledge? Progress? Scaffolding

Construct definition

Preconditions for teaching and learning….affordance

Unpacking Teachers' Assessment Practices
Digging Deeper than Documents

C.f. Bartholomew, S. R. (2016). A Mixed-Method Study of Mobile Devices and Student Self-Directed Learning and Achievement During a Middle School STEM Activity (Doctoral dissertation, Utah State University).

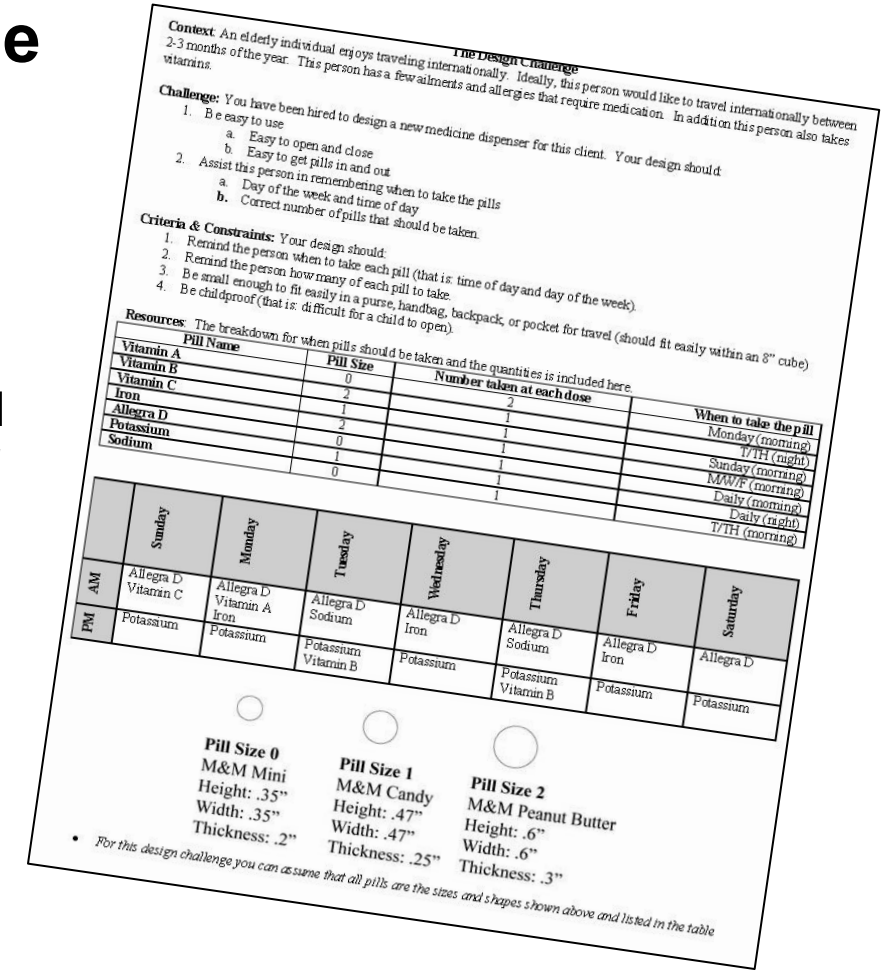
Starting Point
The Design Challenge

An elderly individual enjoys travelling internationally.

Ideally, this person would like to travel internationally between 2-3 months of the year.

This person has a few ailments and allergies that require medication.

In addition this person also takes vitamins.



The Design Challenge

Context: An elderly individual enjoys traveling internationally. Ideally, this person would like to travel internationally between 2-3 months of the year. This person has a few ailments and allergies that require medication. In addition this person also takes vitamins.

Challenge: You have been hired to design a new medicine dispenser for this client. Your design should:

1. Be easy to use
   a. Easy to open and close
   b. Easy to get pills in and out
2. Assist this person in remembering when to take the pills
   a. Day of the week and time of day
   b. Correct number of pills that should be taken.

Criteria & Constraints: Your design should:

1. Remind the person when to take each pill (that is: time of day and day of the week).
2. Remind the person how many of each pill to take.
3. Be small enough to fit easily in a purse, handbag, backpack, or pocket for travel (should fit easily within an 8" cube)
4. Be childproof (that is: difficult for a child to open).

Resources: The breakdown for when pills should be taken and the quantities is included here.

