1946–2021 75 YEARS NFER National Foundation for Educational Research

Virtual Standard Setting: The benefits, the challenges, and the way forward.

Charalambos (Harry) Kollias, Ph.D.

EALTA Webinar, May 18th 2022

standard setting

logistics

consequences

cut score studies …

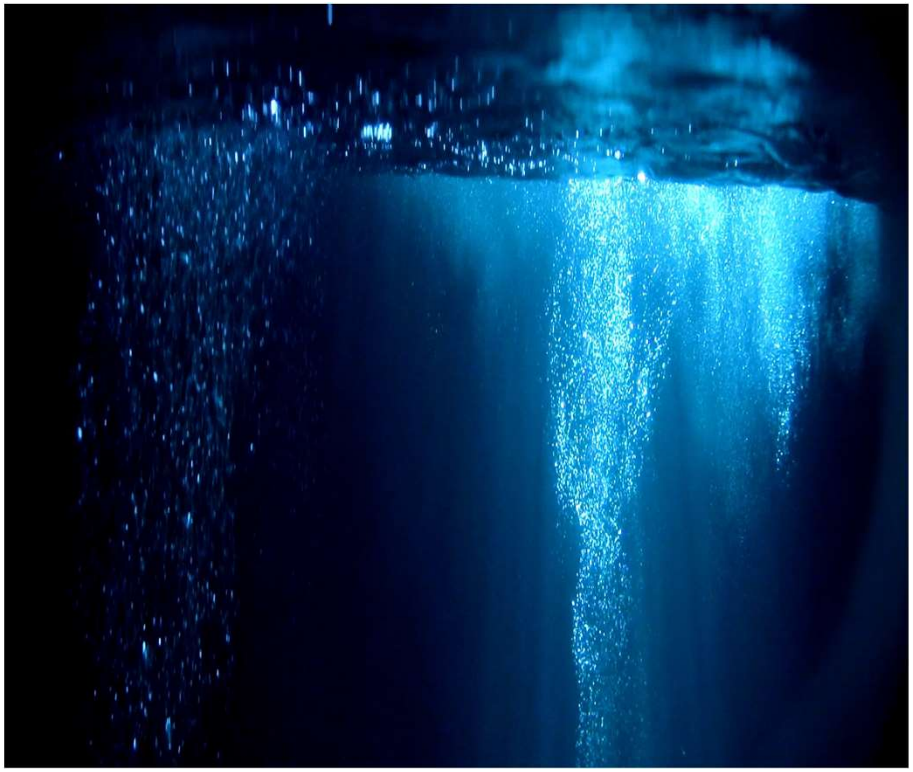
not being conducted

(Tannenbaum, 2013)

not being replicated

(Dunlea & Figueras, 2012)

virtual standard setting (VSS)



literature review (VSS) 1999 – 2014

results comparable with *F2F* (Katz & Tannenbaum; 2014)

few empirical studies

feasible in **<u>series of smaller sessions</u>** …

… through different media:

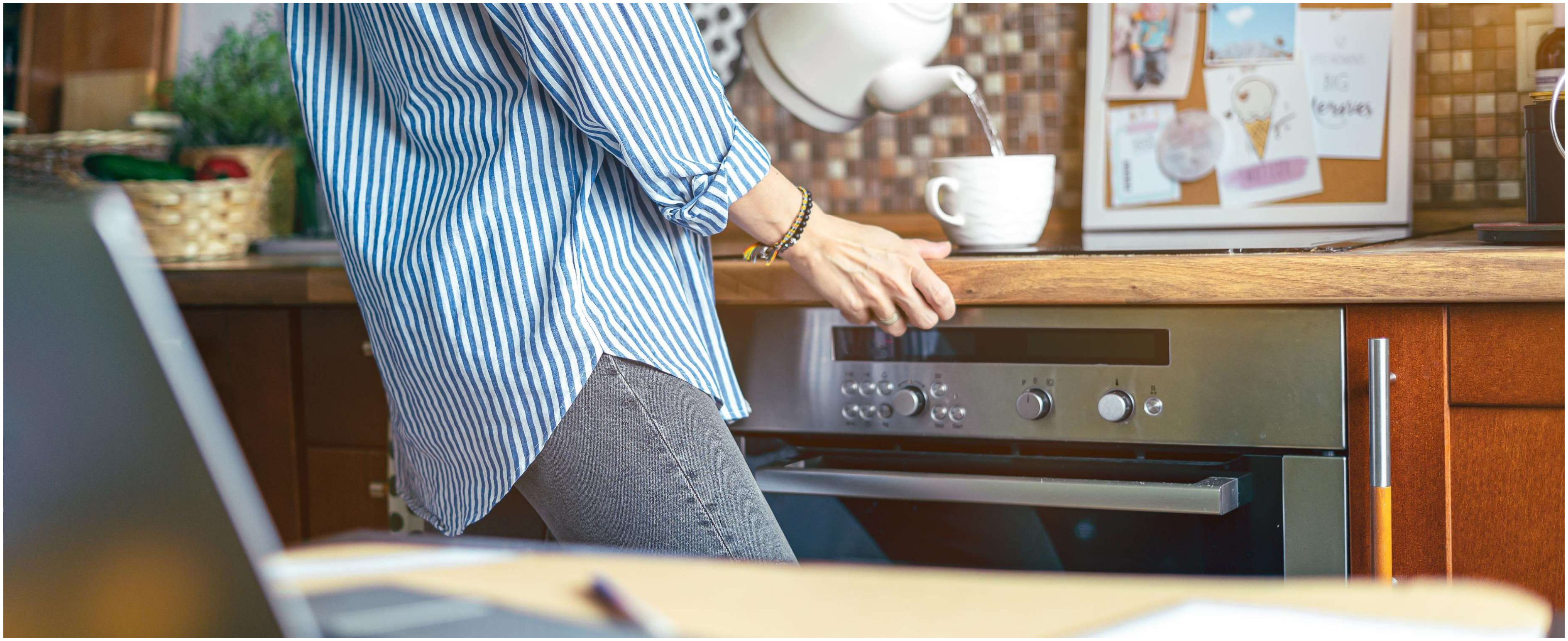
- emails (Harvey & Way, 1999);
- audio-conferencing (Katz & Tannenbaum; 2014);
- combination of audio & video (Katz, Tannenbaum, & Kannan, 2009)].

… in asynchronous environments

… in combined asynchronous & synchronous environments

virtual benefits

virtual benefits cont.

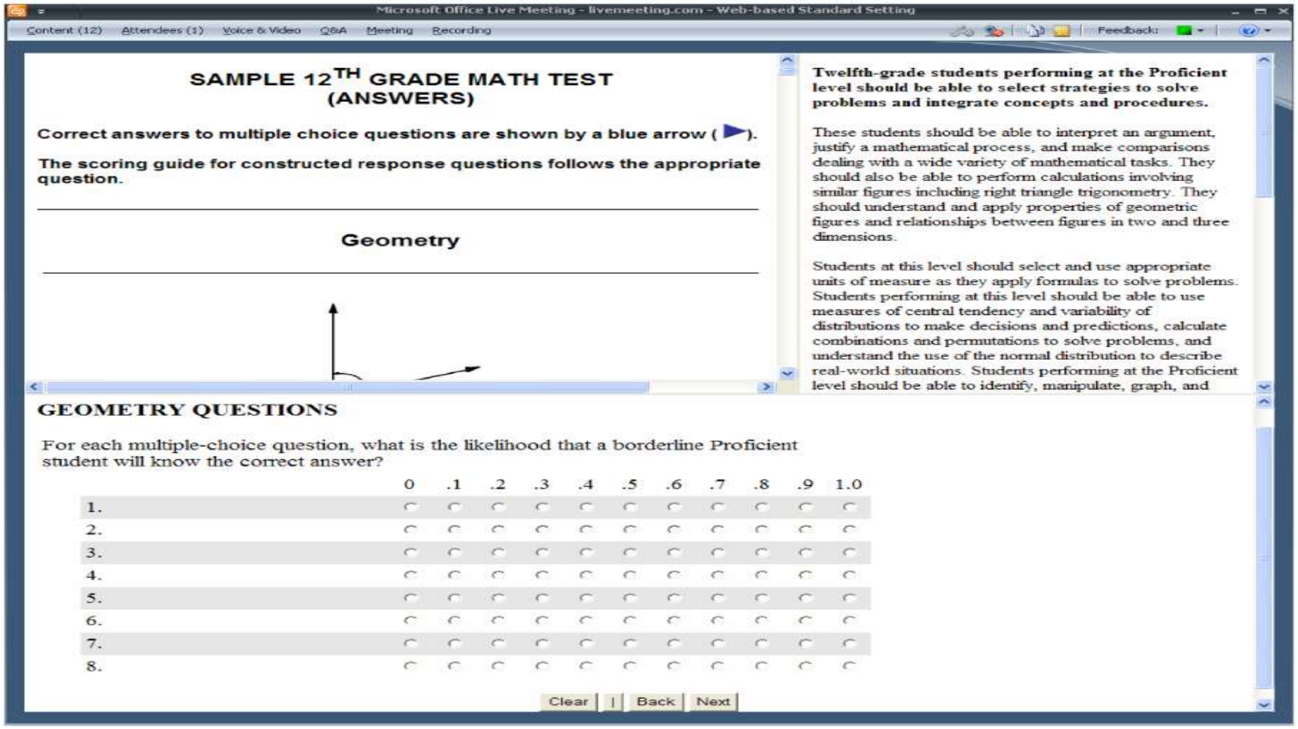


yesterday (f2f)

today (virtual)

choice of method

the Modified Angoff method: ETS platform (2009)



SAMPLE 12TH GRADE MATH TEST (ANSWERS)

Correct answers to multiple choice questions are shown by a blue arrow (▶).

The scoring guide for constructed response questions follows the appropriate question.

Geometry

Twelfth-grade students performing at the Proficient level should be able to select strategies to solve problems and integrate concepts and procedures.

These students should be able to interpret an argument, justify a mathematical process, and make comparisons dealing with a wide variety of mathematical tasks. They should also be able to perform calculations involving similar figures including right triangle trigonometry. They should understand and apply properties of geometric figures and relationships between figures in two and three dimensions.

Students at this level should select and use appropriate units of measure as they apply formulas to solve problems. Students performing at this level should be able to use measures of central tendency and variability of distributions to make decisions and predictions, calculate combinations and permutations to solve problems, and understand the use of the normal distribution to describe real-world situations. Students performing at the Proficient level should be able to identify, manipulate, graph, and

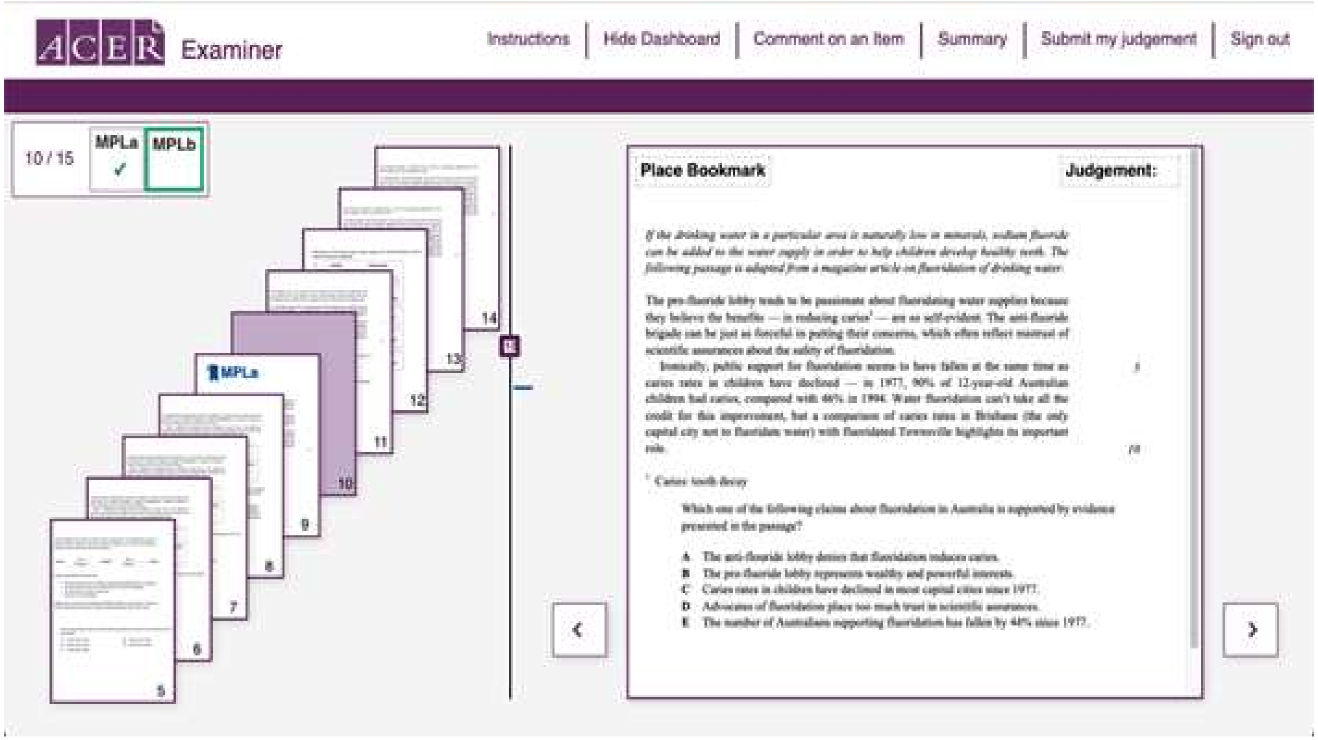
GEOMETRY QUESTIONS

For each multiple-choice question, what is the likelihood that a borderline Proficient student will know the correct answer?

	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
1.	○	○	○	○	○	○	○	○	○	○	○
2.	○	○	○	○	○	○	○	○	○	○	○
3.	○	○	○	○	○	○	○	○	○	○	○
4.	○	○	○	○	○	○	○	○	○	○	○
5.	○	○	○	○	○	○	○	○	○	○	○
6.	○	○	○	○	○	○	○	○	○	○	○
7.	○	○	○	○	○	○	○	○	○	○	○
8.	○	○	○	○	○	○	○	○	○	○	○

Source: Katz, Tannenbaum & Kannan (2009)

the Bookmark method: ACER platform (2022)



If the drinking water in a particular area is naturally low in minerals, sodium fluoride can be added to the water supply in order to help children develop healthy teeth. The following passage is adapted from a magazine article on fluoridation of drinking water.

The pro-fluoride lobby tends to be passionate about fluoridating water supplies because they believe the benefits — in reducing caries[1] — are so self-evident. The anti-fluoride brigade can be just as forceful in putting their concerns, which often reflect mistrust of scientific assurances about the safety of fluoridation.

Ironically, public support for fluoridation seems to have fallen at the same time as caries rates in children have declined — in 1977, 90% of 12-year-old Australian children had caries, compared with 46% in 1994. Water fluoridation can't take all the credit for this improvement, but a comparison of caries rates in Brisbane (the only capital city not to fluoridate water) with fluoridated Townsville highlights its important role.

[1] Caries: tooth decay

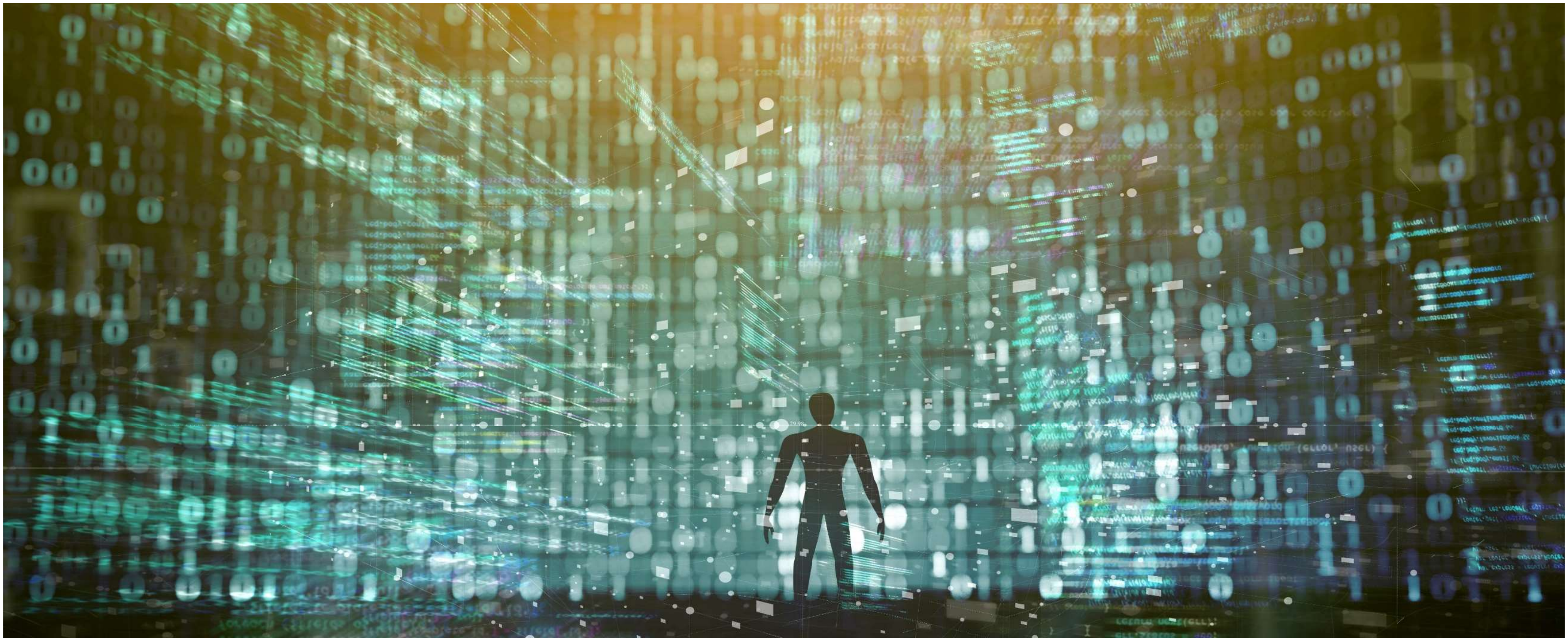
Which one of the following claims about fluoridation in Australia is supported by evidence presented in the passage?

