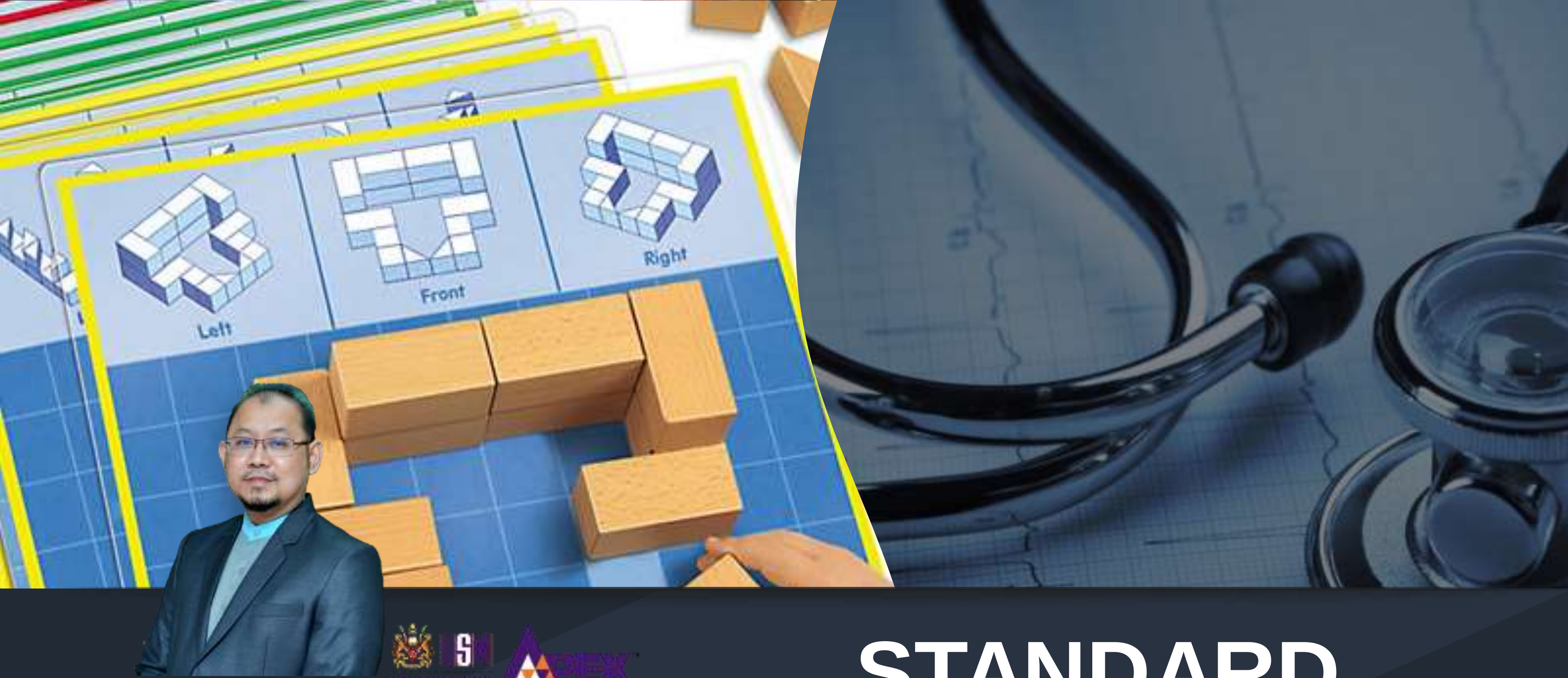


STANDARD SETTING

MUHAMAD SAIFUL BAHRI YUSOFF

Department of Medical Education, School of Medical Sciences, Universiti Sains Malaysia, email: msaiful_bahri@usm.my.

https://www.researchgate.net/profile/Muhamad_Saiful_Bahri_Yusoff

FOCUS

#1 OVERVIEW

An overview on standard setting – what, why, when and how?