Pill Name	Pill Size	Number taken at each dose	When to take the pill
Vitamin A	0	2	Monday (morning)
Vitamin B	2	1	T/TH (night)
Vitamin C	1	1	Sunday (morning)
Iron	2	1	MW/F (morning)
Allegra D	0	1	Daily (morning)
Potassium	1	1	Daily (night)
Sodium	0	1	T/TH (morning)

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
AM	Allegra D Vitamin C	Allegra D Vitamin A Iron	Allegra D Sodium	Allegra D Iron	Allegra D Sodium	Allegra D Iron	Allegra D
PM	Potassium	Potassium	Potassium Vitamin B	Potassium	Potassium Vitamin B	Potassium	Potassium

Pill Size 0
M&M Mini
Height: .35"
Width: .35"
Thickness: .2"

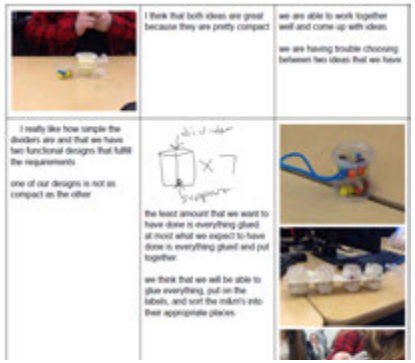
Pill Size 1
M&M Candy
Height: .47"
Width: .47"
Thickness: .25"

Pill Size 2
M&M Peanut Butter
Height: .6"
Width: .6"
Thickness: .3"

- *For this design challenge you can assume that all pills are the sizes and shapes shown above and listed in the table*

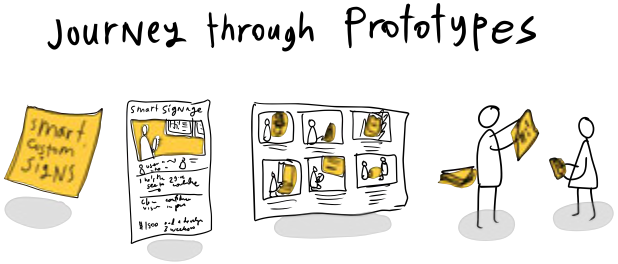
Methods

3 panels of judges from different countries



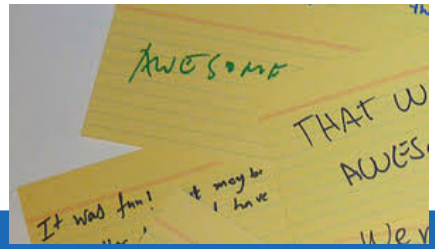
706 12-14 year olds from US worked in 176 groups to complete an open-ended design problem

Pictures of each of the groups prototypes and their portfolios were uploaded into separate ACJ sessions (one for prototypes and one for portfolios)