A The anti-fluoride lobby denies that fluoridation reduces caries.
B The pro-fluoride lobby represents wealthy and powerful interests.
C Caries rates in children have declined in most capital cities since 1977.
D Advocates of fluoridation place too much trust in scientific assurances.
E The number of Australians supporting fluoridation has fallen by 44% since 1977.

Source: https://www.acer.org/gb/discover/article/innovation-in-assessment-standard-setting

Ph.D. thesis: Lancaster University, 2017

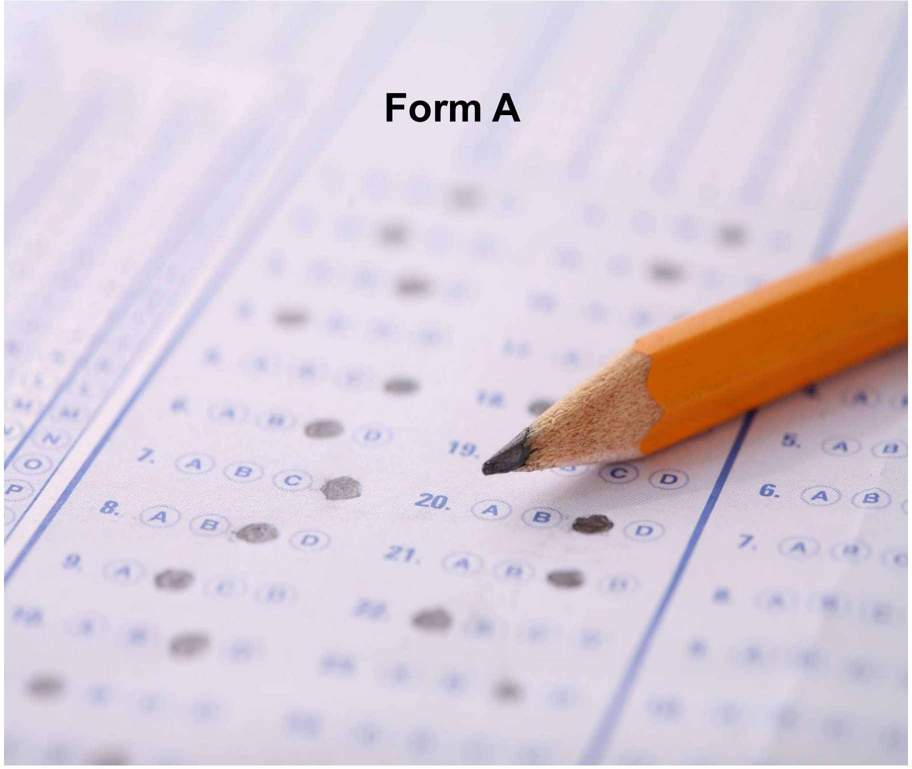
virtual environment



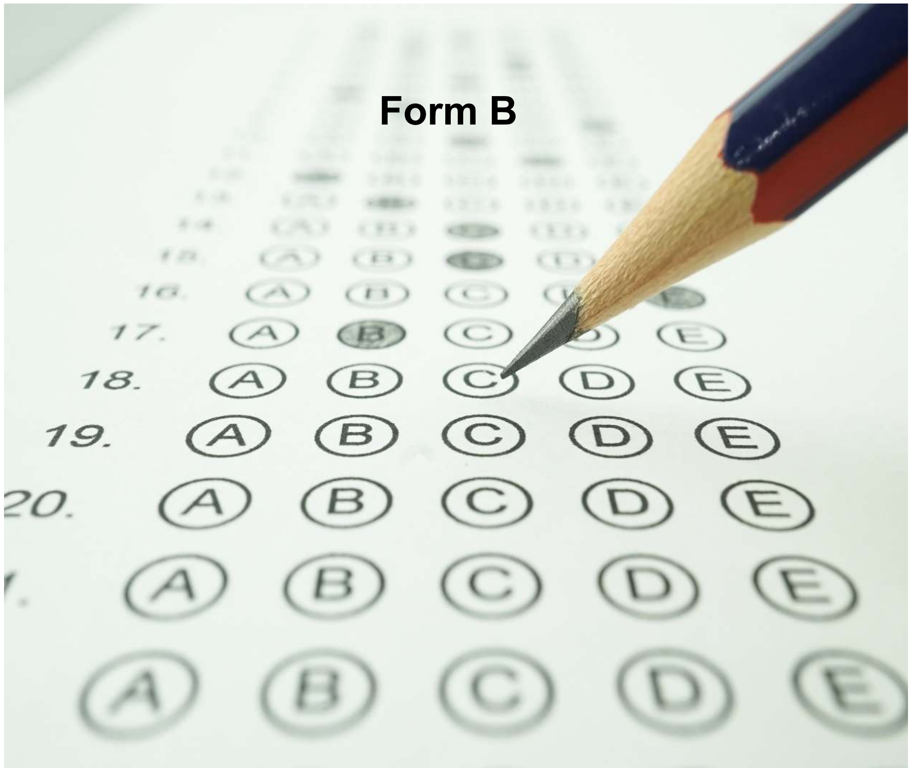
audio medium vs. video medium

45 judges

2 equated CEFR B1 tests



Form A

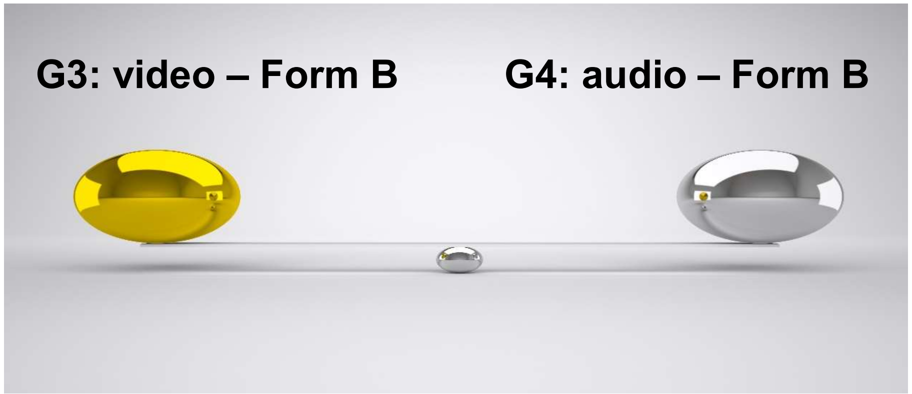


Form B

session 1

G1: audio – Form A

G2: video – Form A



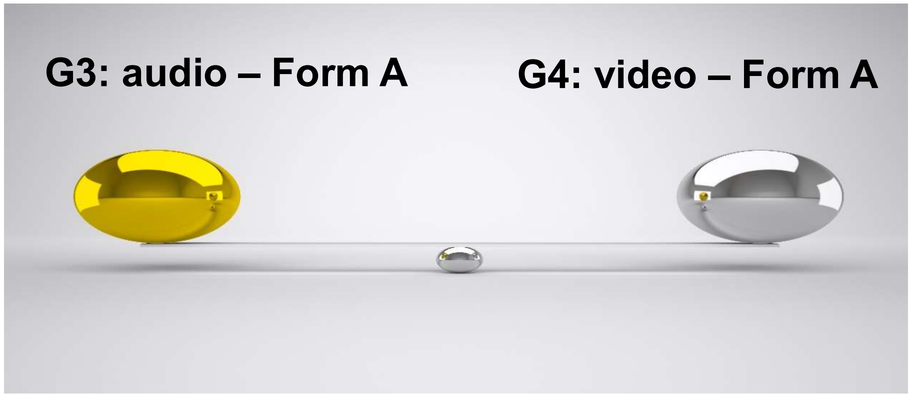
G3: video – Form B

G4: audio – Form B

session 2

G1: video – Form B

G2: audio – Form B



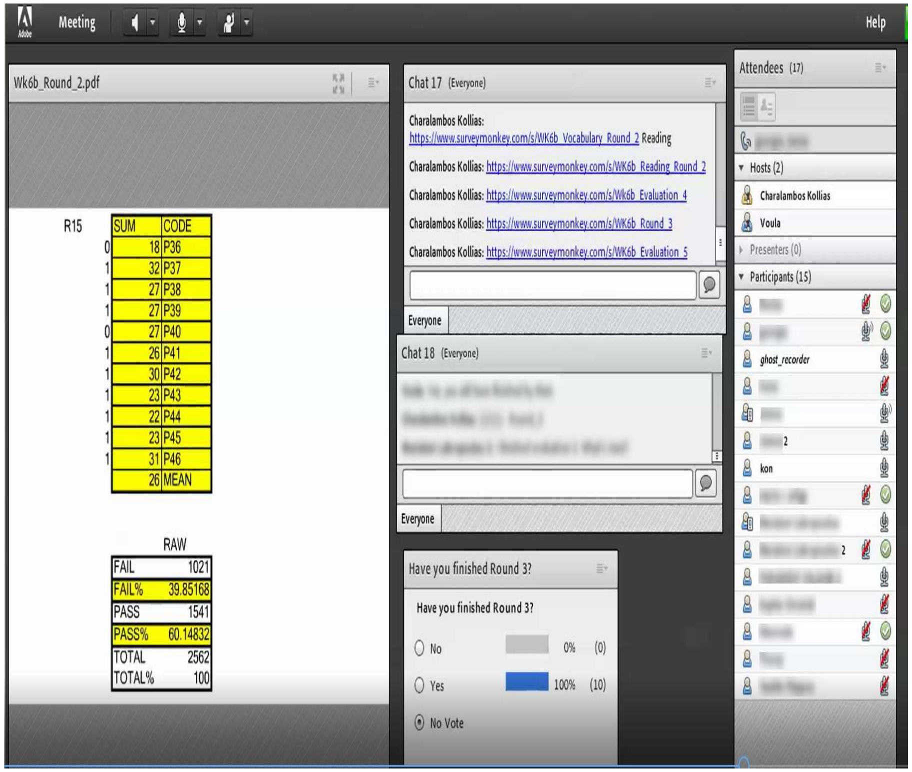
G3: audio – Form A

G4: video – Form A

TRAINING

□ yes □ no
□ yes □ no
□ yes □ no
□ yes □ no
□ yes □ no
□ yes □ no
□ yes □ no
□ yes □ no
□ yes □ no

the research platforms (1)



Wk6b_Round_2.pdf

R15	SUM	CODE
0	18	P36
1	32	P37
1	27	P38
1	27	P39
0	27	P40
1	26	P41
1	30	P42
1	23	P43
1	22	P44
1	23	P45
1	31	P46
	26	MEAN

RAW

FAIL	1021
FAIL%	39.85168
PASS	1541
PASS%	60.14832
TOTAL	2562
TOTAL%	100

Chat 17 (Everyone)

Charalambos Kollias: https://www.surveymonkey.com/s/WK6b_Vocabulary_Round_2 Reading

Charalambos Kollias: https://www.surveymonkey.com/s/WK6b_Reading_Round_2

Charalambos Kollias: https://www.surveymonkey.com/s/Wk6b_Evaluation_4

Charalambos Kollias: https://www.surveymonkey.com/s/WK6b_Round_3

Charalambos Kollias: https://www.surveymonkey.com/s/WK6b_Evaluation_5

Chat 18 (Everyone)

[redacted]

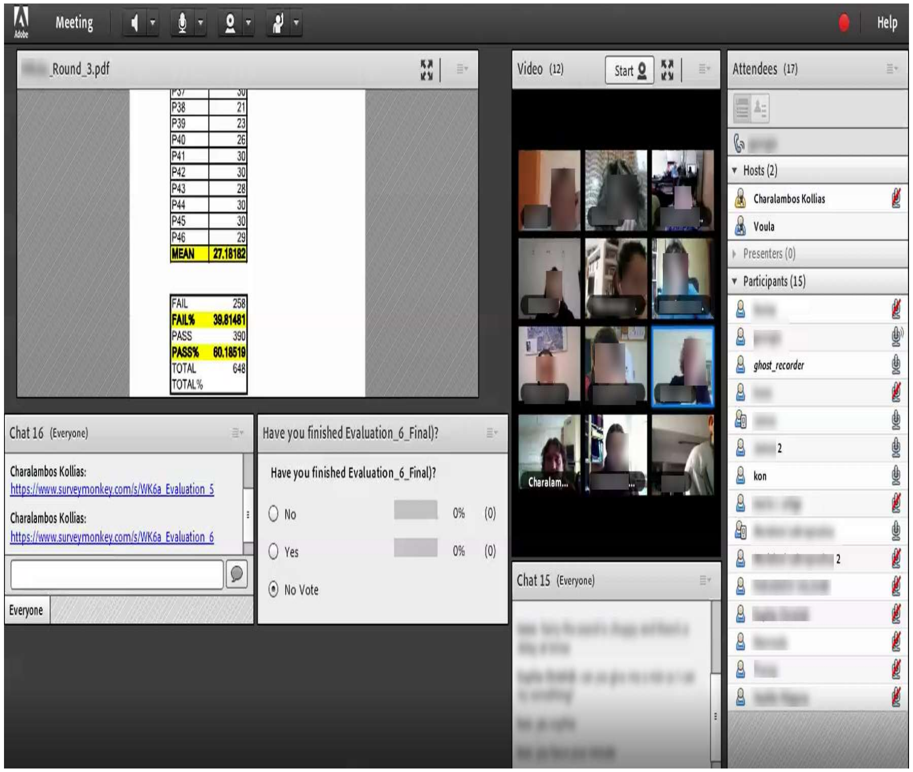
Have you finished Round 3?

Have you finished Round 3?

○ No 0% (0)

○ Yes 100% (10)

● No Vote



[redacted]_Round_3.pdf

P37	30
P38	21
P39	23
P40	26
P41	30
P42	30
P43	28
P44	30
P45	30
P46	29
MEAN	27.18182

FAIL	258
FAIL%	39.81481
PASS	390
PASS%	60.18519
TOTAL	648
TOTAL%	

Chat 16 (Everyone)

Charalambos Kollias: https://www.surveymonkey.com/s/WK6a_Evaluation_5

Charalambos Kollias: https://www.surveymonkey.com/s/WK6a_Evaluation_6

Have you finished Evaluation_6_Final)?

Have you finished Evaluation_6_Final)?

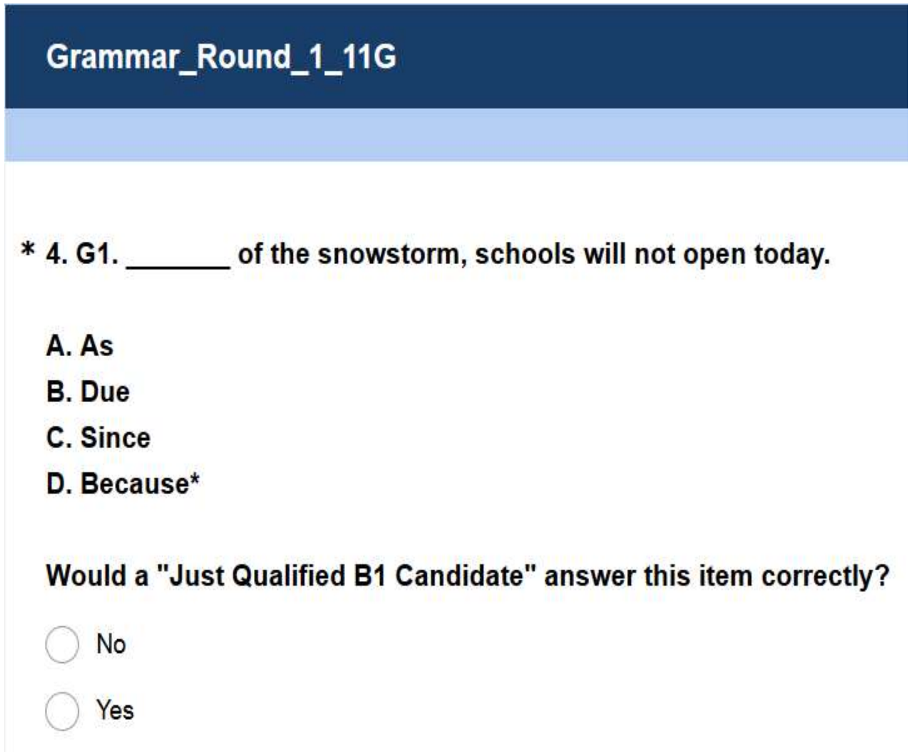
○ No 0% (0)

○ Yes 0% (0)

● No Vote

[redacted]

the research platforms (2)



Grammar_Round_1_11G

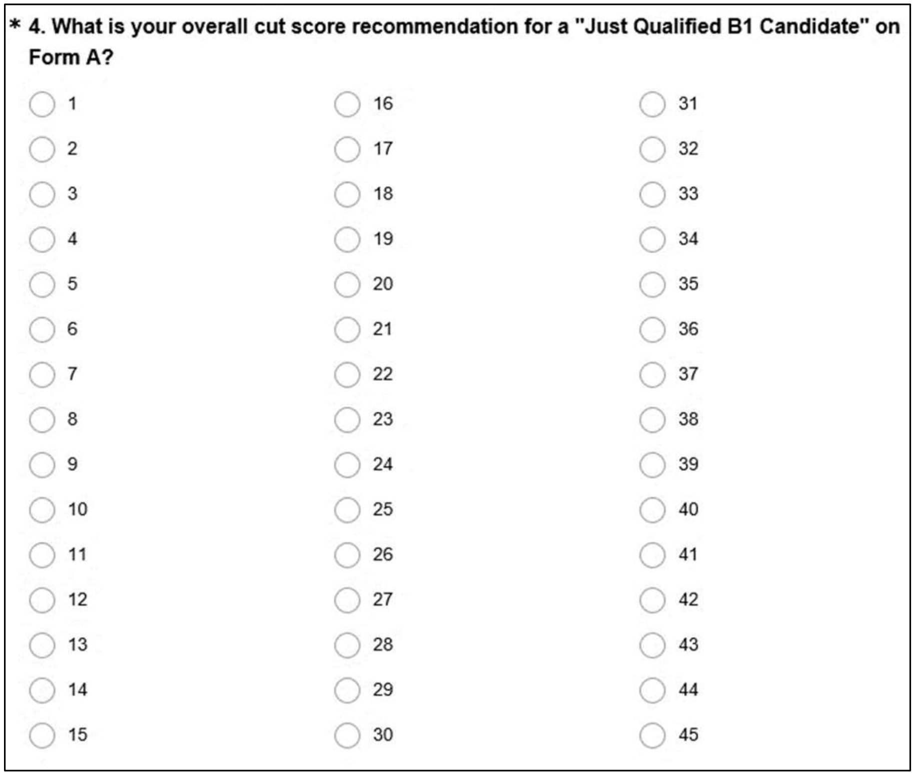
* 4. G1. _______ of the snowstorm, schools will not open today.