#2 PROCESS

Elaboration on the process of different standard setting methods

#3 PRACTICE

Hands-on experience on performing standard setting methods

#4 INSIGHTS

Gaining new insights on appropriate method to set standard for different examination formats

FOCUS

#1 OVERVIEW

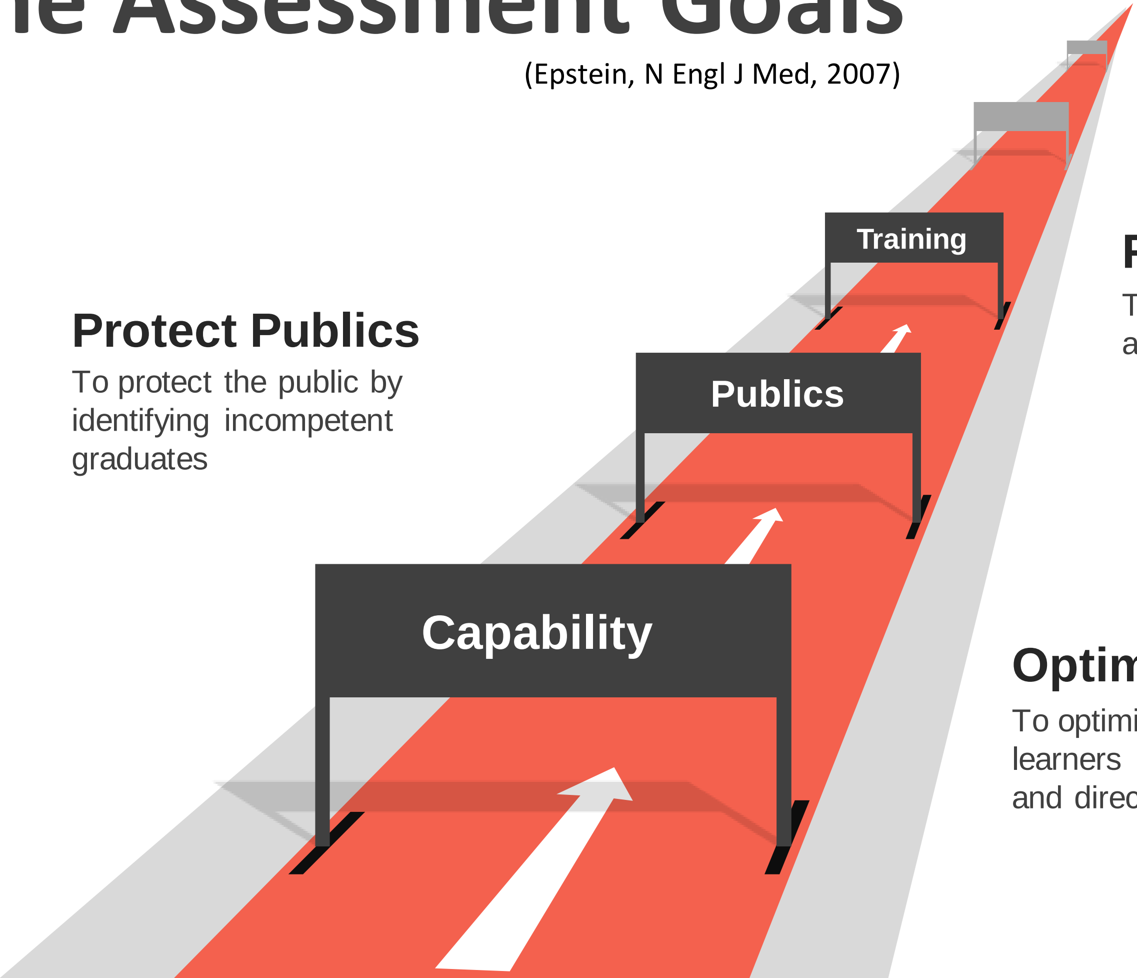
An overview on standard setting – what, why, when and how?

#2 PROCESS

Elaboration on the process of different standard setting methods

The Assessment Goals

(Epstein, N Engl J Med, 2007)



Standard Setting

Standard setting methods are part of the assessment process

(Pearson et al., 2009)