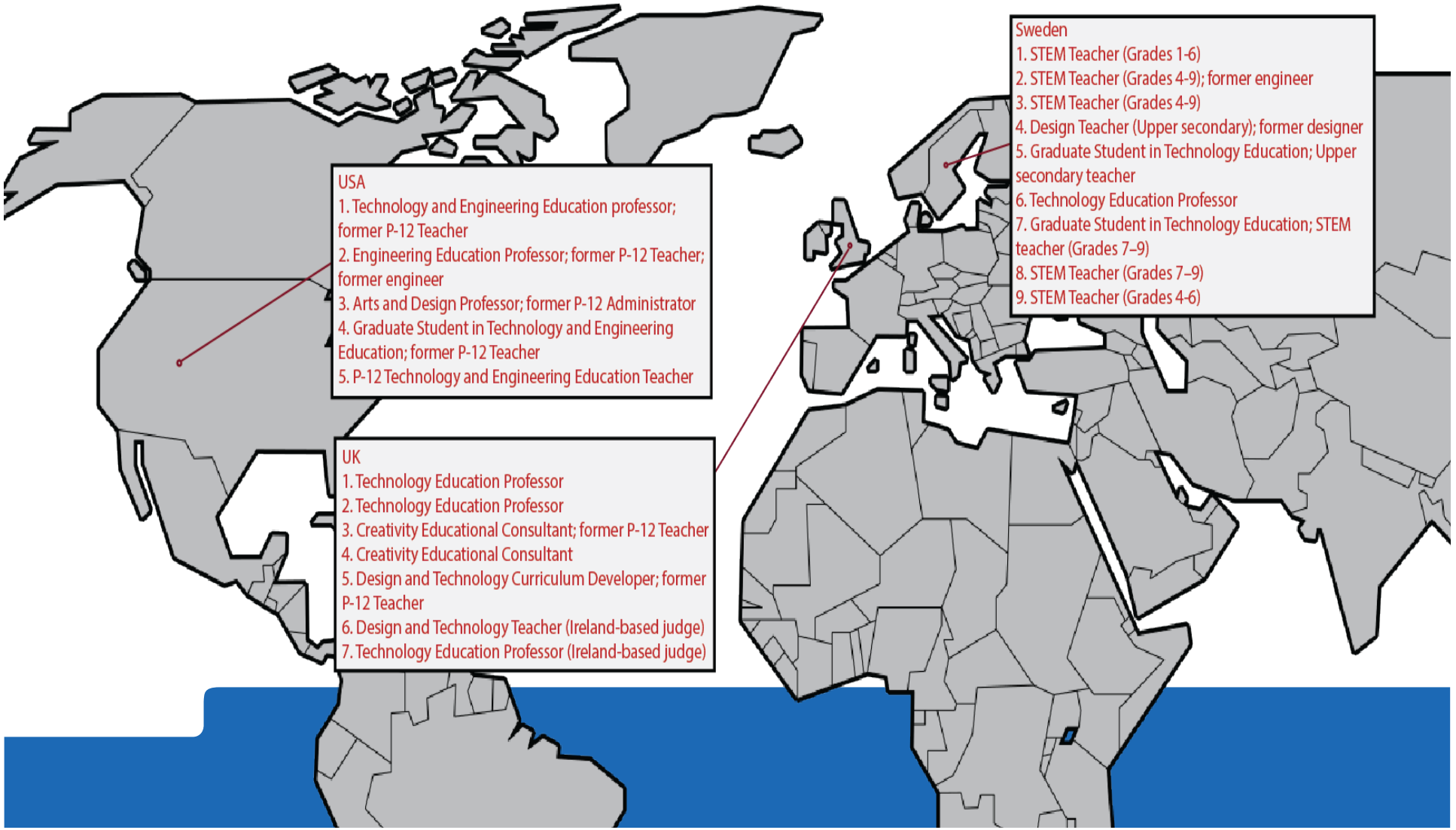


Each of the panels judged both the prototypes and portfolios

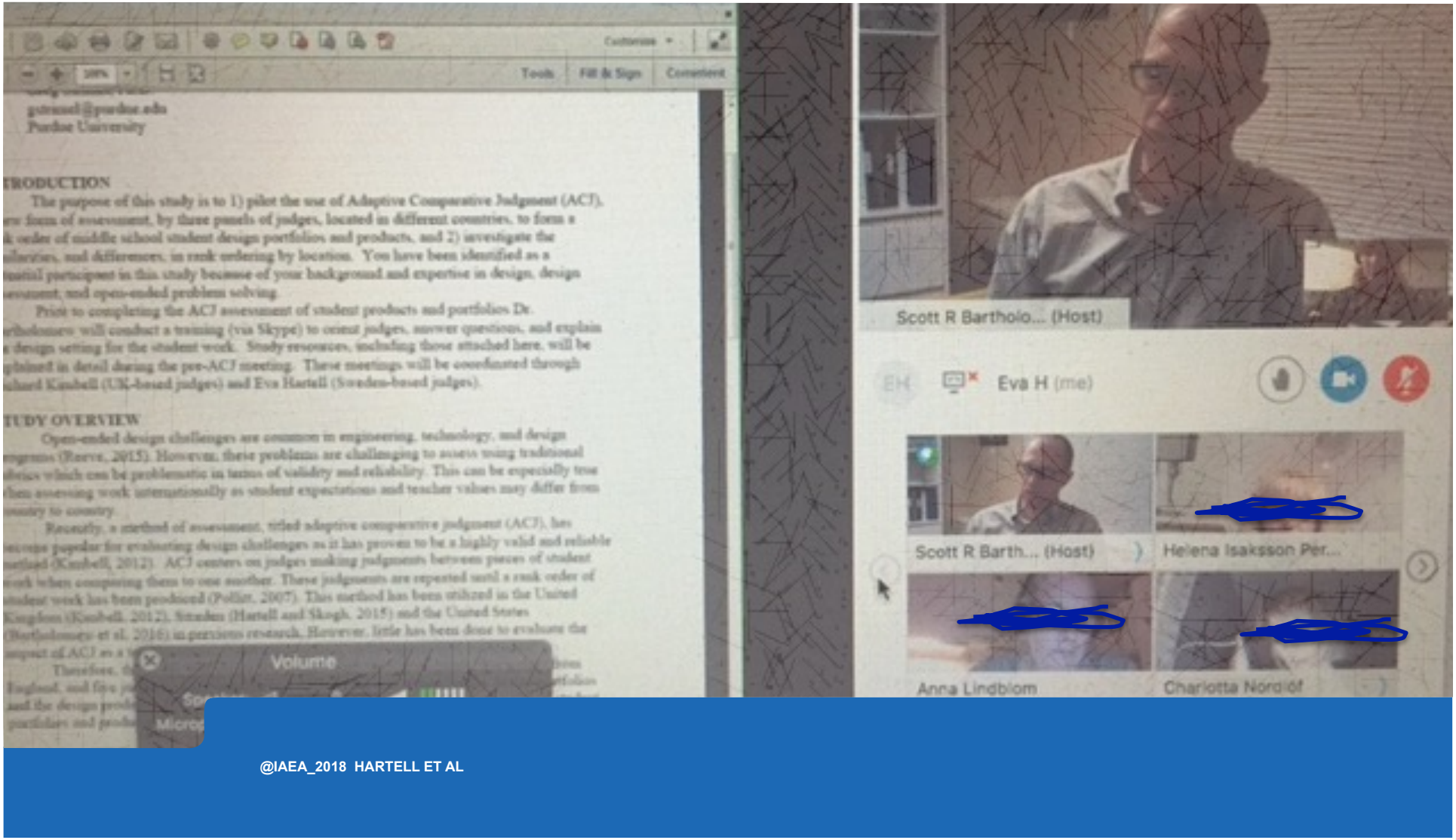
Judge feedback on items was collected and coded to identify themes