A. As
B. Due
C. Since
D. Because*

Would a "Just Qualified B1 Candidate" answer this item correctly?

○ No

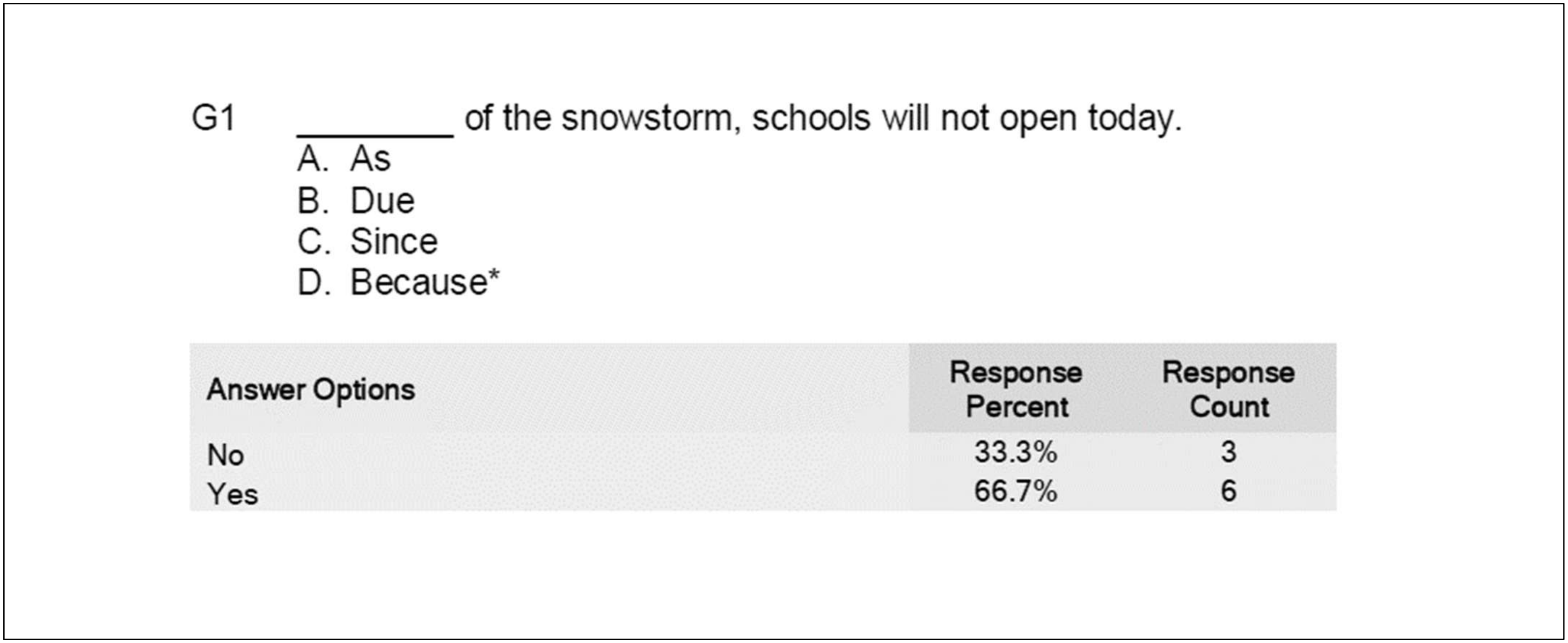
○ Yes



* 4. What is your overall cut score recommendation for a "Just Qualified B1 Candidate" on Form A?

○ 1
○ 2
○ 3
○ 4
○ 5
○ 6
○ 7
○ 8
○ 9
○ 10
○ 11
○ 12
○ 13
○ 14
○ 15
○ 16
○ 17
○ 18
○ 19
○ 20
○ 21
○ 22
○ 23
○ 24
○ 25
○ 26
○ 27
○ 28
○ 29
○ 30
○ 31
○ 32
○ 33
○ 34
○ 35
○ 36
○ 37
○ 38
○ 39
○ 40
○ 41
○ 42
○ 43
○ 44
○ 45

normative information



G1 ________ of the snowstorm, schools will not open today.

A. As
B. Due
C. Since
D. Because*

Answer Options	Response Percent	Response Count
No	33.3%	3
Yes	66.7%	6

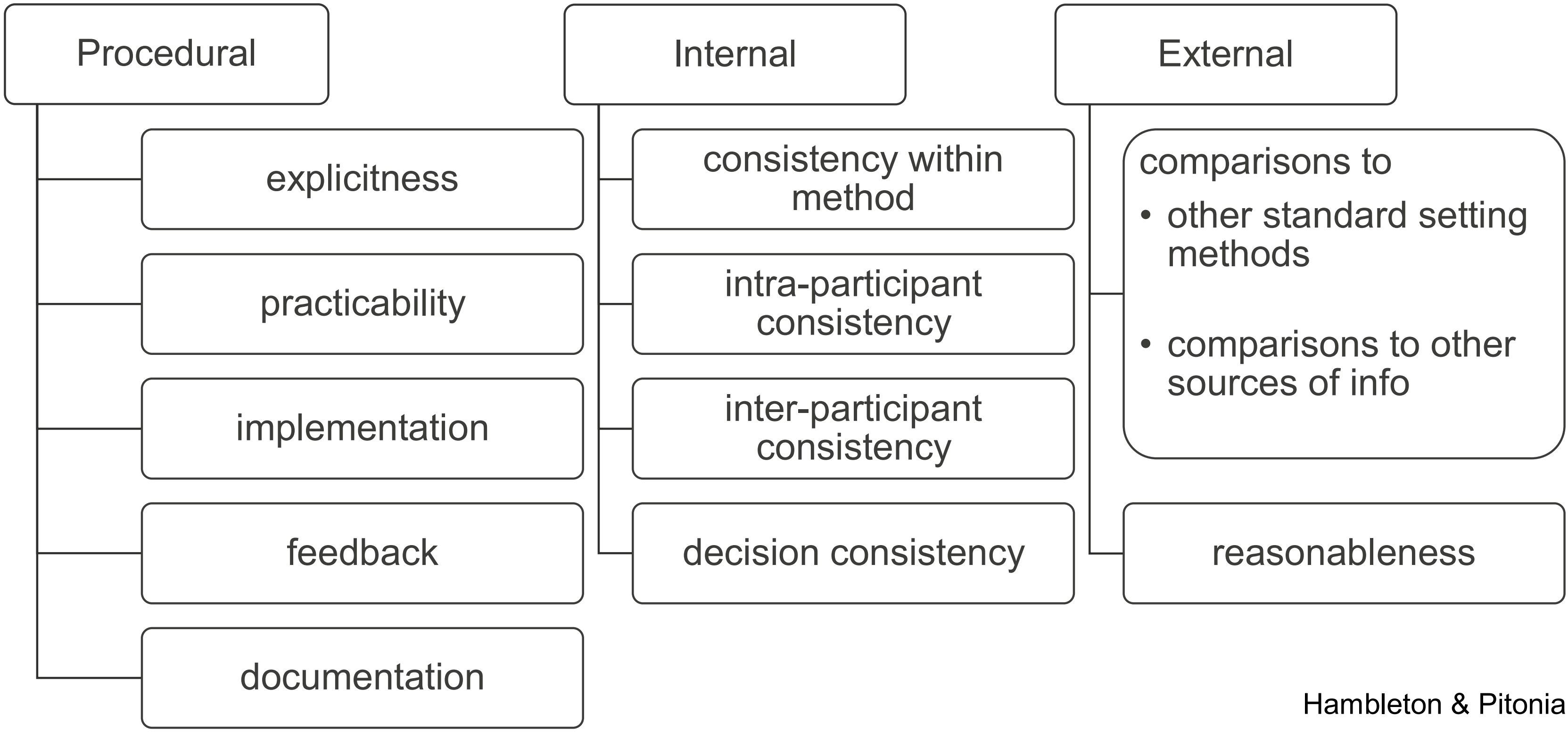
data collection

quantitative

qualitative

EVALUATION

standard setting evaluation elements



Procedural
- explicitness
- practicability
- implementation
- feedback
- documentation

Internal
- consistency within method
- intra-participant consistency
- inter-participant consistency
- decision consistency

External
- comparisons to
  - other standard setting methods
  - comparisons to other sources of info
- reasonableness

Hambleton & Pitoniak, 2006

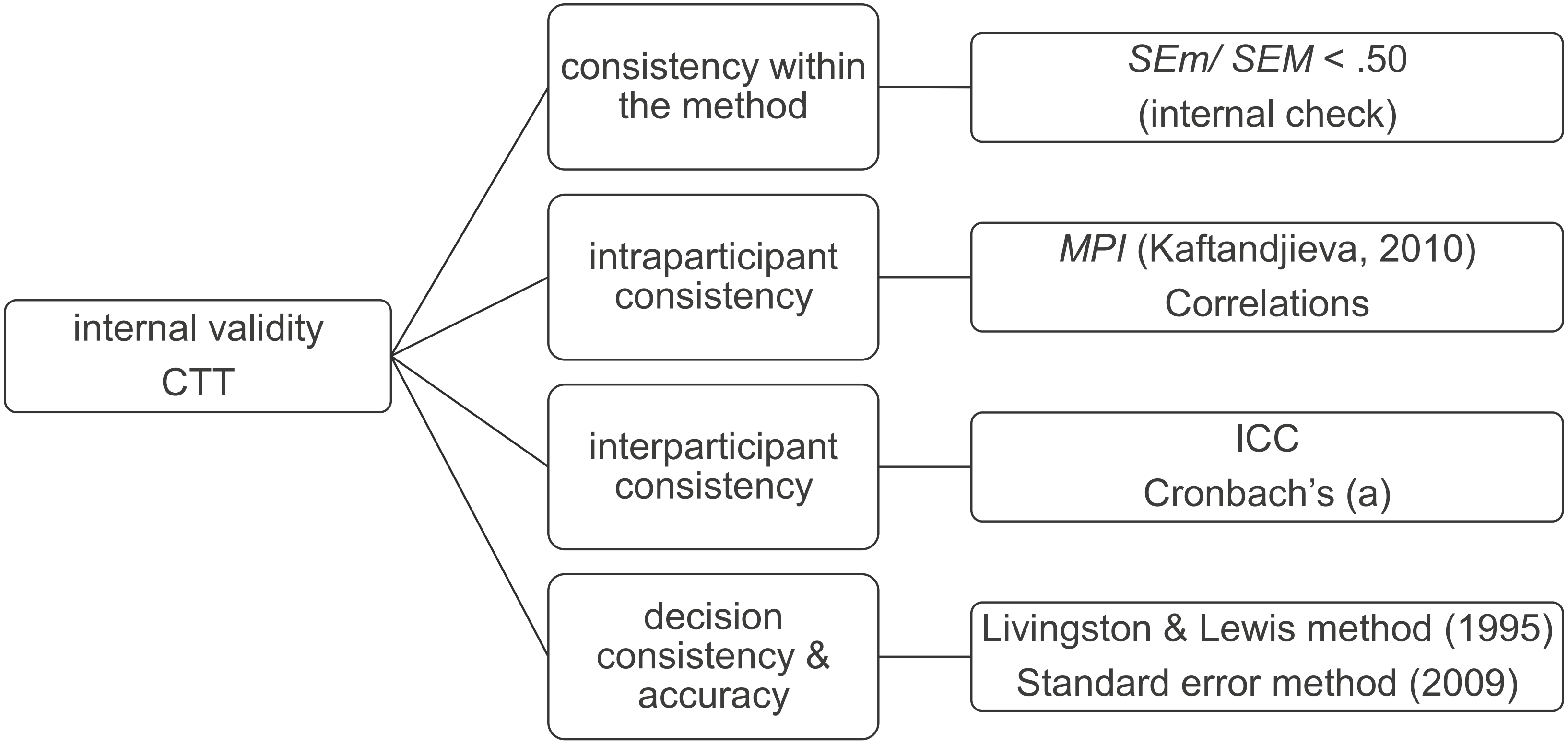
DATA ANALYSIS

quantitative

analysis:

- classical test theory (CTT)
- Rasch measurement theory (RMT)

internal validity: CTT


internal validity
CTT
consistency within the method
intraparticipant consistency
interparticipant consistency
decision consistency & accuracy
SEm/ SEM < .50
(internal check)
MPI (Kaftandjieva, 2010)
Correlations
ICC
Cronbach's (a)
Livingston & Lewis method (1995)
Standard error method (2009)

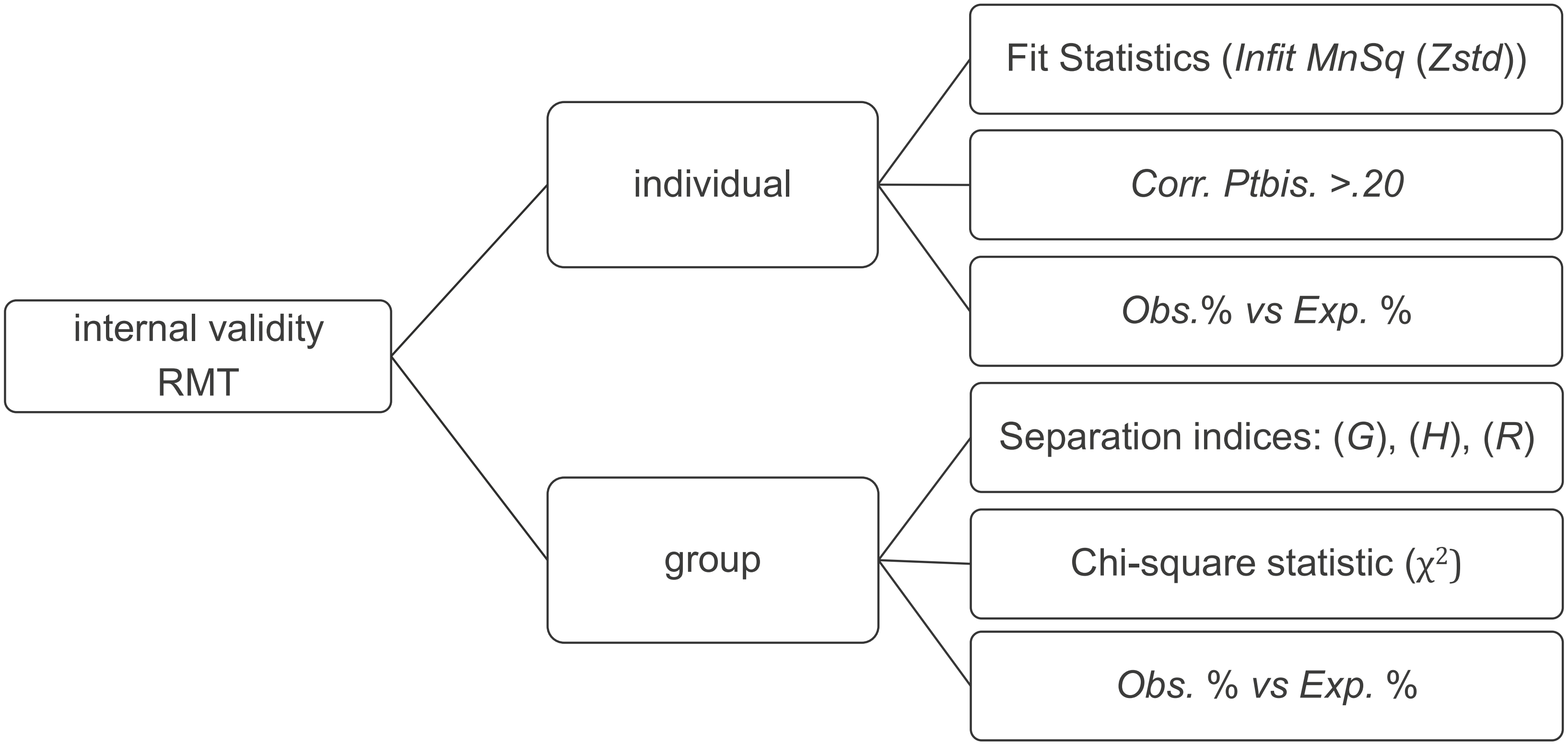
resources CEFR

- **no guidance on Rasch and/or IRT procedures**
- **no framework for evaluating**
  - cut scores set through Rasch and/or IRT
  - intra-/intra-judge consistency within Rasch model

"The basic flaw of many applications of IRT modelling in language testing especially is that there is not enough evidence provided about the model-data fit, which makes the findings of these studies more or less questionable" (p.17).