Further Training

To provide a basis for choosing applicants for advanced training

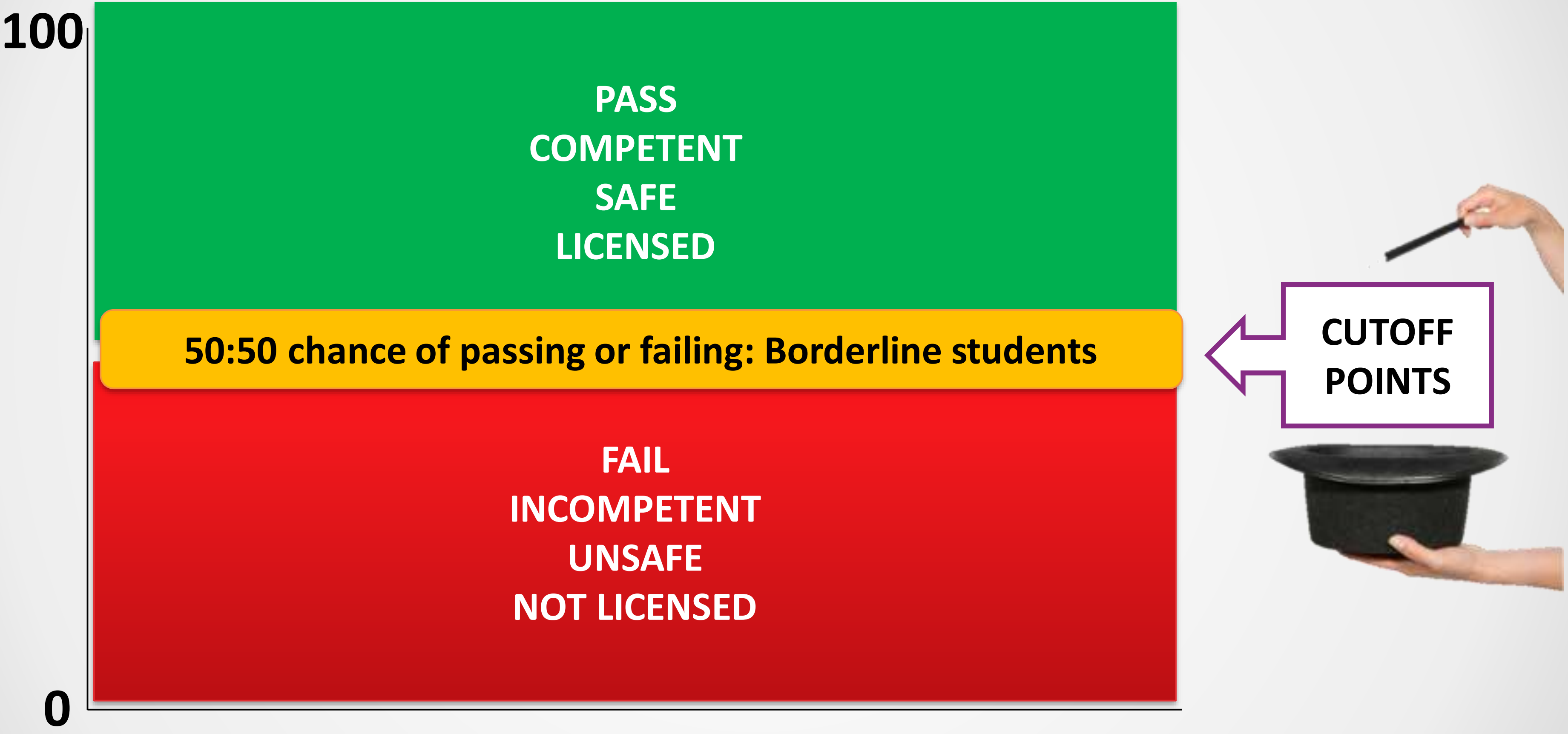
Protect Publics

To protect the public by identifying incompetent graduates

Optimize capabilities

To optimize the capabilities of all learners by providing motivation and direction for future learning