Assessors: Locations & Backgrounds



Training Session with Judges

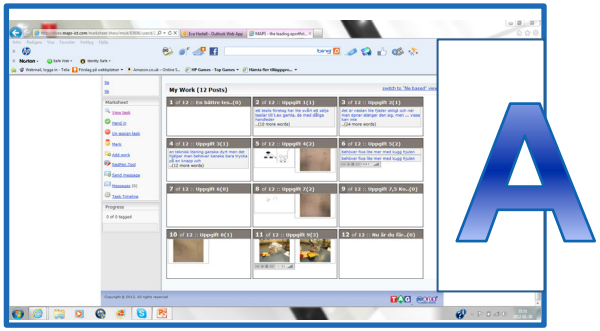


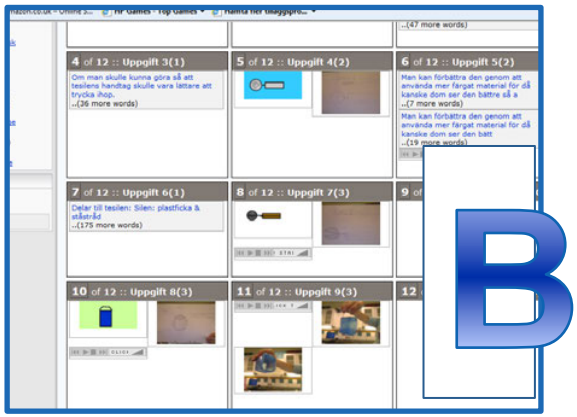
Assessment Method

Adaptive Comparative Judgment:

Relies on pairwise comparisons of work to generate a rank order of all items

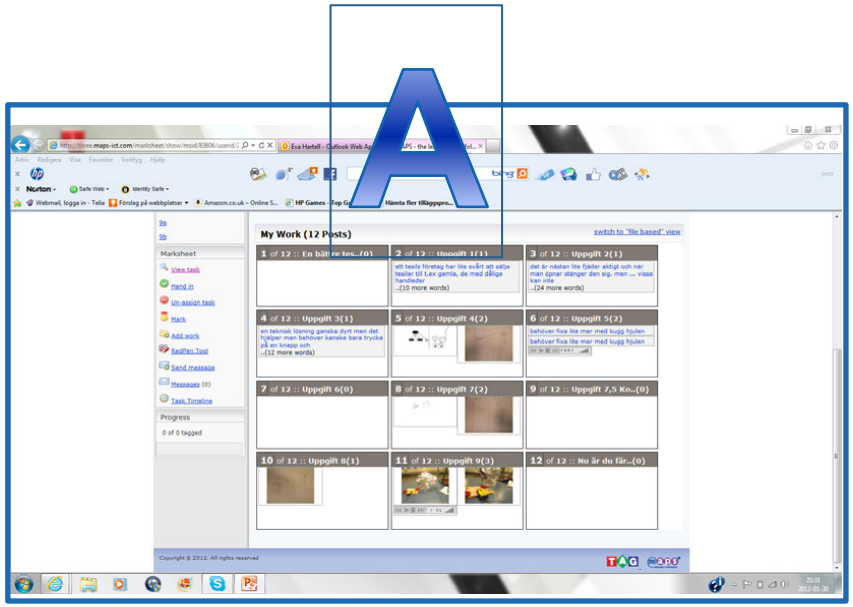
What emerges is a **collective professional consensus** from the group of judges.