(Kaftandjieva, 2004)

internal validity: RMT



internal validity RMT

individual

Fit Statistics (*Infit MnSq* (*Zstd*))

Corr. Ptbis. >.20

Obs.% vs Exp. %

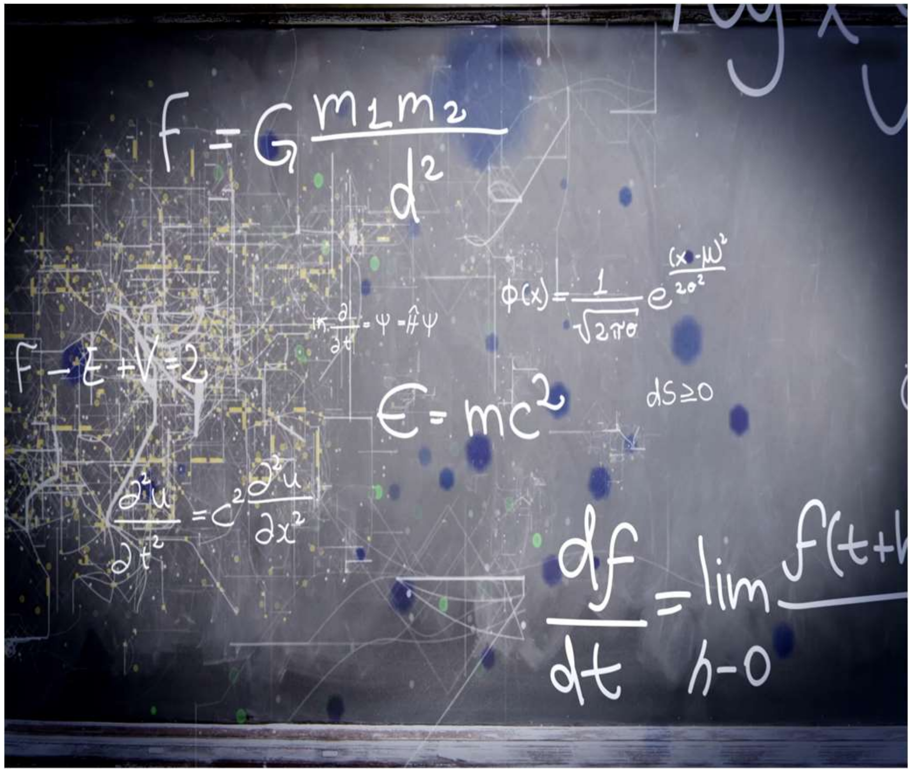
group

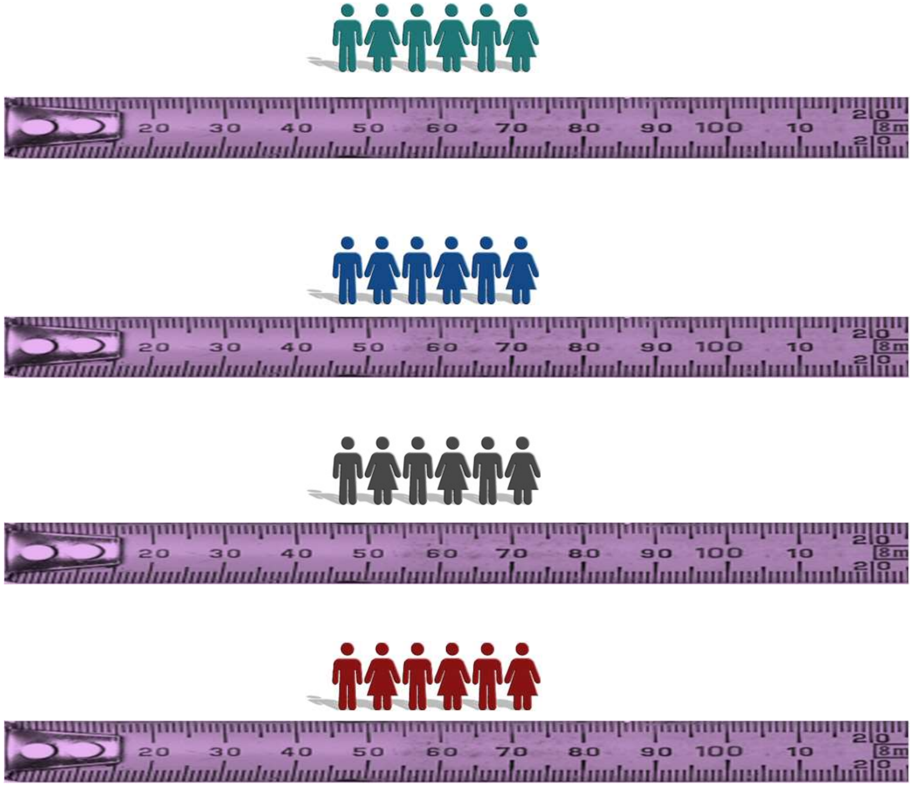
Separation indices: (*G*), (*H*), (*R*)

Chi-square statistic (χ^2)

Obs. % vs Exp. %

measurement model





many-facet Rasch measurement (MFRM) model

The MFRM model (Rounds 1 & 2)

$$\log\left(\frac{P_{nijk1}}{P_{nijk0}}\right) \equiv B_n - D_i - G_m - M_i - O_i - F_t - R_j - D_y$$

P_{nijk1} = prob. “Yes” awarded on item *i* by judge *n*,

P_{nijk0} = prob. “No” awarded on item *i* by judge *n*,

B_n = leniency of judge *n*,

D_i = difficulty of item *i*,

G_m= severity of group *m*,

M_i = difficulty of the medium *i*,

O_i = difficulty of the order *i*,

F_t = difficulty of test form *t*,

R_j = judgment of performance standard for round *j*,

D_y= difficulty of rating a “*Yes*” relative to “*No*”

The MFRM model (Round 3)

$$\log\left(\frac{P_{nijk1}}{P_{nijk-\mathbf{1}}}\right) \equiv B_n - D_i - G_m - M_i - O_i - F_t - R_j - T_{ik}$$

P_{nijk1} = prob. *k* awarded on item *i* by judge *n*,

P_{nijk-} = prob. *k-1* awarded on item *i* by judge *n*,

B_n = leniency of judge *n*,

D_i = difficulty of item *i*,

G_m= severity of group *m*,

M_i = difficulty of the medium *i*,

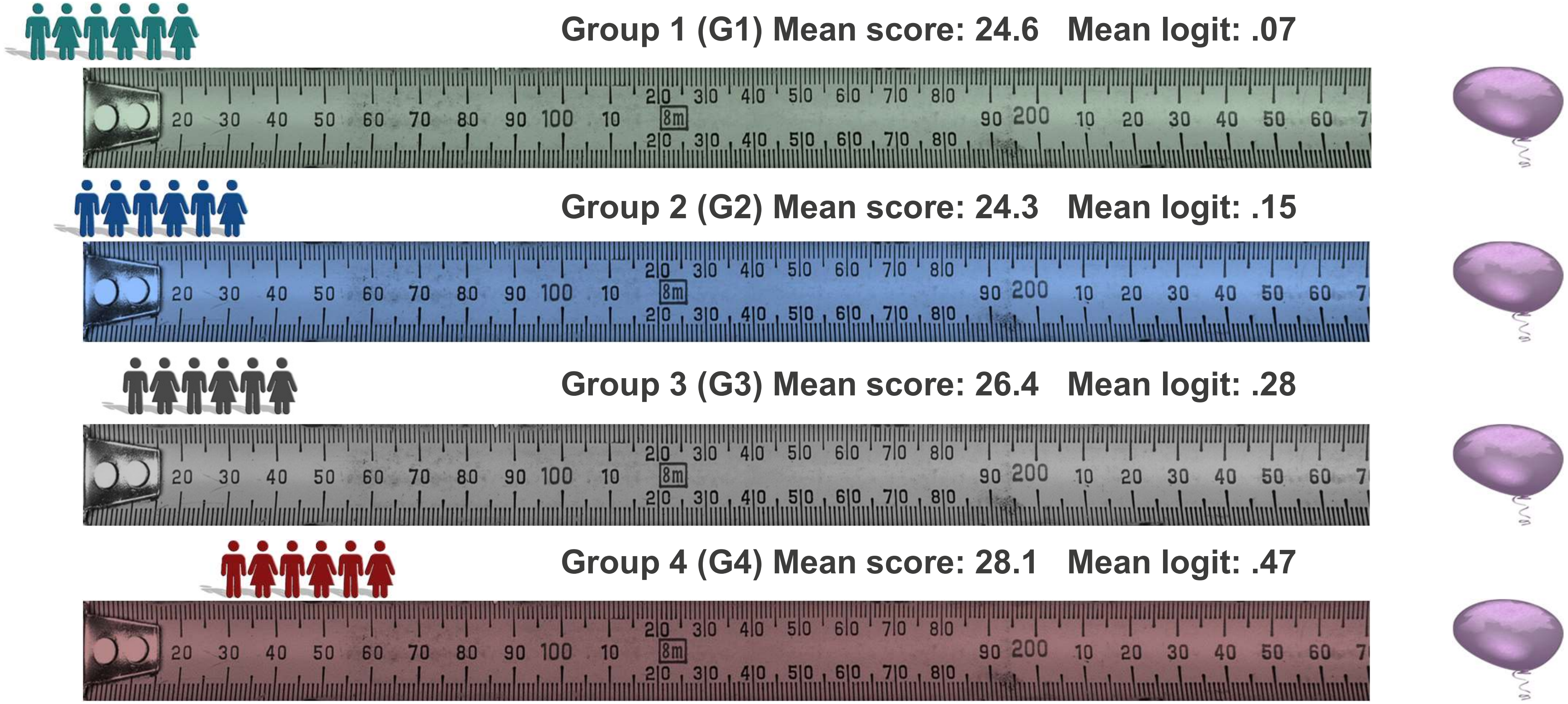
O_i = difficulty of the order *i*,

F_t = difficulty of test form *t*,

R_j = judgment of performance standard for round *j*,

T_{ik}= difficulty of assigning k relative to *k* – 1.

separate analysis



Group 1 (G1) Mean score: 24.6 Mean logit: .07

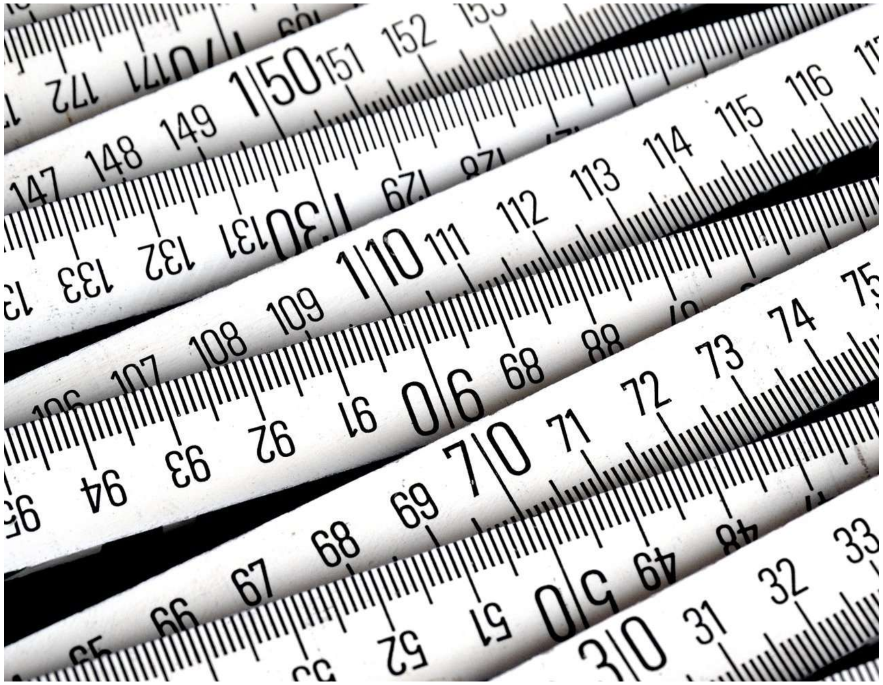
Group 2 (G2) Mean score: 24.3 Mean logit: .15

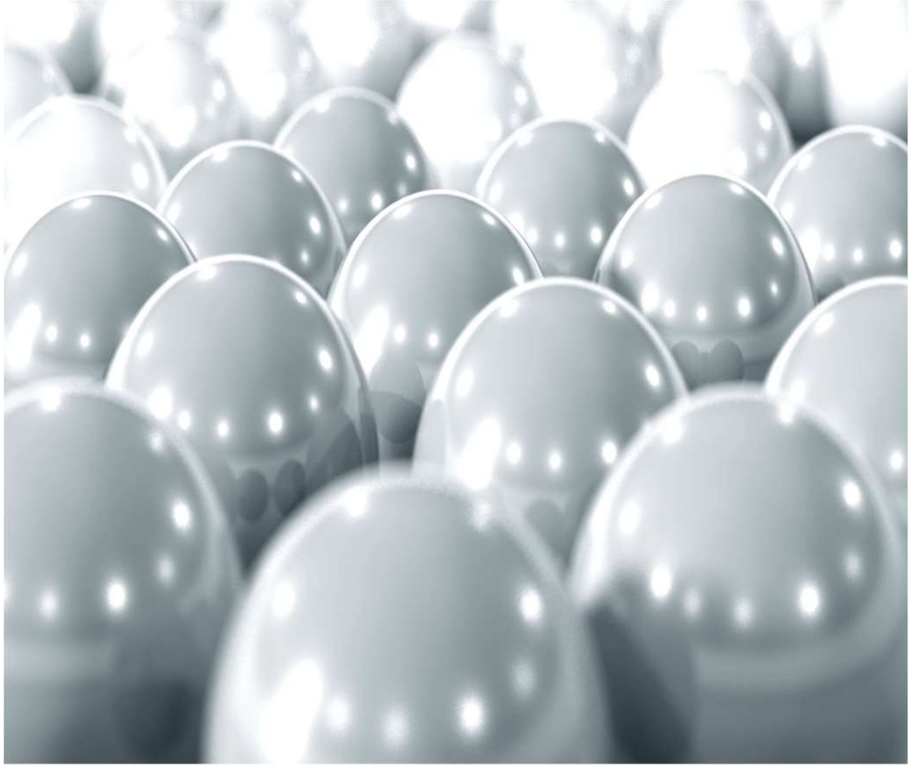
Group 3 (G3) Mean score: 26.4 Mean logit: .28

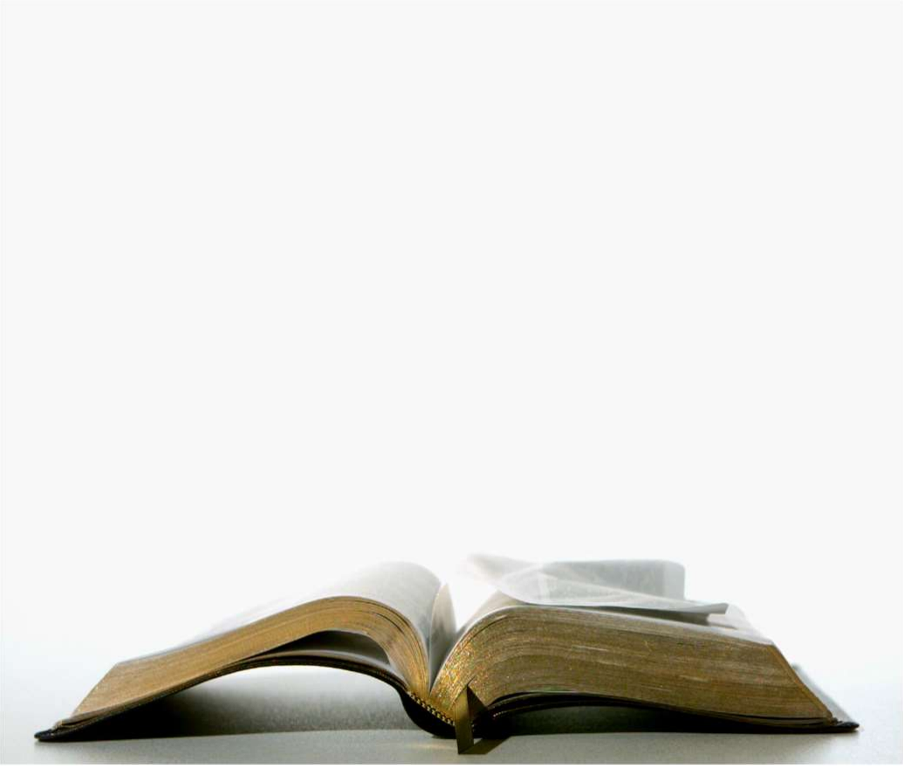
Group 4 (G4) Mean score: 28.1 Mean logit: .47

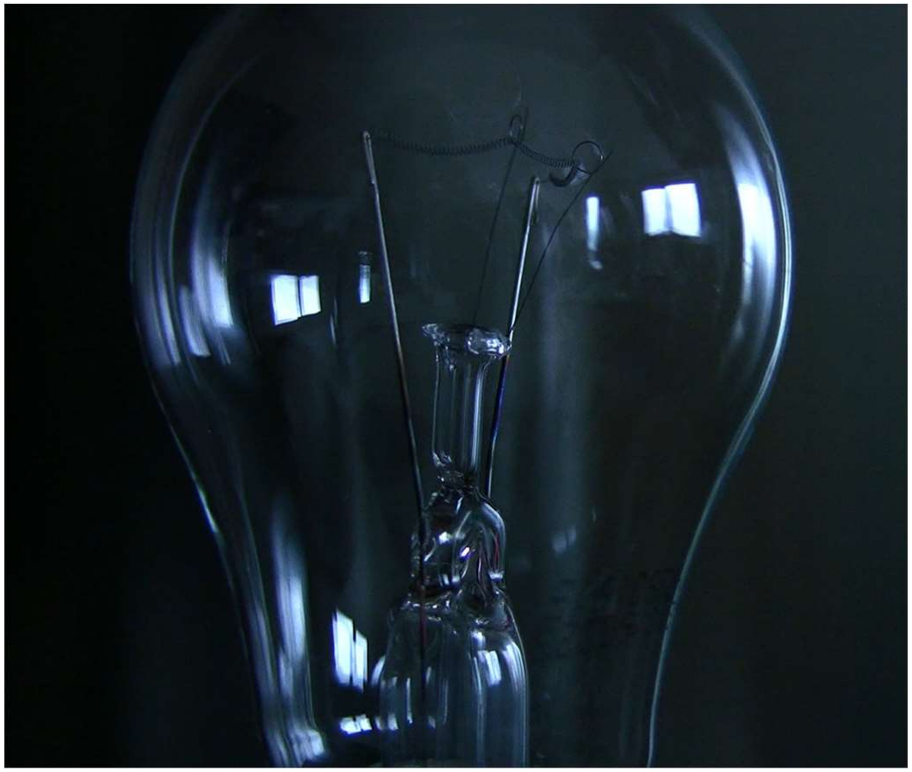
test form A – score table

Raw Score	Logit (S.E)	Raw Score	Logit (S.E)	Raw Score	Logit (S.E)	Raw Score	Logit (S.E)	Raw Score	Logit (S.E)
0	-5.08 (1.83)	10	-1.24 (.37)	20	-.11 (.32)	30	.92 (.33)	40	2.43 (.49)
1	-3.86 (1.02)	11	-1.10 (.36)	21	-.01 (.32)	31	1.03 (.34)	41	2.69 (.54)
2	-3.13 (.73)	12	-.98 (.35)	22	.09 (.32)	32	1.15 (.35)	42	3.02 (.61)
3	-2.69 (.61)	13	-.86 (.34)	23	.19 (.32)	33	1.27 (.35)	43	3.46 (.73)
4	-2.37 (.53)	14	-.74 (.34)	24	.29 (.32)	34	1.40 (.36)	44	4.19 (1.02)
5	-2.11 (.48)	15	-.63 (.33)	25	.39 (.32)	35	1.54 (.38)	45	5.42 (1.84)
6	-1.89 (.45)	16	-.52 (.33)	26	.49 (.32)	36	1.68 (.39)		
7	-1.70 (.42)	17	-.41 (.32)	27	.60 (.32)	37	1.84 (.41)		
8	-1.53 (.40)	18	-.31 (.32)	28	.70 (.33)	38	2.01 (.43)		
9	-1.38 (.39)	19	-.21 (.32)	29	.81 (.33)	39	2.21 (.45)		