STANDARD SETTING: Overview



100

PASS

COMPETENT

SAFE

LICENSED

50:50 chance of passing or failing: Borderline students

FAIL

INCOMPETENT

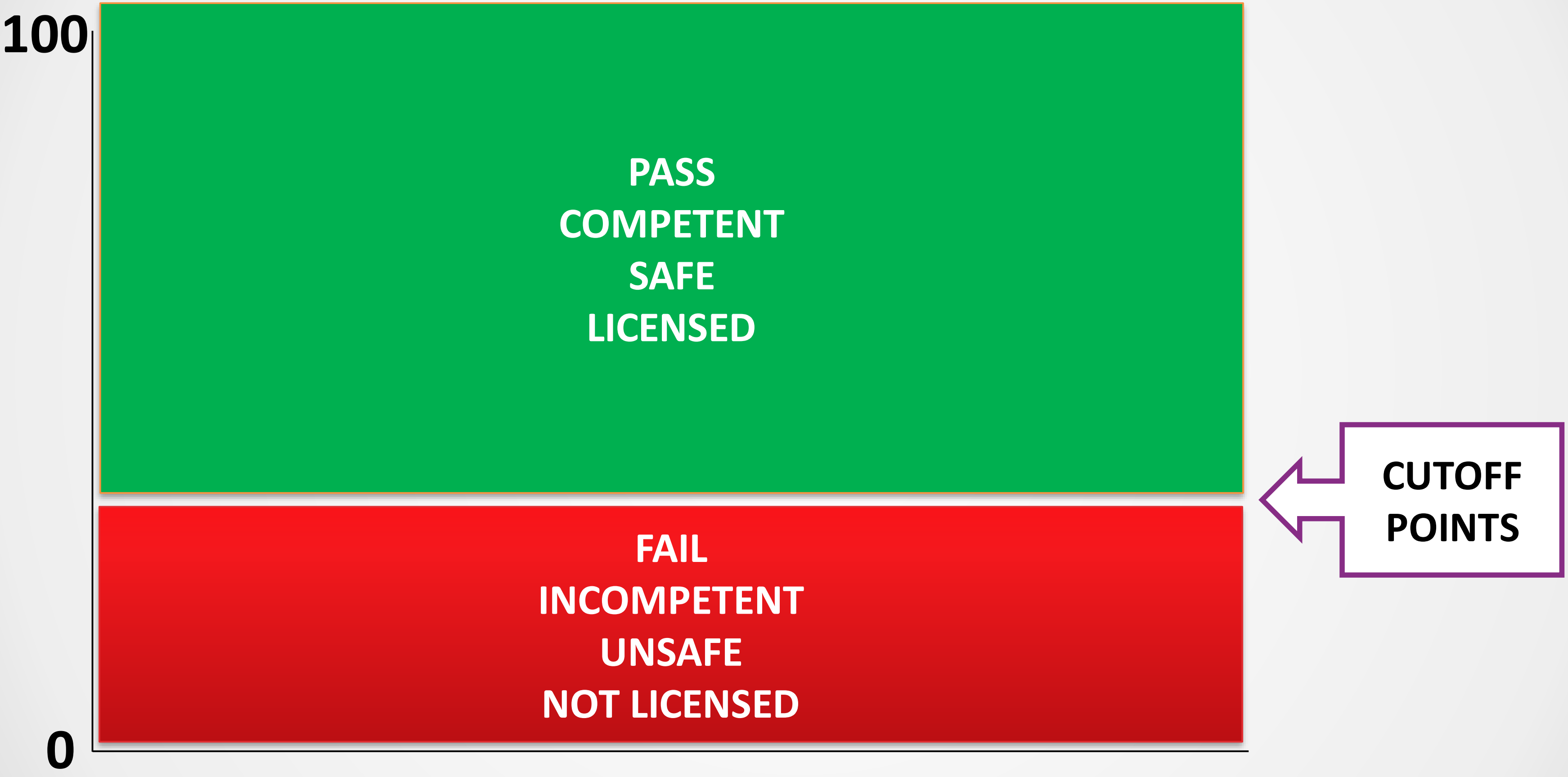
UNSAFE

NOT LICENSED

0

CUTOFF POINTS

STANDARD SETTING: Difficult assessment?