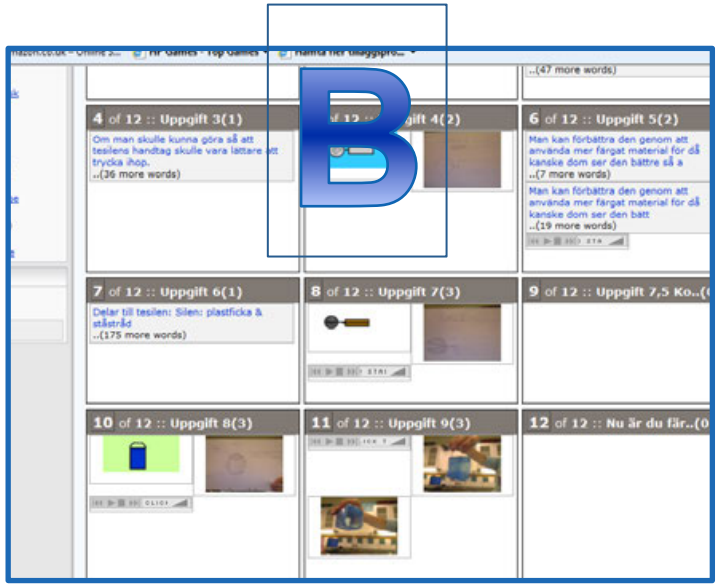




C.f. e.g. Pollitt (2012), Kimbell (2013), Hartell & Skogh, 2015, Bartholomew et al (2017), Canty et al (2017), Lesterhuis, 2017

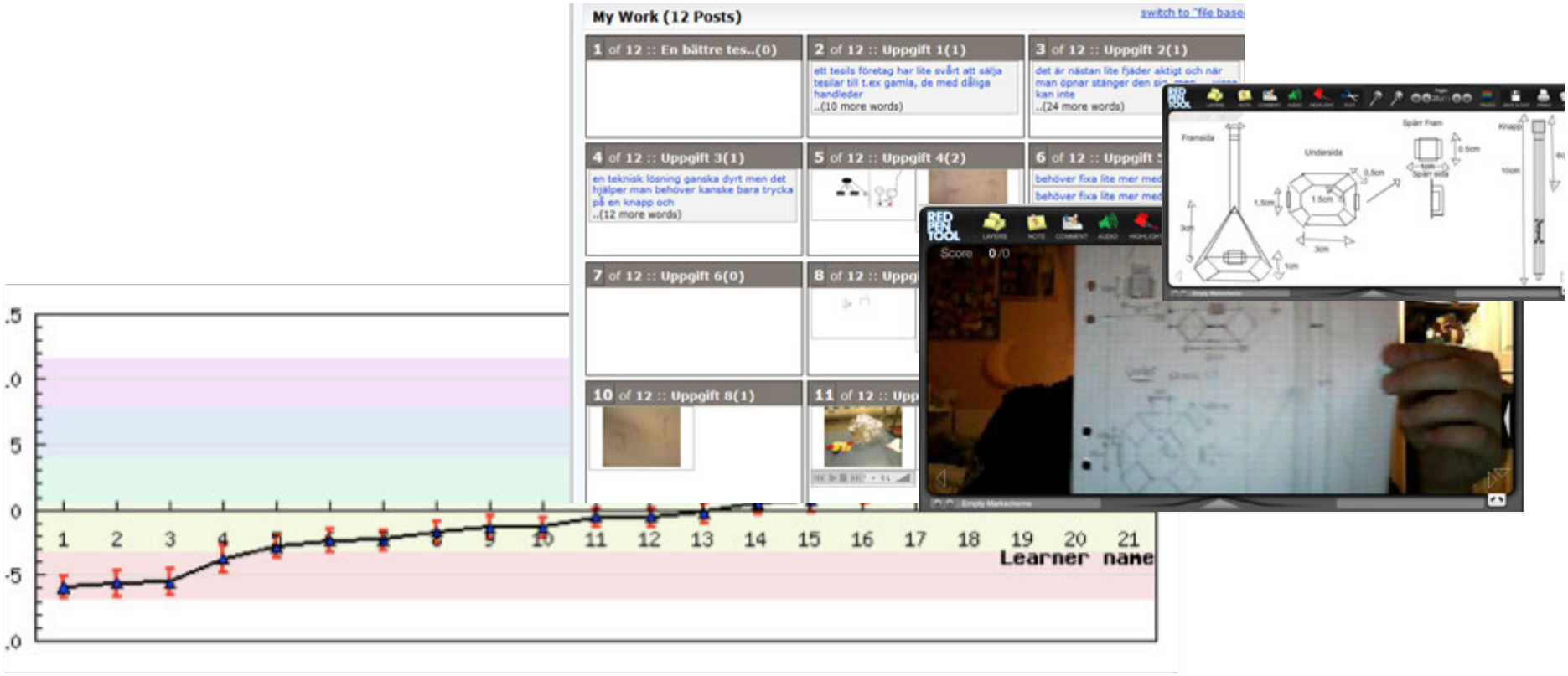
Which One is Better?