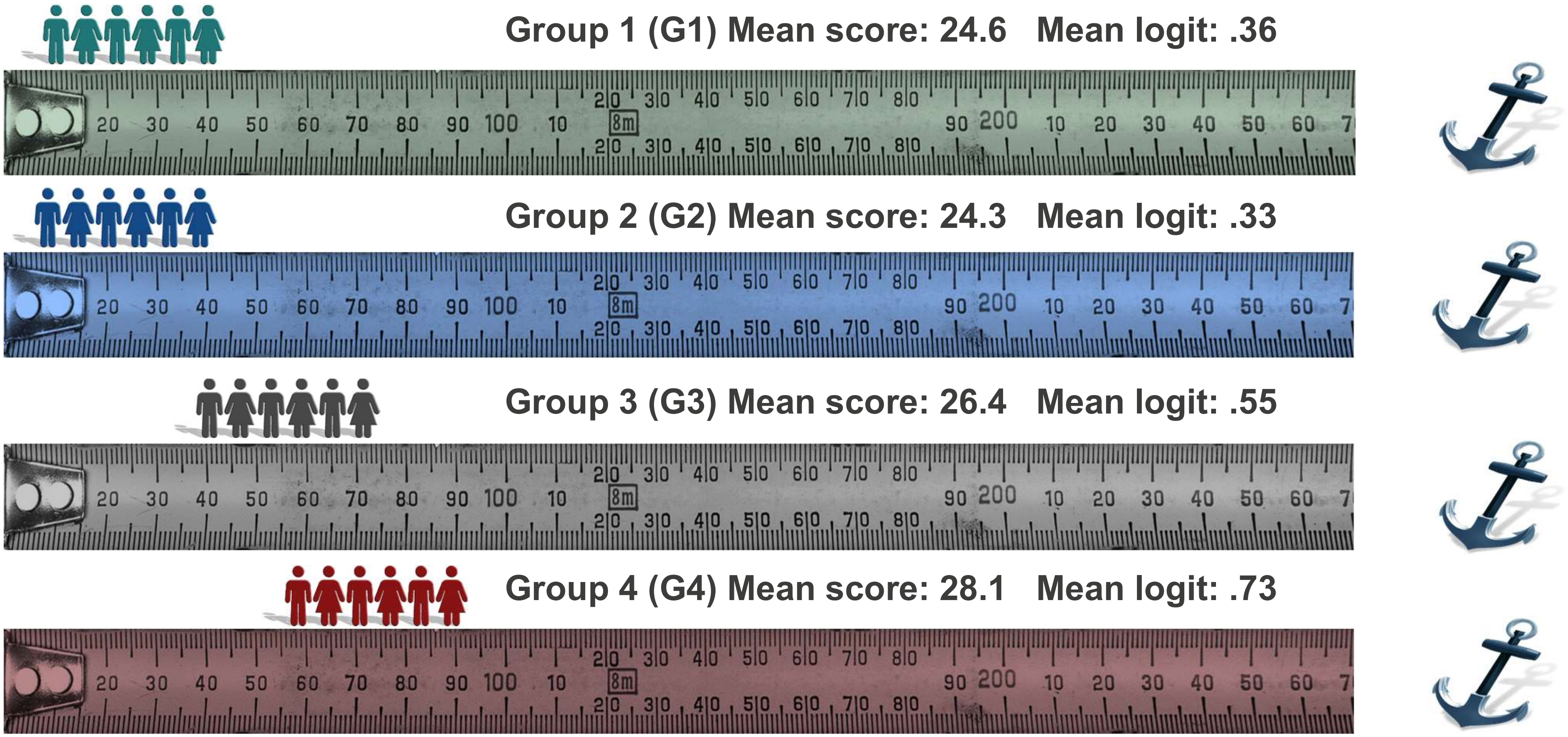








anchored analysis



Group 1 (G1) Mean score: 24.6 Mean logit: .36

Group 2 (G2) Mean score: 24.3 Mean logit: .33

Group 3 (G3) Mean score: 26.4 Mean logit: .55

Group 4 (G4) Mean score: 28.1 Mean logit: .73

test form A – score table

Raw Score	Logit (S.E)	Raw Score	Logit (S.E)	Raw Score	Logit (S.E)	Raw Score	Logit (S.E)	Raw Score	Logit (S.E)
0	-5.08 (1.83)	10	-1.24 (.37)	20	-.11 (.32)	30	.92 (.33)	40	2.43 (.49)
1	-3.86 (1.02)	11	-1.10 (.36)	21	-.01 (.32)	31	1.03 (.34)	41	2.69 (.54)
2	-3.13 (.73)	12	-.98 (.35)	22	.09 (.32)	32	1.15 (.35)	42	3.02 (.61)
3	-2.69 (.61)	13	-.86 (.34)	23	.19 (.32)	33	1.27 (.35)	43	3.46 (.73)
4	-2.37 (.53)	14	-.74 (.34)	24	.29 (.32)	34	1.40 (.36)	44	4.19 (1.02)
5	-2.11 (.48)	15	-.63 (.33)	25	.39 (.32)	35	1.54 (.38)	45	5.42 (1.84)
6	-1.89 (.45)	16	-.52 (.33)	26	.49 (.32)	36	1.68 (.39)		
7	-1.70 (.42)	17	-.41 (.32)	27	.60 (.32)	37	1.84 (.41)		
8	-1.53 (.40)	18	-.31 (.32)	28	.70 (.33)	38	2.01 (.43)		
9	-1.38 (.39)	19	-.21 (.32)	29	.81 (.33)	39	2.21 (.45)		

quantitative

analysis:

- Wilcoxson signed-rank test/ Sign test

qualitative

analysis:

- open-ended

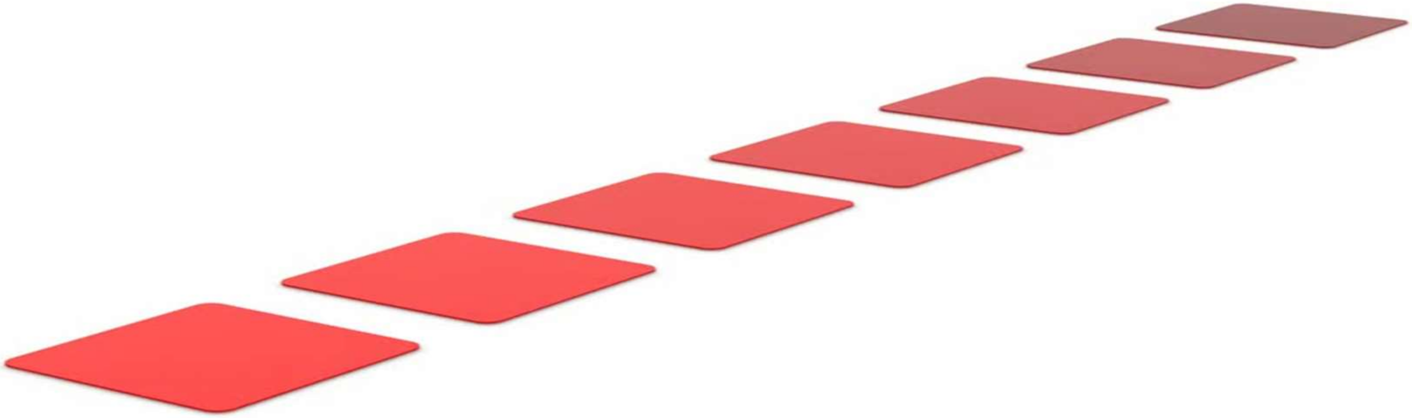
qualitative

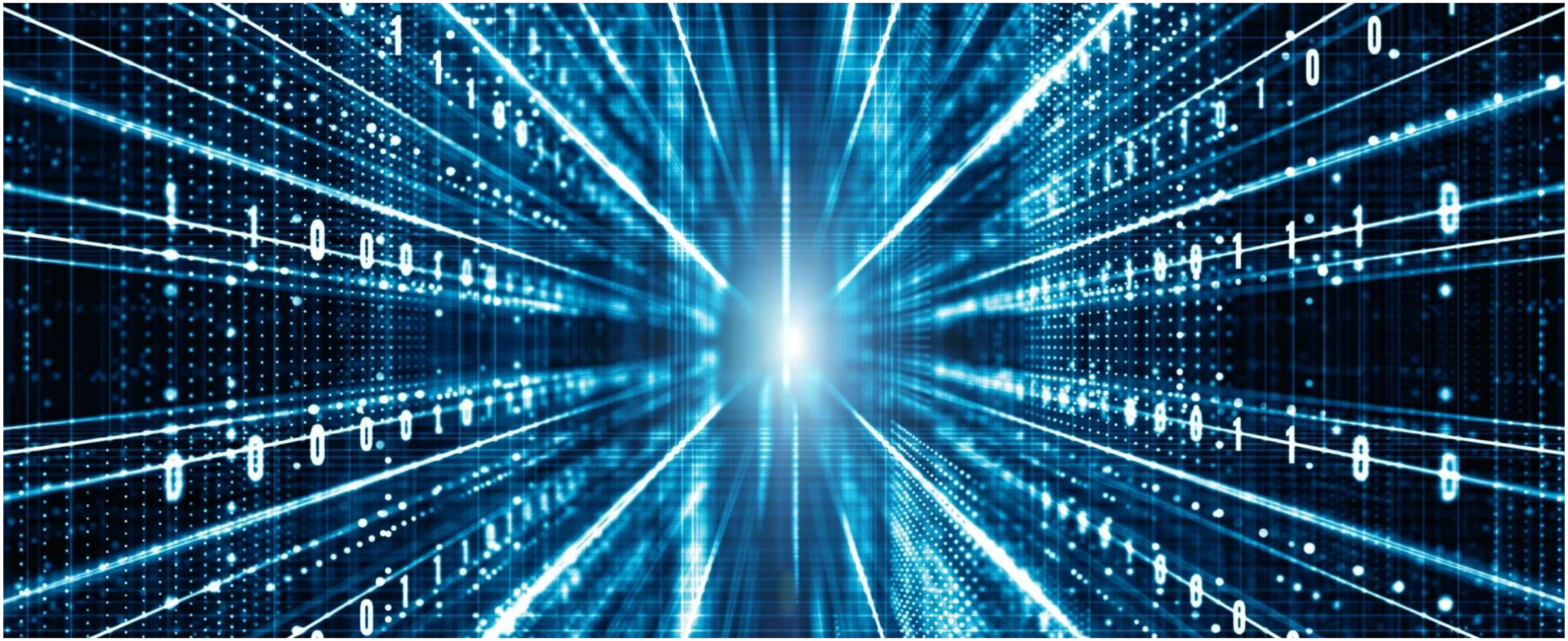
analysis:

constant comparative method (CCM)

FOCUS GROUP

(Corbin & Strauss, 2015; Glaser, 1965; Glaser & Strauss, 1967)





virtual cut scores

```
+-----------------------+---------------------------------------------------------------------------------------------------------------------+
|Measr|               ROUND 1              ||                           ROUND 2                  ||                  ROUND 3                  |
|-----+-----------------+------------------++------------------------------+---------------------++--------------------++---------------------+
|Measr|+AUDIO           |+VIDEO            ||+AUDIO                        |+VIDEO               ||+AUDIO              ||+VIDEO               |
|-----+-----------------+------------------++------------------------------+---------------------++--------------------++---------------------+
|   1 +                 +                  ++                              +                     ++                    ++                     +
|     |                 |                  ||                              |                     ||                    ||                     |
|     |                 |                  ||                              |                     ||                    || G3 (.86)            |
|     | G4(.73)         |                  ||                              | G3 (.75)            ||                    ||                     |
|     |                 |                  || G3 (.61)                     | G2 (.67)            ||                    ||                     |
|     |                 | G3(.55)          || G1 (.57)  G2 (.50)  G4 (.50) | G4 (.56)            || G1 (.57)  G3 (.59) || G2 (.55)            |
|     | G3(.43)         |                  ||                              | G1 (.45)            || G2 (.50)  G4 (.49) || G1 (.49)  G4 (.43)  |
|     | G1(.36)  G2(.35)| G2(.33)          ||                              |                     ||                    ||                     |
|     |                 | G1(.27)  G4(.28) ||                              |                     ||                    ||                     |
|     |                 |                  ||                              |                     ||                    ||                     |
|     |                 |                  ||                              |                     ||                    ||                     |
*   0 *                 *                  **                              *                     **                    **                     *
|-----+-----------------+------------------++------------------------------+---------------------++--------------------++---------------------+
|Measr|+AUDIO           |+VIDEO            ||+AUDIO                        |+VIDEO               ||+AUDIO              ||+VIDEO               |
+-----------------------+---------------------------------------------------------------------------------------------------------------------+
```

round 1: pairwise interactions (DMF)

Group	Group	Sig.
G1-A	G2-A	No
G1-A	G3-A	No
G1-A	G4-A	No
G1-A	G1-V	No
G1-A	G2-V	No
G1-A	G3-V	No
G1-A	G4-V	No
G2-A	G3-A	No
G2-A	G4-A	No
G2-A	G1-V	No
G2-A	G2-V	No
G2-A	G3-V	No
G2-A	G4-V	No
G3-A	G4-A	No
G3-A	G1-V	No
G3-A	G2-V	No
G3-A	G3-V	No
G3-A	G4-V	No
G4-A	G1-V	No
G4-A	G2-V	No
G4-A	G3-V	No
G4-A	G4-V	No
G1-V	G2-V	No
G1-V	G3-V	No
G1-V	G4-V	No
G2-V	G3-V	No
G2-V	G4-V	No
G3-V	G4-V	No

round 2: pairwise interactions (DMF)

Group	Group	Sig.
G1-A	G2-A	No
G1-A	G3-A	No
G1-A	G4-A	No
G1-A	G1-V	No
G1-A	G2-V	No
G1-A	G3-V	No
G1-A	G4-V	No
G2-A	G3-A	No
G2-A	G4-A	No
G2-A	G1-V	No
G2-A	G2-V	No
G2-A	G3-V	No
G2-A	G4-V	No
G3-A	G4-A	No
G3-A	G1-V	No
G3-A	G2-V	No
G3-A	G3-V	No
G3-A	G4-V	No
G4-A	G1-V	No
G4-A	G2-V	No
G4-A	G3-V	No
G4-A	G4-V	No
G1-V	G2-V	No
G1-V	G3-V	No
G1-V	G4-V	No
G2-V	G3-V	No
G2-V	G4-V	No
G3-V	G4-V	No

round 3: pairwise interactions (DMF)

Group	Group	Sig.
G1-A	G2-A	No
G1-A	G3-A	No
G1-A	G4-A	No
G1-A	G1-V	No
G1-A	G2-V	No
G1-A	G3-V	No
G1-A	G4-V	No
G2-A	G3-A	No
G2-A	G4-A	No
G2-A	G1-V	No
G2-A	G2-V	No
G2-A	G3-V	No
G2-A	G4-V	No
G3-A	G4-A	No
G3-A	G1-V	No
G3-A	G2-V	No
G3-A	G3-V	No
G3-A	G4-V	No
G4-A	G1-V	No
G4-A	G2-V	No
G4-A	G3-V	No
G4-A	G4-V	No
G1-V	G2-V	No
G1-V	G3-V	No
G1-V	G4-V	No
G2-V	G3-V	No
G2-V	G4-V	No
G3-V	G4-V	No

virtual cut score comparisons

virtual cut scores

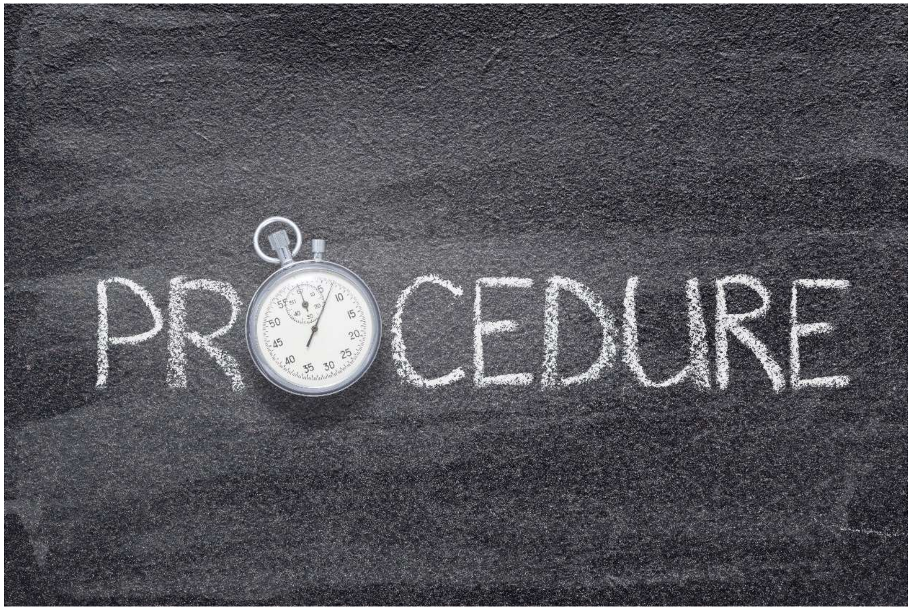
- reliable
- comparable
- valid

virtual panels

- no differential medium functioning (DMF)

survey items

items



perception survey frequency data

	Audio	Video
1 (Strongly Disagree)	10 (0.29%)	2 (0.00%)
2 (Disagree)	11 (0.32%)	15 (0.43%)
3 (Slightly Disagree)	71(2.05%)	58 (1.67%)
4 (Slightly Agree)	253 (7.30%)	229 (6.61%)
5 (Agree)	2061 (59.48%)	2149 (62.02%)
6 (Strongly Agree)	1059 (33.02%)	1010 (29.15%)
Missing	0 (0.00%)	2 (0.06%)
Total	3465 (100%)	3465 (100%)

92.50%

91.17%

perception survey frequency data cont.