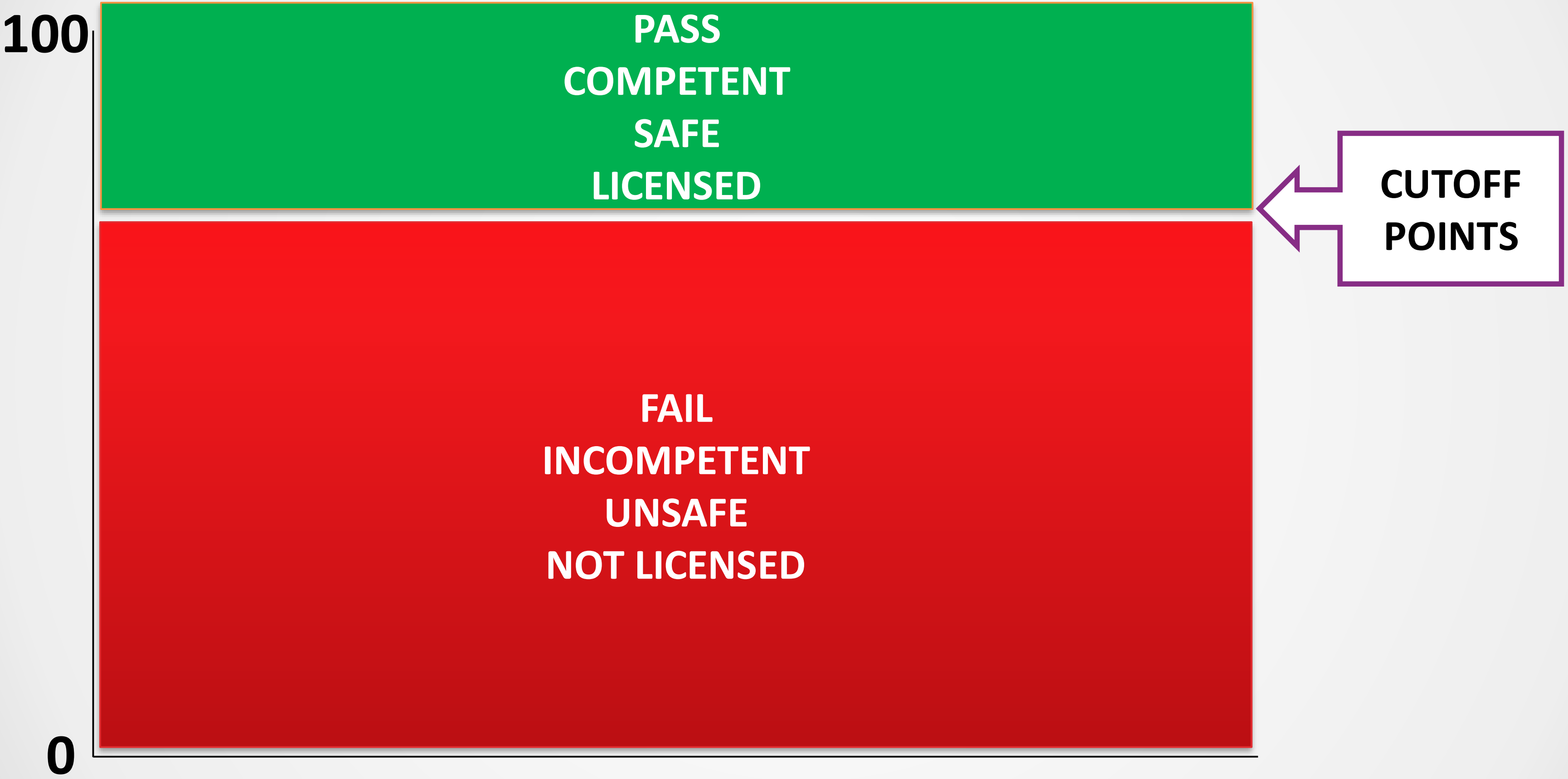


PASS
COMPETENT
SAFE
LICENSED

FAIL
INCOMPETENT
UNSAFE
NOT LICENSED

CUTOFF
POINTS

STANDARD SETTING: Easy assessment?



PASS
COMPETENT
SAFE
LICENSED

FAIL
INCOMPETENT
UNSAFE
NOT LICENSED

CUTOFF
POINTS

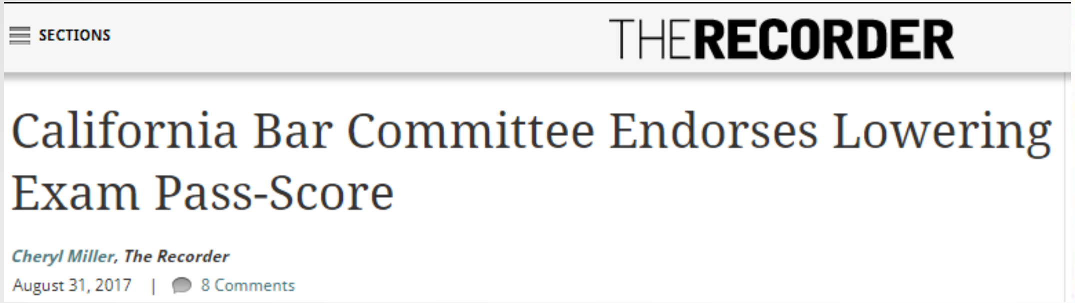
STANDARD SETTING: Definition

> "The proper following of a prescribed, rational system of rules or procedures resulting in the assignment of a number to differentiate between two or more states or degrees of performance"
>
> (Cizek, 1993)

COMPETENCE → PASSING SCORE

(Kane, 1994; Norcini, 1994)