Rank Order? Not Really!

Instead, a collective professional consensus from the group of judges (teachers, students, etc.)



Why Did the Assessors Choose They Way They Did?

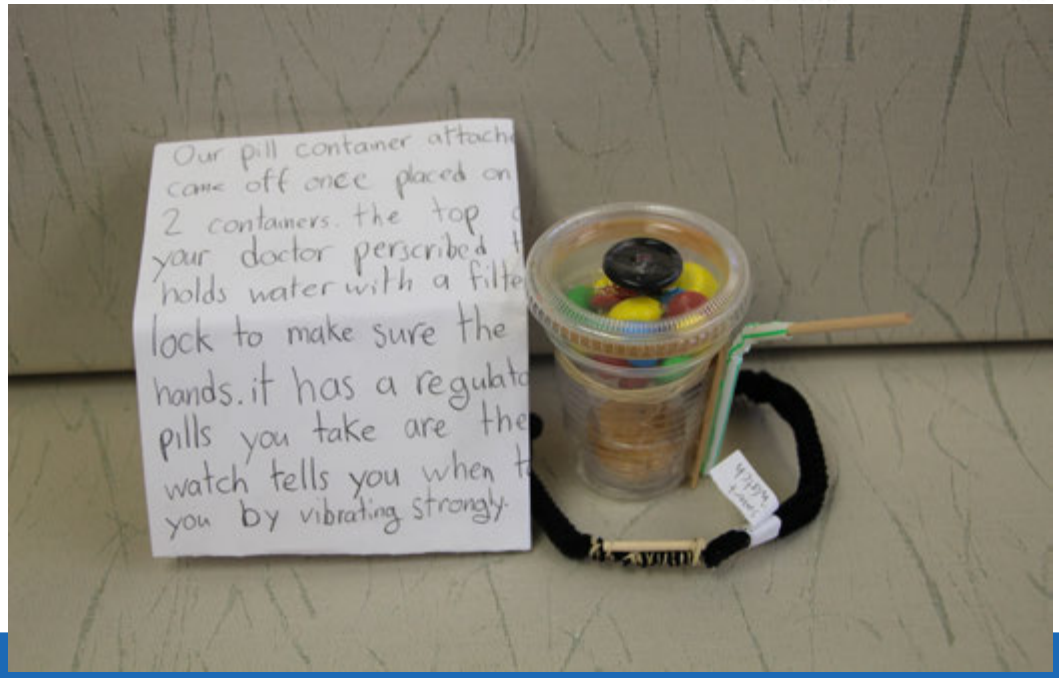
Qualitative Analysis of Similarities & Differences

(Product Comments)

Looks **easier** to use	An example of wacky vs practical? I chose A because it looked **more exciting**, is that a spiral dispensing system or just a random pipe cleaner? *This is so frustrating.*	a is a simple idea and has a **compact** design/ shape. Good to have in purse/ bag
Easier to use	B won as it is a bit different and there seems to be some **thought** given to it.	a is **compact** in its design, it is good to have if you want to bring it along in a bag
Clearer, **easier** to use	looks like a **richer journey**	**stylish** design, aesthetically thorough
Seems more **functional**	B - slightly **more developed**, but equally chaotic!	easy to **pack**
	2 very similar 'looking' photos. choice = well made or **wacky**. I went with wacky???	a is **simple** and smooth