	Audio	Video
1 (Strongly Disagree)	10 (0.29%)	2 (0.06%)
2 (Disagree)	11 (0.32%)	15 (0.43%)
3 (Slightly Disagree)	71(2.05%)	58 (1.67%)
4 (Slightly Agree)	253 (7.30%)	229 (6.61%)
5 (Agree)	2061 (59.48%)	2149 (62.02%)
6 (Strongly Agree)	1059 (30.56%)	1010 (29.15%)
Missing	0 (0.00%)	2 (0.06%)
Total	3465 (100%)	3465 (100%)

2.66%

2.16%

procedural survey frequency data

	Audio	Video
1 (Strongly Disagree)	6 (0.26%)	8 (0.34%)
2 (Disagree)	11 (0.47%)	5 (0.21%)
3 (Slightly Disagree)	33 (1.41%)	14 (0.60%)
4 (Slightly Agree)	188 (8.03%)	132 (6.22%)
5 (Agree)	1237 (52.86%)	1333 (56.97%)
6 (Strongly Agree)	858 (36.67%)	840 (35.90%)
Missing	7 (0.30%)	8 (0.34%)
Total	2340 (100%)	2340 (100%)

89.53%

92.86%

procedural survey frequency data cont.

	Audio	Video
1 (Strongly Disagree)	6 (0.26%)	8 (0.34%)
2 (Disagree)	11 (0.47%)	5 (0.21%)
3 (Slightly Disagree)	33 (1.41%)	14 (0.60%)
4 (Slightly Agree)	188 (8.03%)	132 (6.22%)
5 (Agree)	1237 (52.86%)	1333 (56.97%)
6 (Strongly Agree)	858 (36.67%)	840 (35.90%)
Missing	7 (0.30%)	8 (0.34%)
Total	2340 (100%)	2340 (100%)

2.14%

1.15%

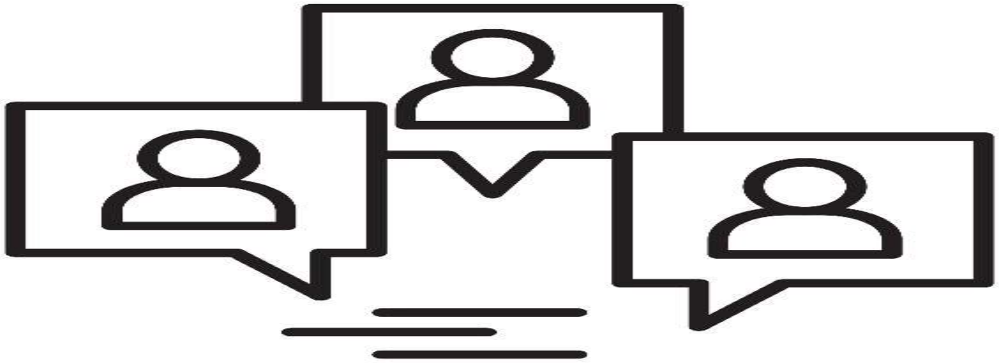
quantitative & qualitative

quantitative

- no preference towards specific medium

qualitative

- preference towards video medium



FOCUS GROUP

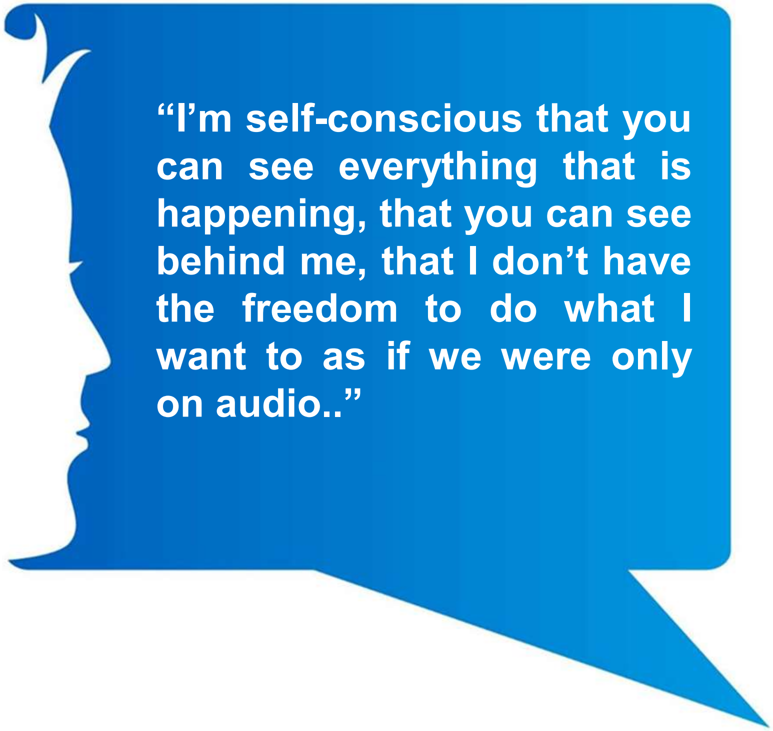
1. psychological aspects;
2. interaction;
3. technical aspects;
4. convenience;
5. decision-making process.

psychological aspects

few distractions: audio medium (+)

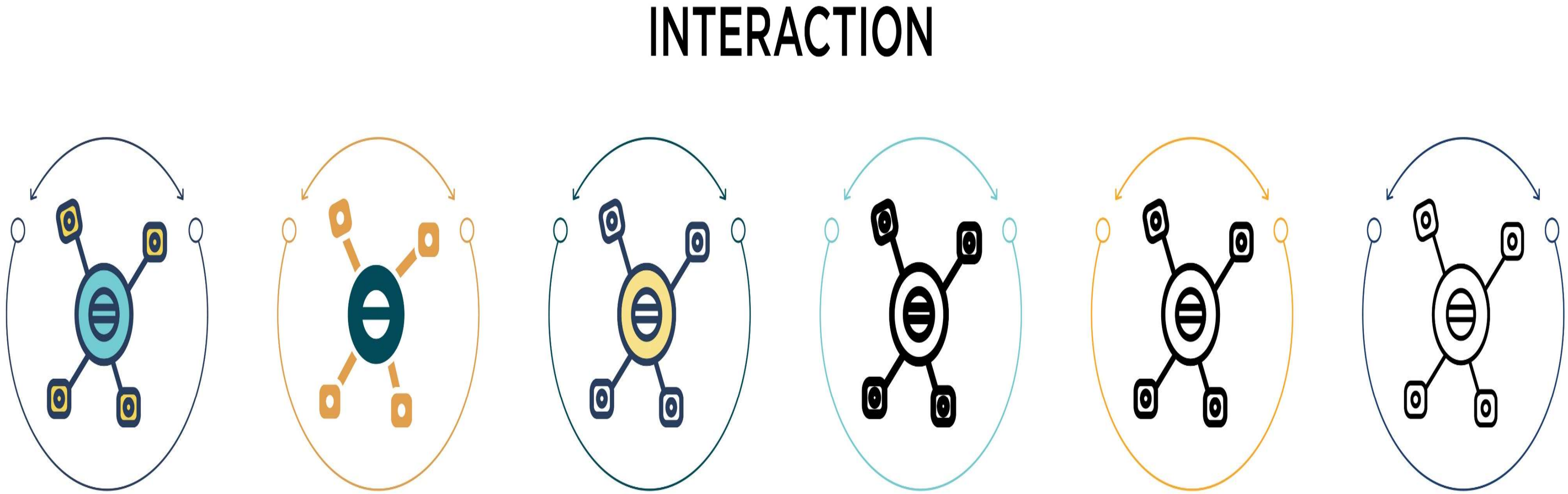
“… but when we used audio we were not so distracted so much, we were more concentrated on what we were supposed to do.”

self-awareness: audio medium (+)



"I'm self-conscious that you can see everything that is happening, that you can see behind me, that I don't have the freedom to do what I want to as if we were only on audio.."

interaction



lack of small talk: f2f *vs* virtual environment

"… when you meet someone … you have some time to get to know one another other … so it becomes a bit more personal … the positive aspect of this system [online communication] is that it is more professional, on the other hand, it is less personal…".

fewer digressions: f2f *vs* virtual environment

" … I didn't feel that at any point our discussion went off topic whereas this may happen in face-to-face situations. We were always on topic and very focused on what we were discussing … "

technical aspects

FROZEN!

convenience

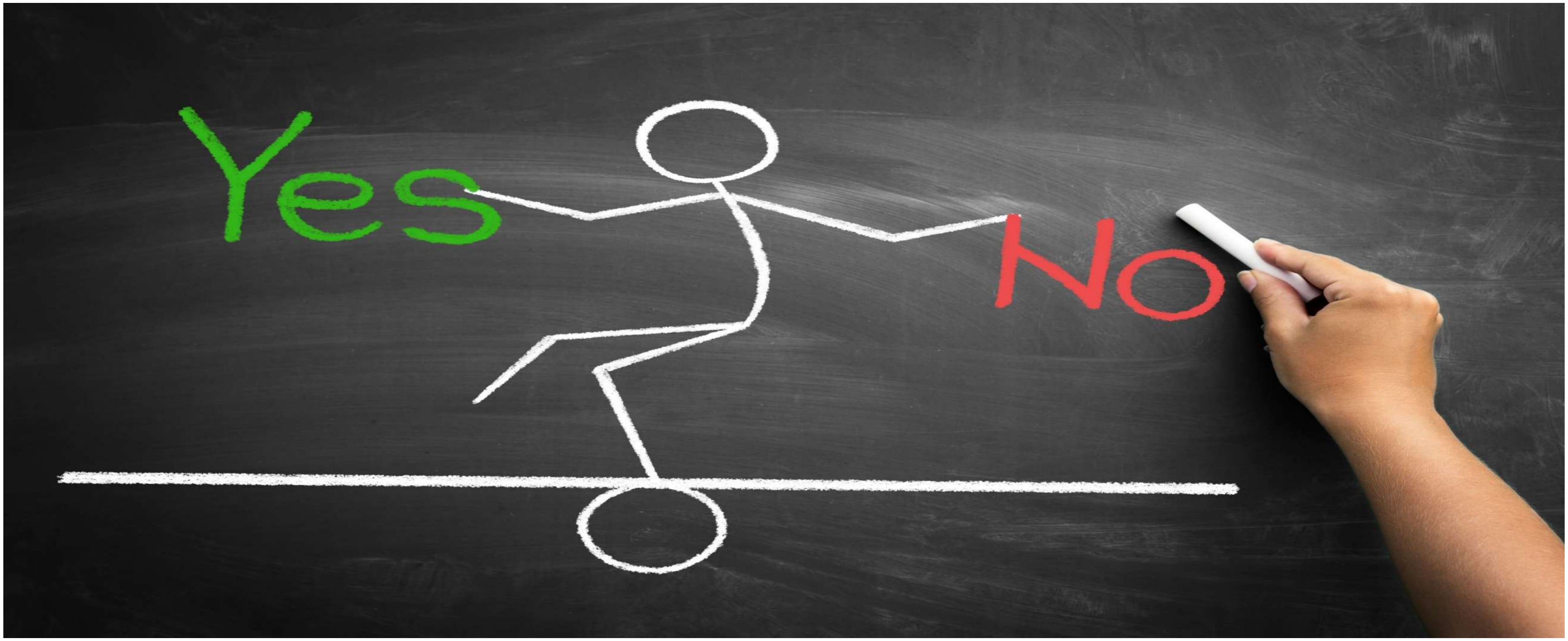
time-saving

“… we don't have to travel to a place or come back home or wait for busses and other means of transport …”

less fatigue

"… the fact that I was at home and I could do the whole thing in the comfort of my home was very convenient for me. I mean, I would have been exhausted if there were an equivalent workshop face-to face. So yes, I was tired, but not too tired".

decision-making process



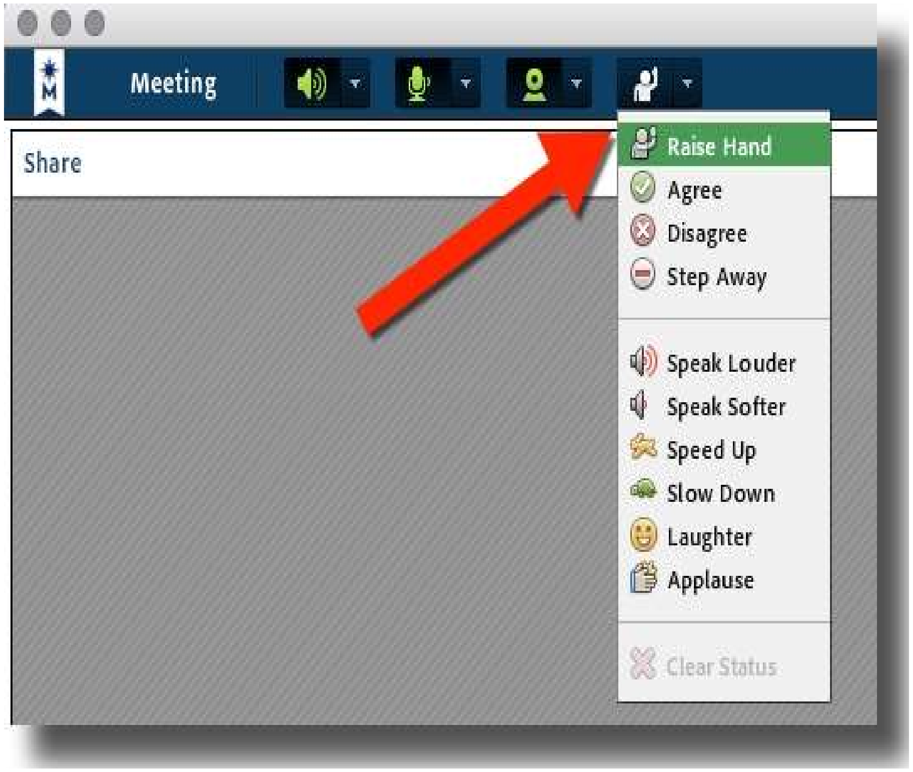
decision-making process

"Um, personally, for me, it didn't. It was the empirical data that you showed us that influenced my original opinions … I think no one changed their opinion because they were able to look at someone actually saying something else. So no".

NETIQUETTE
LOADING...

turn-taking system

“ … felt that the raising hand symbol was very convenient because it enabled us to speak whenever we wanted to … express our opinion, etc. And, … it helped the whole process so it worked very well”.



qualitative

qualitative

- preference towards video medium

synchronous full-day workshop feasible;

judges' preference towards video;

both media equally appropriate for setting virtual cut scores;

virtual cut scores reliable, comparable, and valid;

virtual media not hindering communication.

judges' decision-making processes not hindered by virtual environment;

facilitators to select virtual environment that best …

… **suits** workshop needs;

... **meets** technical & pragmatic geographical limitations (facilitator/ panellists);

… **caters** for panellists' video reservations

a wider panellists selection;

reduction in associated F2F costs;

cut score studies conducted and/or replicated;

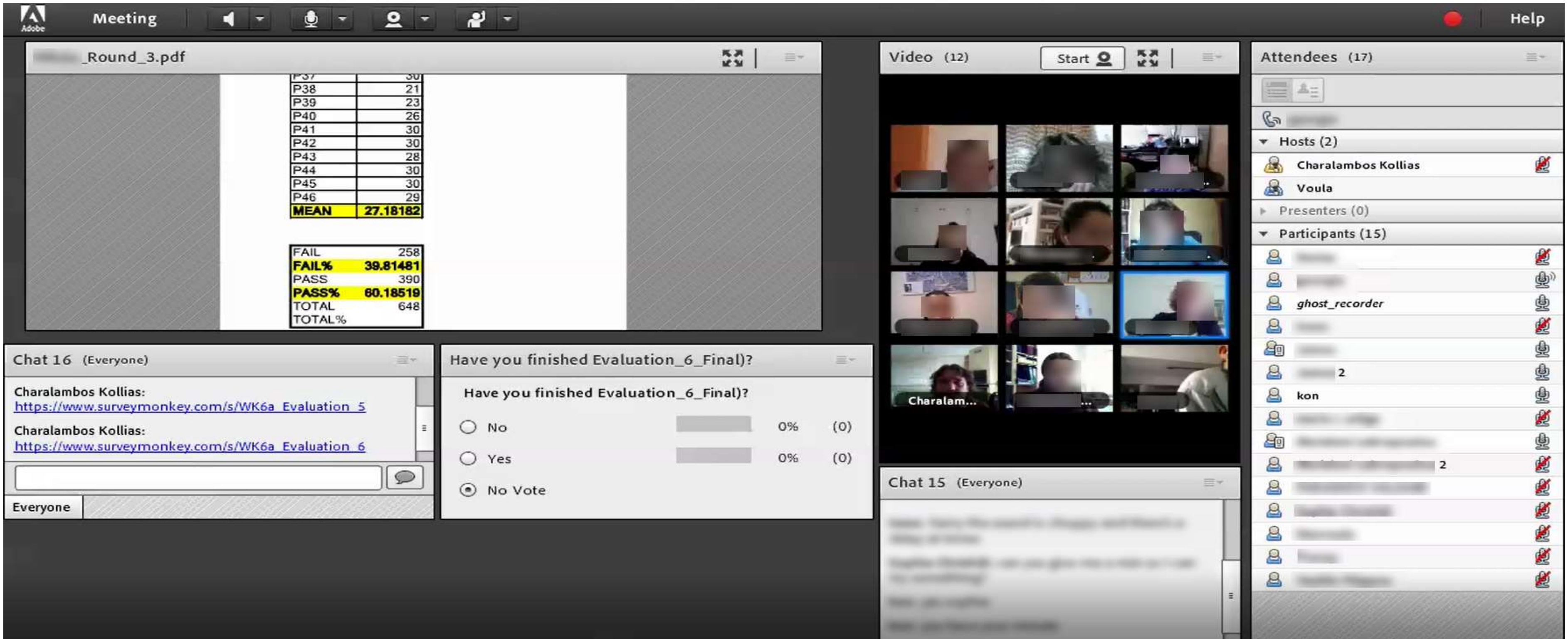
future research

research opportunities . . .

- test security
- other SS methods
- other CEFR levels
- other skills:
  - listening
  - speaking
  - writing

- discussion in virtual environment
- concurrent verbal reports of judge ratings
  - break out rooms
- comparison of audio, video, and F2F

judges' preference towards video medium



[redacted]_Round_3.pdf

P37	30
P38	21
P39	23
P40	26
P41	30
P42	30
P43	28
P44	30
P45	30
P46	29
MEAN	**27.18182**

FAIL	258
FAIL%	**39.81481**
PASS	390
PASS%	**60.18519**
TOTAL	648
TOTAL%	

Chat 16 (Everyone)

Charalambos Kollias:
https://www.surveymonkey.com/s/WK6a_Evaluation_5

Charalambos Kollias:
https://www.surveymonkey.com/s/WK6a_Evaluation_6

Everyone

Have you finished Evaluation_6_Final)?