STANDARD SETTING: An Accountability



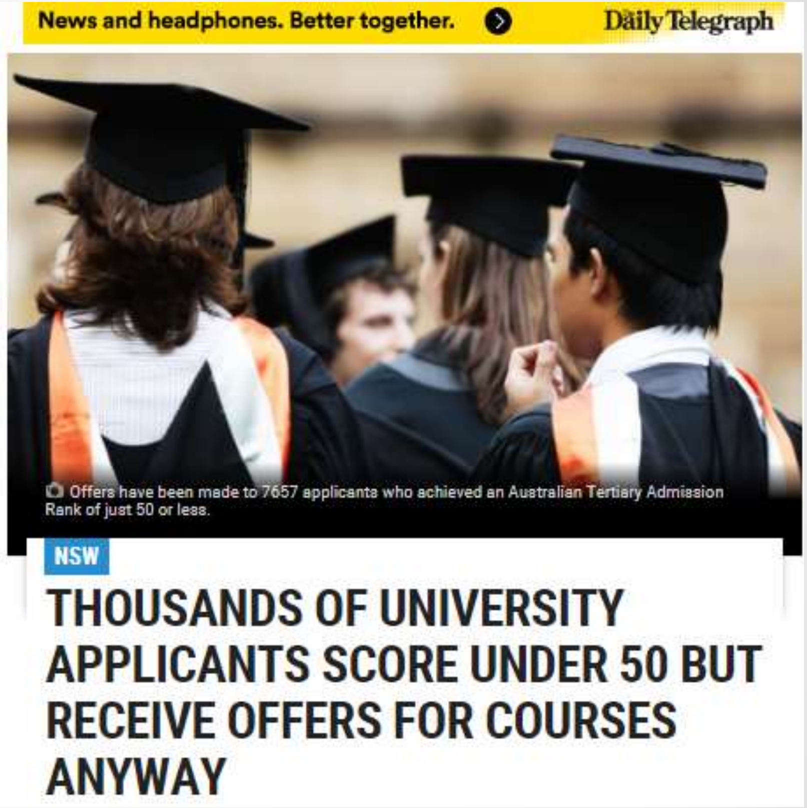
SECTIONS

THERECORDER

California Bar Committee Endorses Lowering Exam Pass-Score

Cheryl Miller, The Recorder

August 31, 2017 | 8 Comments



Daily Telegraph

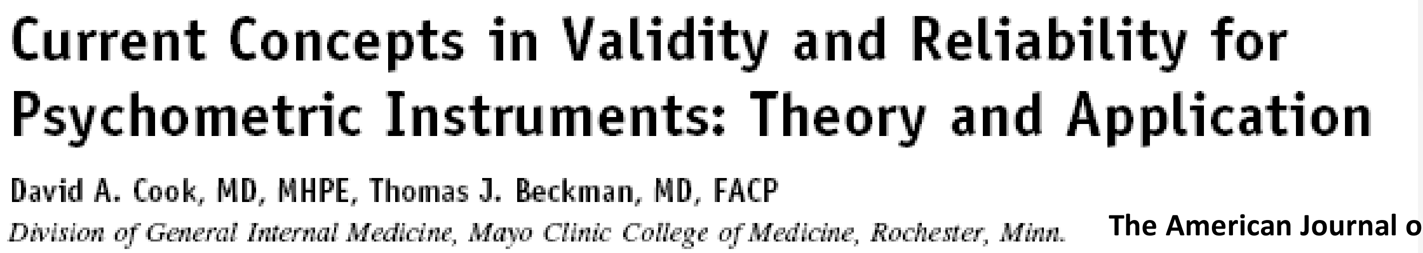
Offers have been made to 7657 applicants who achieved an Australian Tertiary Admission Rank of just 50 or less.

NSW

THOUSANDS OF UNIVERSITY APPLICANTS SCORE UNDER 50 BUT RECEIVE OFFERS FOR COURSES ANYWAY

"The question of what the appropriate cut score should be has come into sharp focus, and intense debate, over the last year as the exam's pass rate has tumbled"