Top Ranked Prototype – United Kingdom

Comment	Code
chose B as the pots were arranged differently! Is it a better pill holder - no idea???	innovation
Marginal - A more complete concept	developed
wins because it is a more developed solution	developed
no idea what [is] going on but different	innovation
potentially more user friendly	usability



Top Ranked Prototype – Sweden

Swedish Comment	Translation	Code
a verkar enkel och funktionell.	a seems simple and functional	usability
A is smaler	A is smaller	size
B har en spännande formgivning	B has an exciting design	design
A är en enkel idé och har en kompakt form. Bra i väskan.	A is a simple idea and has a compact design/shape. Good to have in purse/bag	design; size

Top Ranked Prototype – United States

Comment	Code
Looks easier to use	usability
More compact. user-friendly	size; usability
love the idea looks like it can hold all the days etc.	design

Findings– Themes by Country

Country	Prototype	Portfolio
United Kingdom	1. Innovation 2. Developed 3. Usability	1. Developed 2. Innovation 3. Follow Through
Sweden	1. Usability 2. Size 3. Design	1. Communication 2. Design Process 3. Complete
United States	1. Usability 2. Size 3. Design	1. Criteria 2. Complete 3. Reflection

Conclusion

Adaptive Comparative Judgment can:

- act as a assessment tool to discover design values
- be useful for international comparisons in open ended design scenarios (Task design is very important)
- Serve as a catalyst for discussion
- Serve as a useful tool to
  - unpack teachers' assessment practices and uncover design values
  - dig deeper than documents
  - explicate criteria for success

Conclusion

We see possible many use of ACJ on larger scale to find out and explicate criteria for success in open-ended design tasks to inform formative assessment practices.

See:

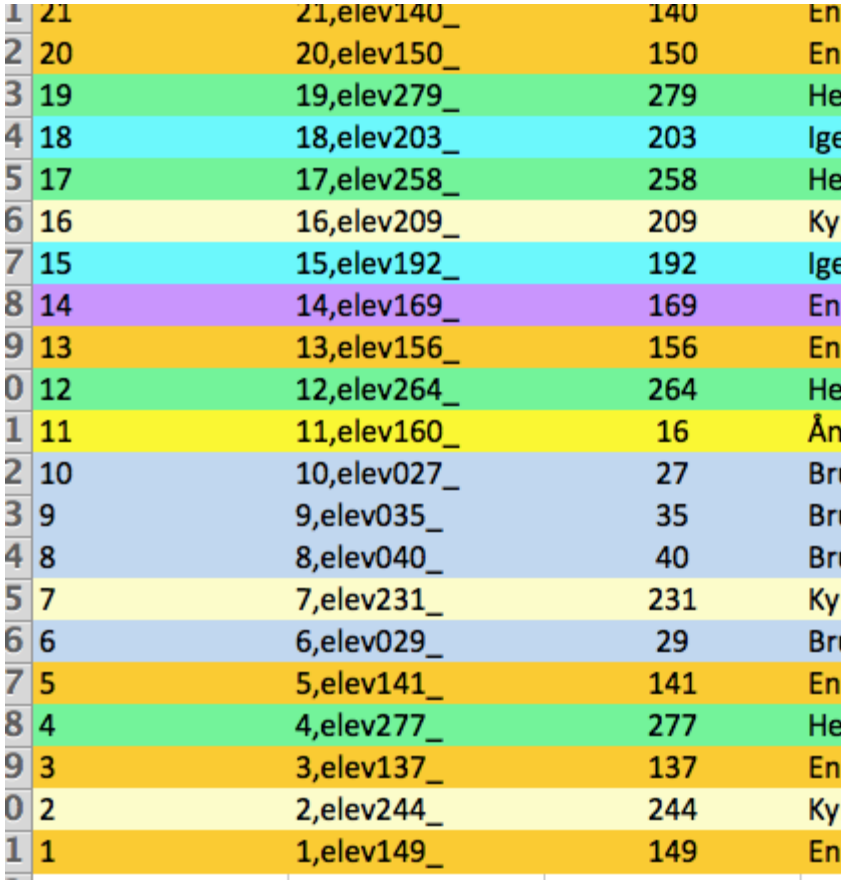
- Bartholomew, Hartell & Strimel (2017)
- Hartell, Strimel & Bartholomew (2017)
- Bartholomew, Yoshikawa, Hartell & Strimel (2018)