Have you finished Evaluation_6_Final)?

○ No	0%	(0)
○ Yes	0%	(0)
● No Vote		

Video (12) Start

Chat 15 (Everyone)

[redacted]

Attendees (17)

[redacted]

Hosts (2)
- Charalambos Kollias
- Voula

Presenters (0)

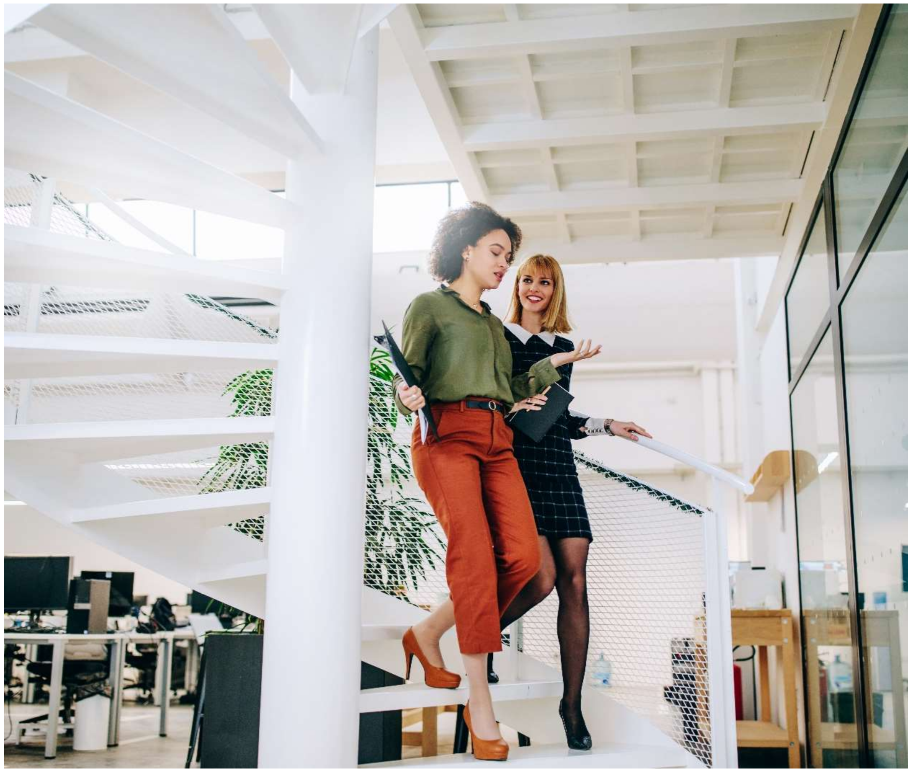
Participants (15)
- [redacted]
- [redacted]
- ghost_recorder
- [redacted]
- [redacted]
- [redacted] 2
- kon
- [redacted]
- [redacted]
- [redacted] 2
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

Kollias, C. (May, 2022). Virtual Standard Setting: The benefits, the challenges, and the way forward.

INTERPRETATION

media naturalness theory (MNT)

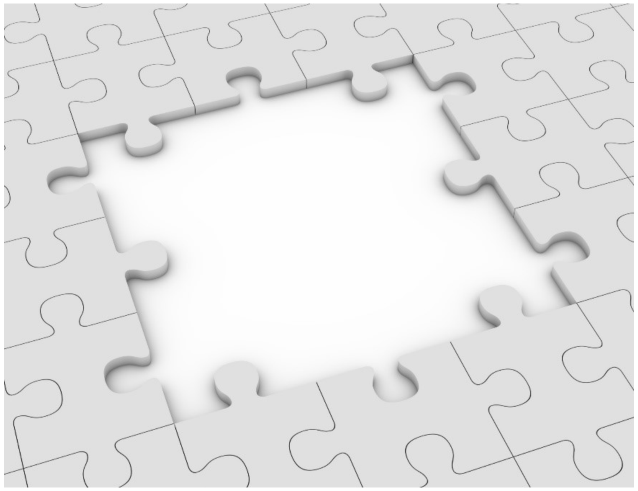
5 media naturalness elements (Kock, 2005/ 2010)



5 media naturalness elements cont.

5 media naturalness elements cont.

decrease in naturalness (Kock, 2005/ 2010)



co-location: video environment (+)

"It was as natural – after the initial 5 mins – as being physically in a room as we could see everyone …".

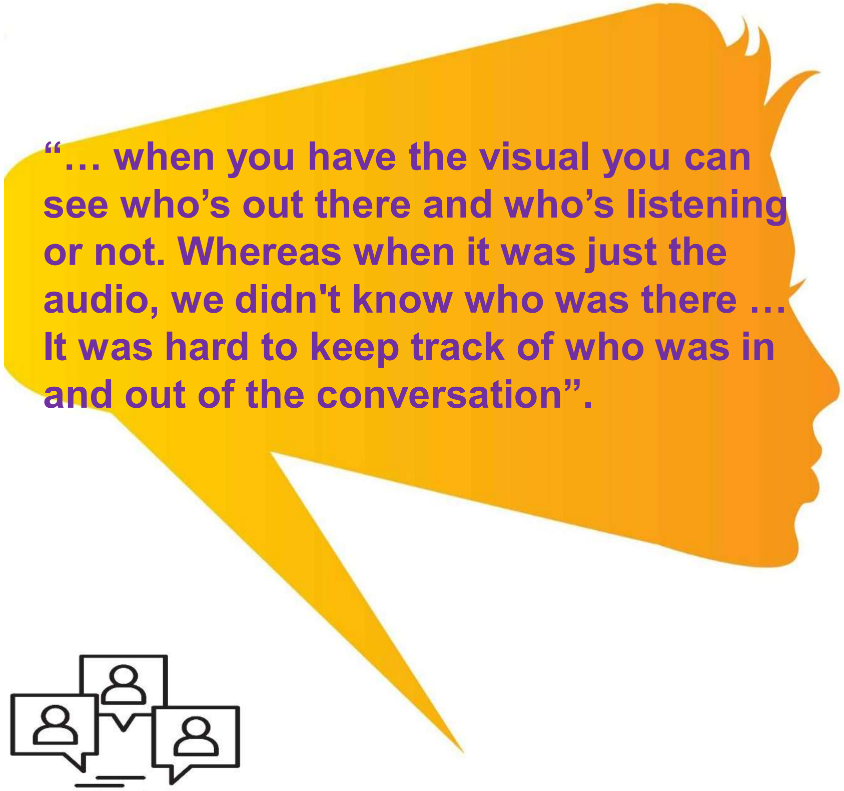
able to employ & detect facial expressions: video medium (+)

"... video helps as well, because you can see the expressions on other people's face if they agree, disagree if they want to say something".

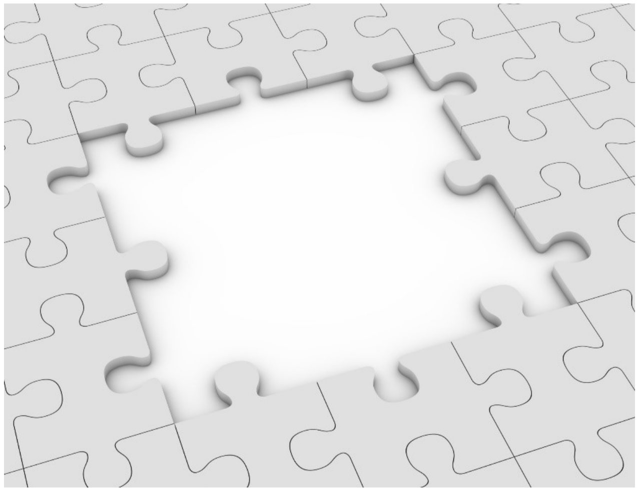
able to exchange verbal ques quickly: audio medium (+)

"… the audio helped us communicate quickly by listening to each other's thoughts and explanations".

increase in cognitive ambiguity: audio medium (-)



"… when you have the visual you can see who's out there and who's listening or not. Whereas when it was just the audio, we didn't know who was there … It was hard to keep track of who was in and out of the conversation".



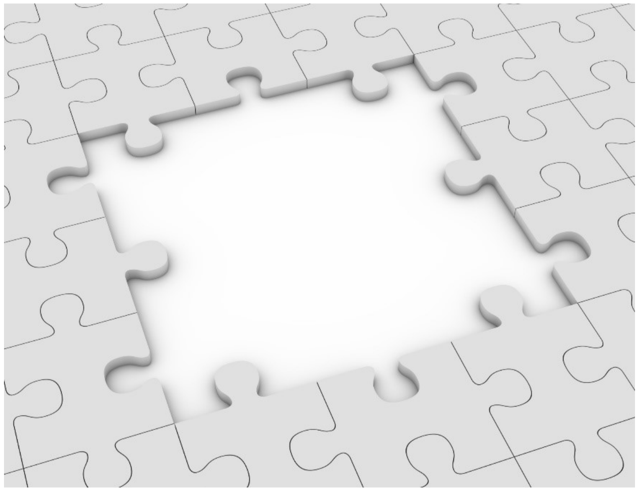
increase in cognitive effort: audio medium (-)

“… it was difficult for me to concentrate on just the voice without seeing anything on the screen. It felt like I had to concentrate twice in order to understand what was going on … it”.

decrease in physiological arousal: audio medium (-)

"… confused without the camera for some reason [and not feeling] like talking most of the time".

increase in cognitive ambiguity: audio medium (-)

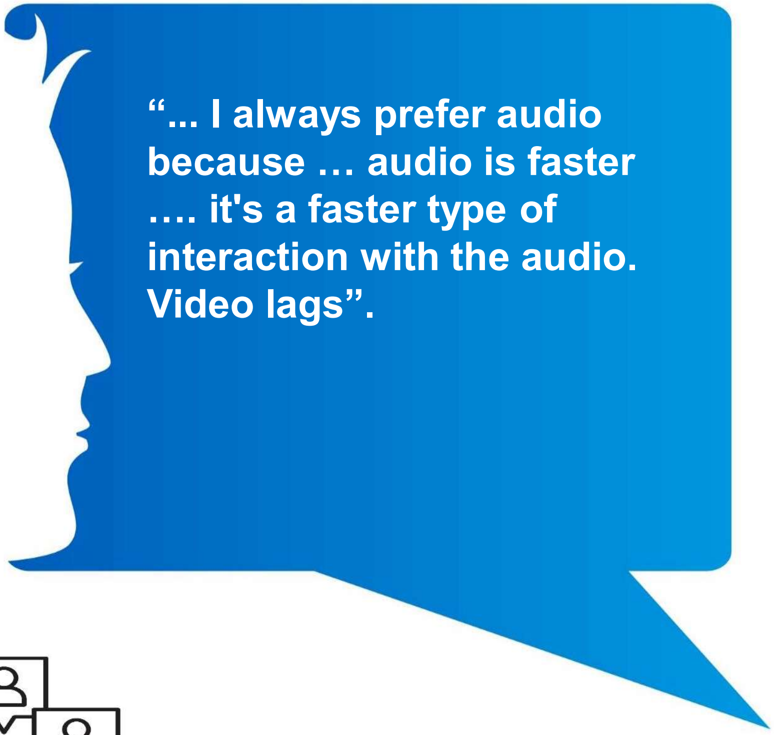


"… sometimes I couldn't understand who was speaking and I think that is more natural, more friendly to see who I'm talking to".

decrease in physiological arousal: virtual environment (-)

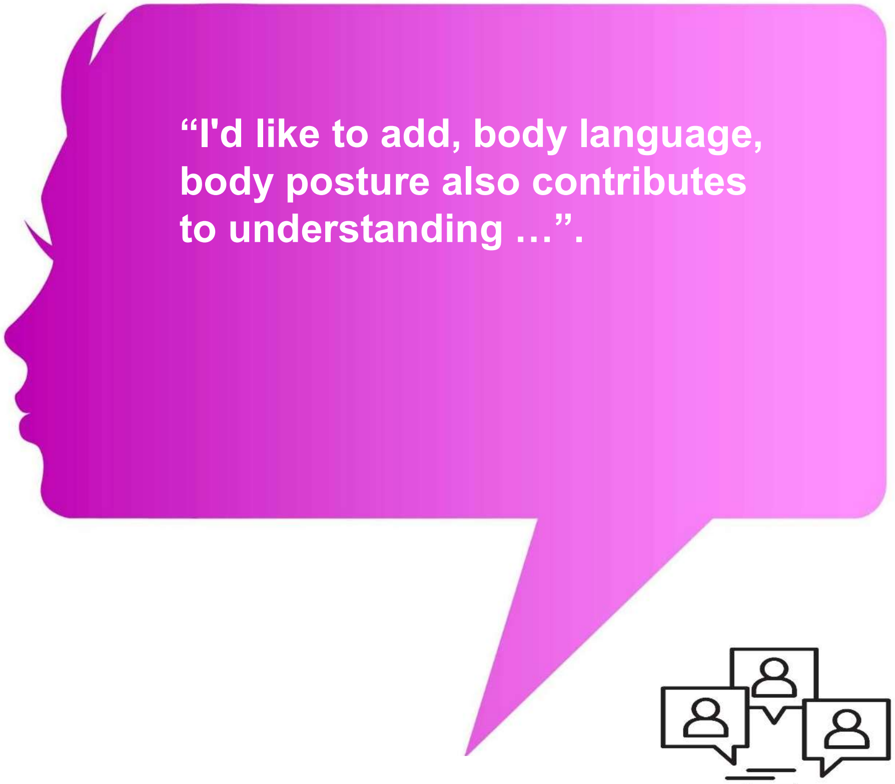
"No, no I don't think the discussion was enough … I think if they were F2F, people are more forth [sic. forthcoming] to express their opinion …".

able to exchange verbal ques quickly: audio medium (+)



"... I always prefer audio because ... audio is faster …. it's a faster type of interaction with the audio. Video lags".

able to employ & detect body language: video medium (+)

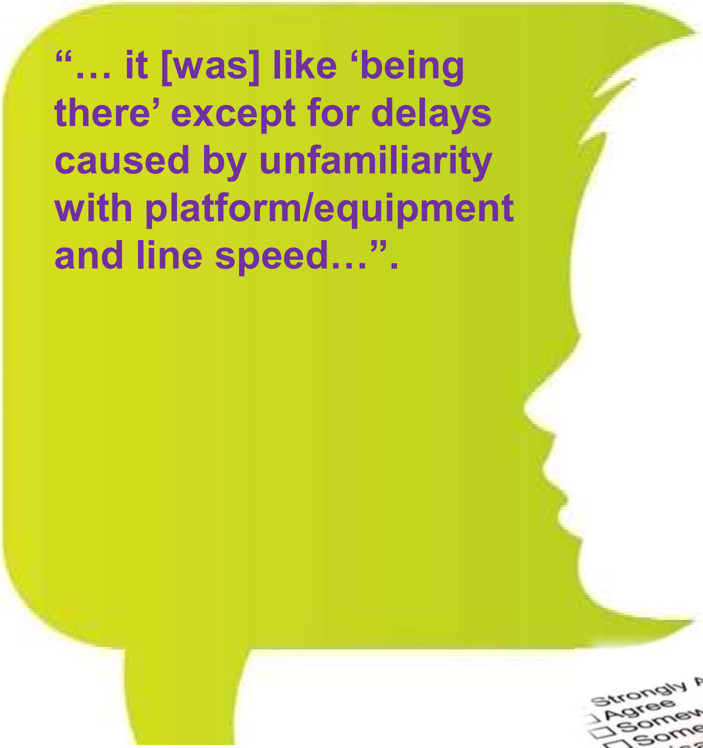


“I'd like to add, body language, body posture also contributes to understanding …”.

able to employ & detect body language: video medium (+)

“… communicate better with someone when [looking] at him, since body language helps [the judge] understand others better”.

co-location: audio medium (+)



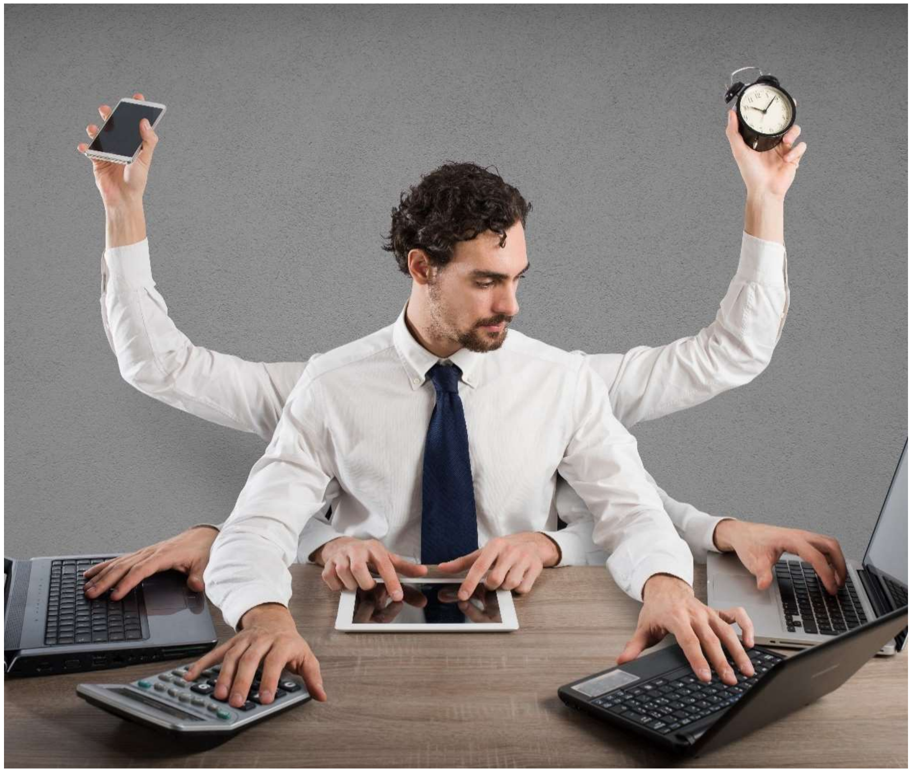
"… it [was] like 'being there' except for delays caused by unfamiliarity with platform/equipment and line speed…".

the way forward

knowledge, skills, and abilities (KSAs)