STANDARD SETTING: As validity evidence

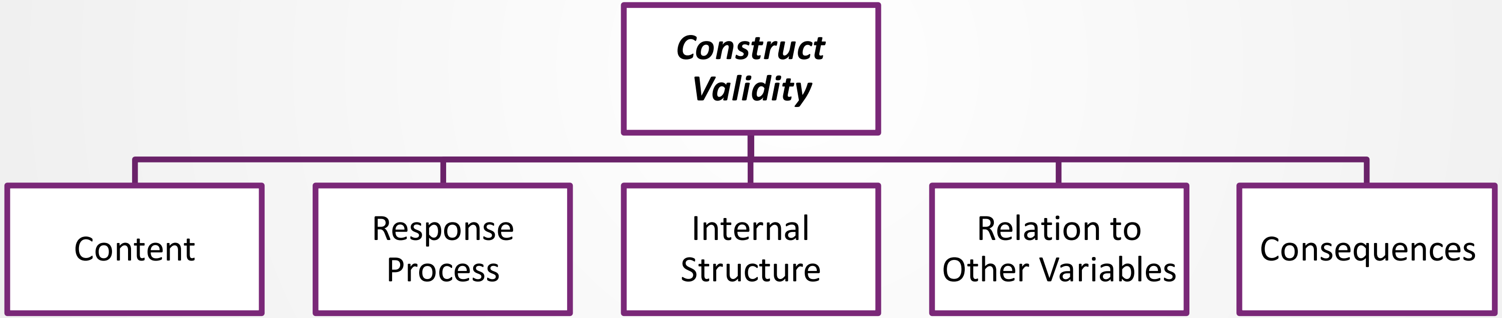


Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application

David A. Cook, MD, MHPE, Thomas J. Beckman, MD, FACP

Division of General Internal Medicine, Mayo Clinic College of Medicine, Rochester, Minn.

The American Journal of Medicine, 2006



Construct Validity

Content

Response Process

Internal Structure

Relation to Other Variables

Consequences

Methods of determining exam pass/fail score

STANDARD SETTING: Standards

Statement about whether the examination performance fit for a particular purpose.

Based on judgement on candidate performances against education constructs.

Examples

i. **Ready for graduation**
ii. **Competent to move to practical years**

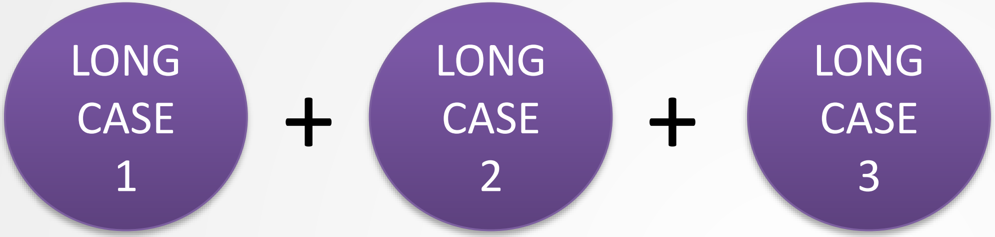
"The graduate of this medical program should demonstrate **adequate knowledge for safe clinical decision and management**, be able to work with supervision, equipped with **standard clinical skills**, and **conduct themselves professionally**."

(USM Medical School Pro II Standard)

STANDARD SETTING: Types of standard

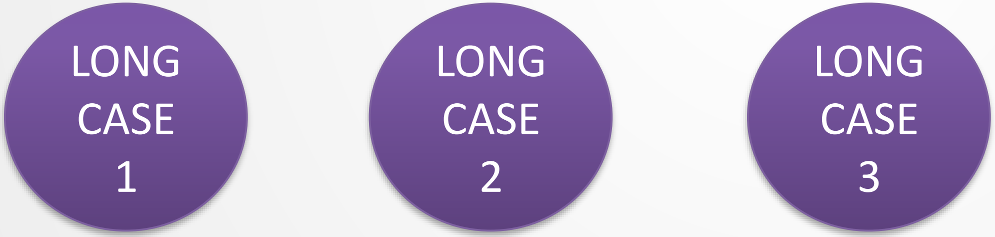
Relative	Absolute	Compromise
Norm-referenced	Criterion-referenced	Combine both
“Top 60% will pass”	“Candidate who gets more than 60% pass”	
‘Limited seats’ - Admission	High stakes examination	