The Potential of Comparative Judgment in Open-Ended Tasks

- Data is collected during “ordinary” lesson activities
- Students collect evidence of learning (validity & teachers work load)
- Decision driven data collection instead of data driven decision making since tasks design
- Reliable results
- Judge consistency
- Inviting other professionals to your classroom and you get to visit theirs “without too much trouble” (cloud-based)

- *The power of the collective & the profession*

Many Applications of Comparative Judgment



1	21	21,elev140_	140	En
2	20	20,elev150_	150	En
3	19	19,elev279_	279	He
4	18	18,elev203_	203	Ige
5	17	17,elev258_	258	He
6	16	16,elev209_	209	Ky
7	15	15,elev192_	192	Ige
8	14	14,elev169_	169	En
9	13	13,elev156_	156	En
0	12	12,elev264_	264	He
1	11	11,elev160_	16	Ån
2	10	10,elev027_	27	Br
3	9	9,elev035_	35	Br
4	8	8,elev040_	40	Br
5	7	7,elev231_	231	Ky
6	6	6,elev029_	29	Br
7	5	5,elev141_	141	En
8	4	4,elev277_	277	He
9	3	3,elev137_	137	En
0	2	2,elev244_	244	Ky
1	1	1,elev149_	149	En

- Both summative and formative assessment purposes
- Track progress
- Peer and self-assessment
- Teacher training!
- Connoisseurship
- Building assessment literacy and self-efficacy
- Moderation (yourself / peer)
- ~~Ranking schools~~

- **Research method**

Future Work…..

1. Potential for ACJ as an International Collaboration Tool
2. Usefulness/ appropriateness for different purposes of assessment
3. Outliers
4. Expanding International Partnerships
5. Investigating designs made by students in other regions
6. International students do the same task or different?
7. Moderation
8. Tool for building assessment literacy and self-efficacy?
9. Connoisseur of STEM?

Thank you! To be continued

Dr. Eva Hartell
ehartell@kth.se

Dr. Greg Strimel
gstrimel@purdue.edu

Dr. Scott Bartholomew
sbartho@purdue.edu

Emily Yoshikawa
emilyyoshikawa@gmail.com

@EvaHartell
@strimelg
@IAEA_2018

References

- Bartholomew, S., Hartell, E., & Strimel, G. (2017). ACJ: *A Tool for International Assessment Collaboration*. In PATT34 Technology & Engineering Education – Fostering the Creativity of Youth Around The Globe, Millersville University, Pennsylvania, USA 10–14 July, 2017.
- Bartholomew, S. R., Strimel, G. S., & Yoshikawa, E. (2018). *Using Adaptive Comparative Judgment for Student Formative Feedback and Learning During a Middle School Open-ended Design Challenge.* International Journal of Technology & Design Education. https://doi.org/10.1007/s10798-018-9442-7
- Canty, D., Seery, N., Hartell, E. & Doyle, E. (2017), "*Integrating Peer Assessment in Technology Education through Adaptive Comparative Judgment,*" In PATT34 Technology & Engineering Education – Fostering the Creativity of Youth Around The Globe, Millersville University, Pennsylvania, USA. 10–14 July, 2017.
- Hartell, E., and Skogh, I. B. (2015). *Criteria for Success: A Study of Primary Technology Teachers' Assessment of Digital Portfolios*. Australasian Journal of Technology Education, 2(1) http://ajte.org/index.php/AJTE/article/view/27
- Pollitt, A. (2012). *The method of adaptive comparative judgment. Assessment in Education: Principles, Policy & Practice*, 19(3), s. 281-300.
- Seery, N., Canty, D., & Phelan, P. (2012). *The validity and value of peer assessment using adaptive comparative judgement in design driven practical education.* International Journal of Technology and Design Education, 22(2), 205-226.
- Seery, N., & Canty, D. (2018). *Assessment and Learning: The Proximal and Distal Effects of Comparative Judgment.* In, Springer International Handbook of Technology Education, s. 735–748.
- Statens skolverk (2017) *Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011* (reviderad 2017).
- van Daal, T., Lesterhuis, M., Coertjens, L., Donche, V., & De Maeyer, S. (2016). *Validity of comparative judgement to assess academic writing: examining implications of its holistic character and building on a shared consensus*. Assessment in Education: Principles, Policy & Practice, s. 1-16.