PREPARATION
SUCCESS

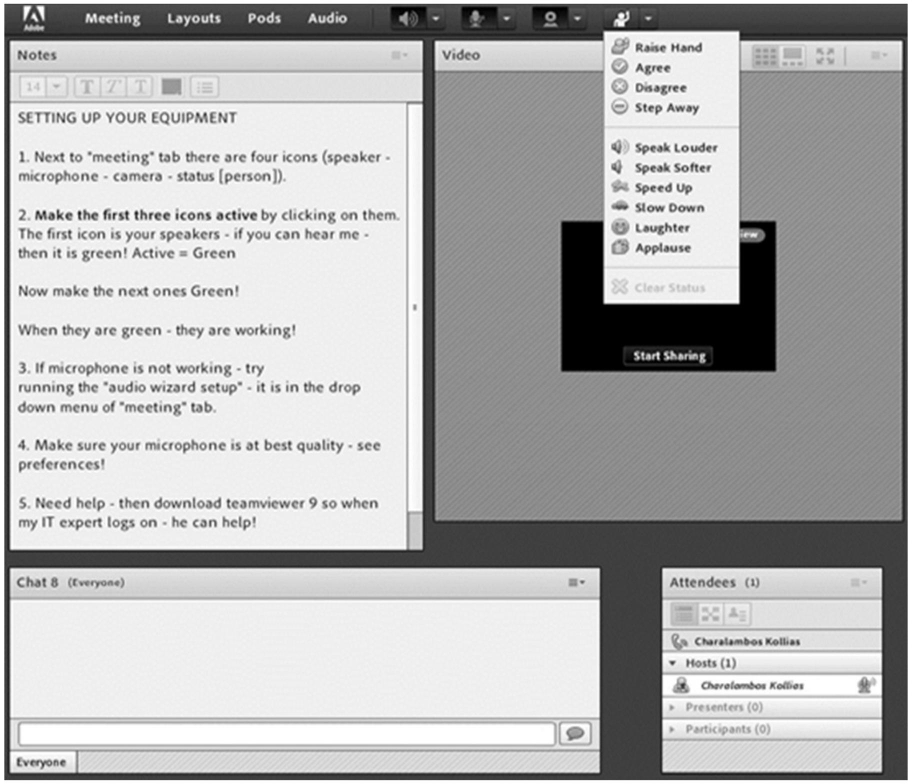
facilitator KSAs



facilitator will need to …

- establish netiquette;
- be able to multi-task;
- be thoroughly prepared;
- engage judges throughout;
- have familiarity with platform & tools;
- understand nature of technical issues.

training in platform & netiquette



SETTING UP YOUR EQUIPMENT

1. Next to "meeting" tab there are four icons (speaker - microphone - camera - status [person]).

2. **Make the first three icons active** by clicking on them. The first icon is your speakers - if you can hear me - then it is green! Active = Green

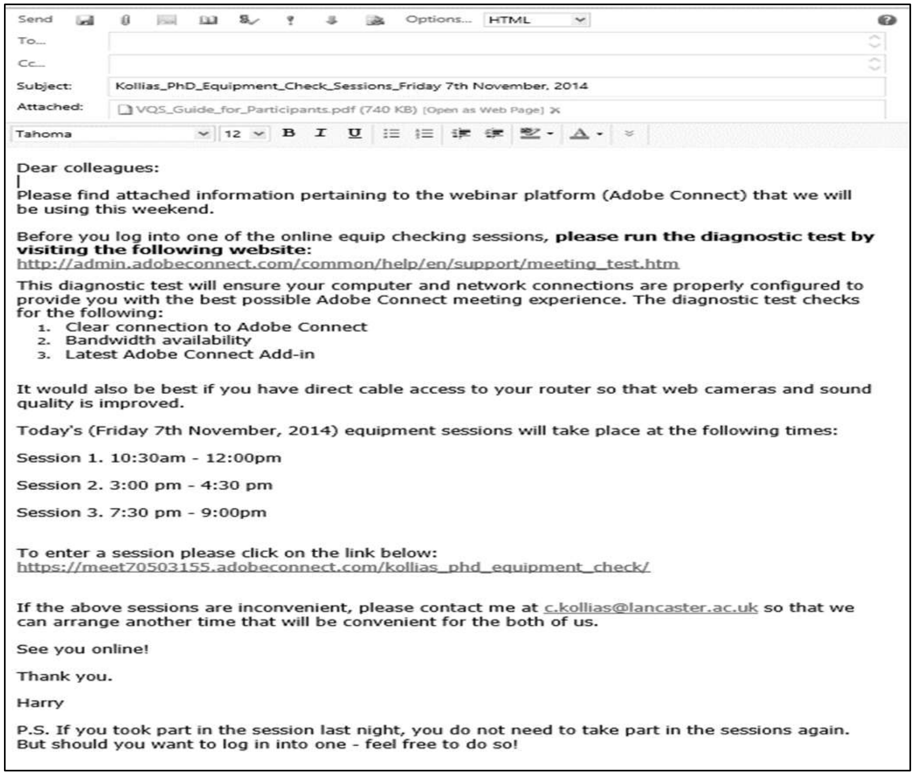
Now make the next ones Green!

When they are green - they are working!

3. If microphone is not working - try running the "audio wizard setup" - it is in the drop down menu of "meeting" tab.

4. Make sure your microphone is at best quality - see preferences!

5. Need help - then download teamviewer 9 so when my IT expert logs on - he can help!



Subject: Kollias_PhD_Equipment_Check_Sessions_Friday 7th November, 2014

Attached: VQS_Guide_for_Participants.pdf (740 KB) [Open as Web Page]

Dear colleagues:

Please find attached information pertaining to the webinar platform (Adobe Connect) that we will be using this weekend.

Before you log into one of the online equip checking sessions, **please run the diagnostic test by visiting the following website:**
http://admin.adobeconnect.com/common/help/en/support/meeting_test.htm

This diagnostic test will ensure your computer and network connections are properly configured to provide you with the best possible Adobe Connect meeting experience. The diagnostic test checks for the following:

1. Clear connection to Adobe Connect
2. Bandwidth availability
3. Latest Adobe Connect Add-in

It would also be best if you have direct cable access to your router so that web cameras and sound quality is improved.

Today's (Friday 7th November, 2014) equipment sessions will take place at the following times:

Session 1. 10:30am - 12:00pm

Session 2. 3:00 pm - 4:30 pm

Session 3. 7:30 pm - 9:00pm

To enter a session please click on the link below:
https://meet70503155.adobeconnect.com/kollias_phd_equipment_check/

If the above sessions are inconvenient, please contact me at c.kollias@lancaster.ac.uk so that we can arrange another time that will be convenient for the both of us.

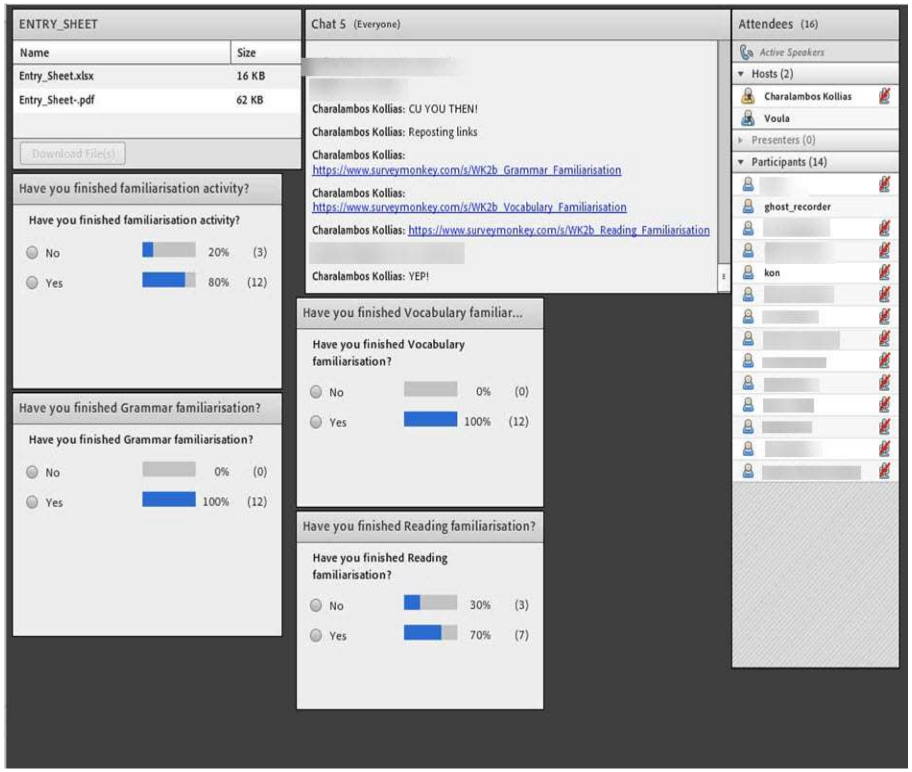
See you online!

Thank you.

Harry

P.S. If you took part in the session last night, you do not need to take part in the sessions again. But should you want to log in into one - feel free to do so!

panellist engagement



ENTRY_SHEET

Name	Size
Entry_Sheet.xlsx	16 KB
Entry_Sheet-.pdf	62 KB

Have you finished familiarisation activity?

Have you finished familiarisation activity?

○ No	20%	(3)
○ Yes	80%	(12)

Have you finished Grammar familiarisation?

Have you finished Grammar familiarisation?

○ No	0%	(0)
○ Yes	100%	(12)

Chat 5 (Everyone)

[redacted]

Charalambos Kollias: CU YOU THEN!

Charalambos Kollias: Reposting links

Charalambos Kollias: https://www.surveymonkey.com/s/WK2b_Grammar_Familiarisation

Charalambos Kollias: https://www.surveymonkey.com/s/WK2b_Vocabulary_Familiarisation

Charalambos Kollias: https://www.surveymonkey.com/s/WK2b_Reading_Familiarisation

[redacted]

Charalambos Kollias: YEP!

Have you finished Vocabulary familiar...

Have you finished Vocabulary familiarisation?

○ No	0%	(0)
○ Yes	100%	(12)

Have you finished Reading familiarisation?

Have you finished Reading familiarisation?

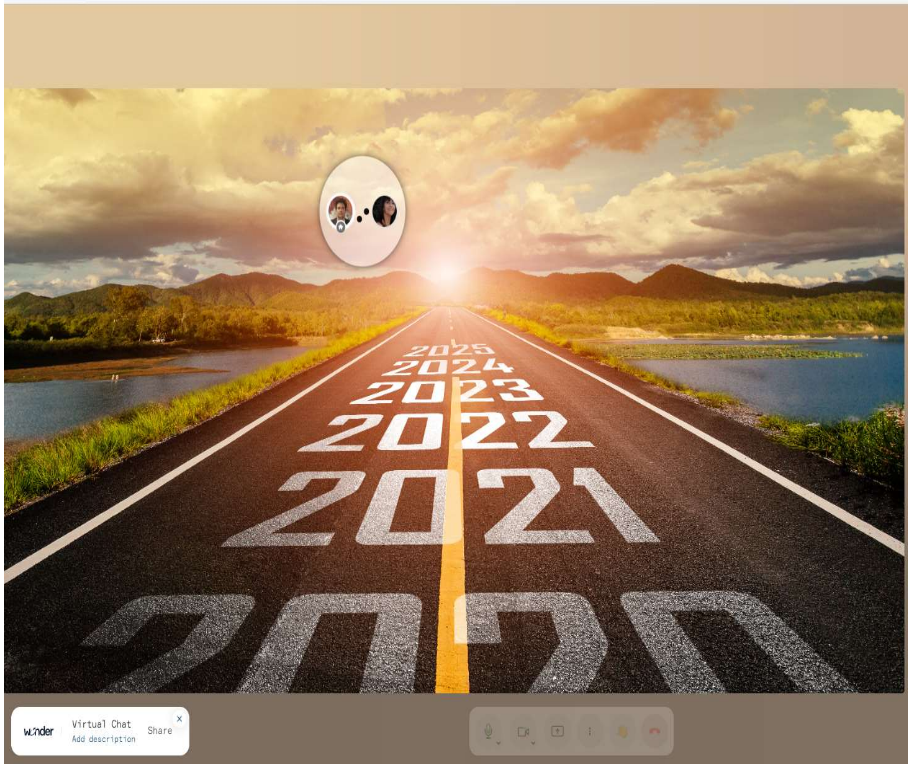
○ No	30%	(3)
○ Yes	70%	(7)

Attendees (16)

Active Speakers

- Hosts (2)
  - Charalambos Kollias
  - Voula
- Presenters (0)
- Participants (14)
  - [redacted]
  - ghost_recorder
  - [redacted]
  - [redacted]
  - kon
  - [redacted]
  - [redacted]
  - [redacted]
  - [redacted]
  - [redacted]
  - [redacted]
  - [redacted]
  - [redacted]
  - [redacted]

chat area


2025
2024
2023
2022
2021
2020
wonder
Virtual Chat
Add description
Share

SMALL
TALK

FRAME
WORK

virtual standard setting platform framework

Stage	Description of stage	Judge medium	Judge platform	Facilitator medium	Facilitator platform
Orientation	introductions	video		video	
	familiarisation activities	audio	speakers & mic muted (video paused)		microphone muted
	feedback on activities	video			
Training in the method	method training	video		video	
	training items discussion				

virtual standard setting platform framework cont.

Stage	Description of stage	Judge medium	Judge platform	Facilitator medium	Facilitator platform
Round 1	Round 1 ratings	audio	speakers & mic muted (video paused)	video	mic muted
	Round 1 feedback/ discussion	video			
Round 2	Round 2 ratings	audio	speakers & mic muted (video paused)	Video	mic muted
	Round 2 feedback/ discussion	video			

virtual standard setting platform framework cont.

Stage	Description of stage	Judge medium	Judge platform	Facilitator medium	Facilitator platform
Round 3 (when applicable)	Round 3 Ratings	audio	speakers & mic muted (video paused)	video	mic muted
	Round 3 feedback/ discussion	video			
Wrap-up	Video				

references

Corbin, J., & Strauss, A. (2015). Basics of qualitative research: Techniques and procedures for developing grounded theory. California: Sage Publications, Inc.

Dunlea, J., & Figueras, N. (2012). Replicating results from a CEFR test comparison project across continents. In D. Tsagari, & C. Ildikó (Eds.), Collaboration in language testing and assessment (Vol. 26, pp. 31-45). Frankfurt: Peter Lang.

Glaser, B. G. (1965). The constant comparative method of qualitative analysis. Social Problems, 12(4), 436-445. Retrieved from http://www.jstor.org/stable/798843

Glaser, B., & Strauss, A. (1967). The discovery of grounded theory. Chicago: Aldine.

Harvey, A. L., & Way, W. D. (1999). A comparison of web-based standard setting and monitored standard setting. Montreal: Paper presented at the annual meeting of the National Council of Measurement in Education.

Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 433-470). Westport: Praeger Publishers.

Kaftandjieva, F. (2004). : Standard setting. In S. Takala (Ed.), Reference Supplement to the Preliminary Pilot version of the Manual for Relating Language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (Section B). Strasbourg: Council of Europe/Language Policy Division.

references cont.

Katz, I. R., & Tannenbaum, R. J. (2014). Comparison of web-based and face-to-face standard setting using the Angoff method. Journal of Applied Testing Technology, 15(1), 1-17. Retrieved from www.jattjournal.com/index.php/atp/article/view/52751.

Katz, I. R., Tannenbaum, R. J., & Kannan, P. (2009). Virtual standard setting. Clear Exam Review, 20(2), 19-27.

Kock, N. (2005). Media richness or media naturalness? The evolution of our biological communication apparatus and its influence on our behavior toward e-communication tools. IEEE Transactions on Professional Communication, 48(2), 117-130. doi:10.1109/TPC.2005.849649

Kock, N. (2010). Evolutionary psychology and information systems theorizing. In N. Kock (Ed.), Evolutionary psychology and information systems research (pp. 3-38). New York: Springer.

Kollias, C. (forthcoming).Virtual Standard Setting: Setting cut scores in synchronous audio and video media. Manuscript in production: Peter Lang.

Kollias, C. (2022, April). 'Virtual Standard Setting: Which medium do judges prefer during each stage'. Paper presented in the 2022 annual meeting of the National Council of Measurement in Education (NCME), Virtual.

Kollias, C. (2021a, June). 'Virtual Standard Setting: Applying the Many-facet Rasch Measurement (MFRM) Model'. Paper presented in the 2021 annual meeting of the National Council of Measurement in Education (NCME), Virtual.

references cont.

Kollias, C. (2021b, June). ‘Variations in setting cut scores: How comparable are cut scores across media, methods, time, and instrument length?’ Paper presented in the 17th annual conference of the European Association for Language Testing and Assessment (EALTA), Online.

Kollias, C. (2021a, February). ‘Evaluating virtual standard setting workshops through a Many-facet Rasch measurement (MFRM) framework’. Paper presented in the International Objective Measurement Workshop (IOMW) 2020 Virtual Conference.

Kollias, C. (2019). ‘Using the CEFR Rasch and/or IRT resources: The benefits, the challenges, and the ellipsis’. Paper presented in the Common European Framework of Reference (CEFR) Special Interest Group (SIG) of the European Association for Language Testing and Assessment (EALTA), Dublin.

Kollias, C. (2018, May). ‘Virtual standard setting in language testing: Exploring the use of synchronous audio and audio-visual environment’. Paper presented in the 15th annual conference of the European Association for Language Testing and Assessment (EALTA), Bochum, Germany.

Kollias, C. (2017, February). ‘Conducting a synchronous virtual standard setting workshop to set a CEFR cut score’. Paper presented in the Common European Framework of Reference (CEFR) Special Interest Group (SIG) of the European Association for Language Testing and Assessment (EALTA), Kaplan International College, London, UK.

references cont.

Kollias, C. & Kanistra, P. (2019, March). ‘Setting cut scores and evaluating standard setting judgments through the Many-Facet Rasch Measurement (MFRM) model”. Paper presented in the 13th Annual UK Rasch User Group Meeting, Cambridge, UK.

Tannenbaum, R. J. (2013). Setting standards on the TOEIC(R) listening and reading test and the TOEIC(R) speaking and writing tests: A recommended procedure. In The research foundation for the TOEIC tests: A compendium of studies (Vol. II, pp. 8. 1-8.12). Princeton: Educational Testing Service.

disclaimer

The opinions expressed in this webinar are those of the presenter (Charalambos Kollias). They do not purport to reflect the *opinions* or *views* of the NFER or its members.

Thank You

Any questions?

1946–2021 75 YEARS

NFER

National Foundation for Educational Research

Evidence for excellence in education