STANDARD SETTING: Strategies



LONG CASE 1 + LONG CASE 2 + LONG CASE 3 = more than 50%

COMPENSATORY



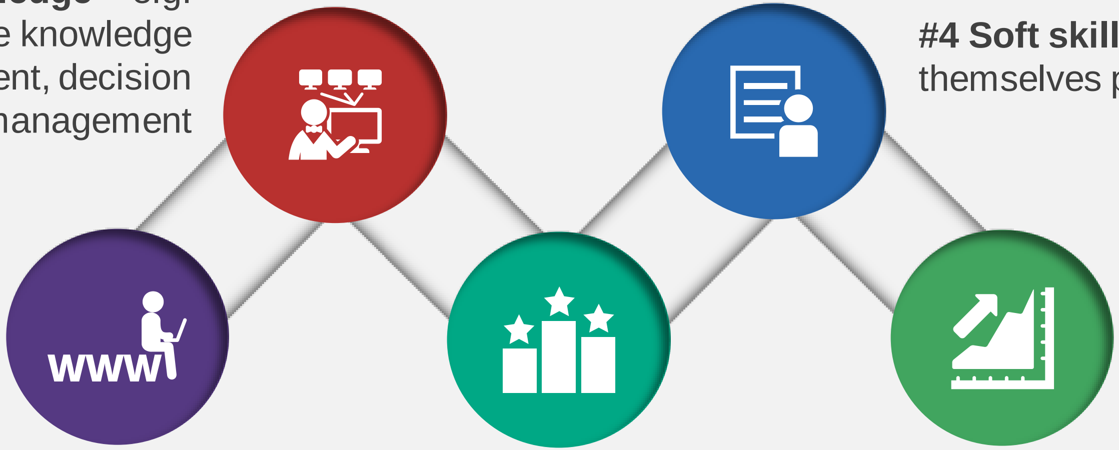
LONG CASE 1 LONG CASE 2 LONG CASE 3 = must pass separately

CONJUNCTIVE

STANDARD SETTING: Method Selection

	Absolute	**Compromise**
Test item based 	**Angoff families** **Ebel** **Nedelsky** **Bookmark**	Cohen
Test examinees based 	**Borderline group/** **Borderline regression** Contrasting group	Hofstee

STANDARD SETTING: Guides to define Borderline



#1 Setting – e.g. graduate of the ophthalmology program

# 2 Knowledge – e.g. demonstrate adequate knowledge for safe clinical judgment, decision making and management

#3 Skills – e.g. be able to work with moderate supervision, equipped with acceptable technical ability

#4 Soft skills - e.g. conduct themselves professionally

#5 Errors (considering the forgivable or unforgivable) – e.g. safe clinical judgment, decision making and management

(Mills, Melican & Ahluwalia, 1991)

STANDARD SETTING: Borderline Standards

Setting

Errors
Forgivable, non-forgivable

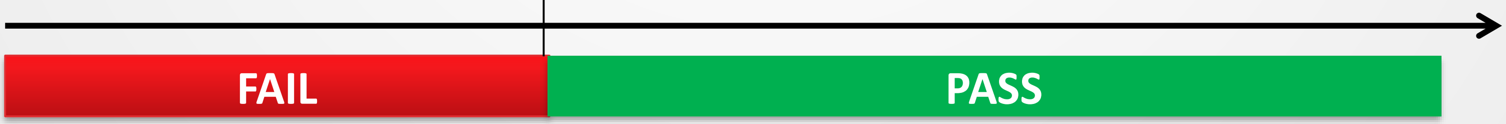
“The borderline **graduate of undergraduate medical program** should demonstrate **adequate fundamental knowledge for safe clinical judgment** and **decision making**, be able to **work under supervision**, competent in **basic clinical skills**, and **conduct themselves professionally**.”

Knowledge

Skills

Attitude

(UNIMAS, 17 April 2018, Standard Setting Workshop)



STANDARD SETTING: Borderline Standards

Setting

Errors
Forgivable, non-forgivable

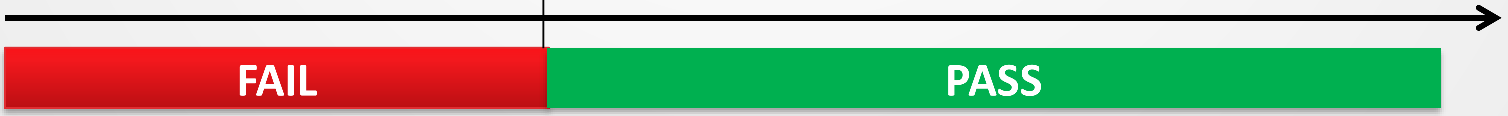
"The borderline **graduate of the ophthalmology program** should demonstrate **adequate knowledge for safe clinical judgment**, **decision making and management**, be able to **work with moderate supervision**, equipped with **acceptable technical ability**, and **conduct themselves professionally**."

Knowledge

Skills

Attitude

(MUCCO, 20-22 Aug 2014, A Workshop on Examination Questions Preparation, Kuala Lumpur)



STANDARD SETTING: Borderline Standards

Setting

Knowledge

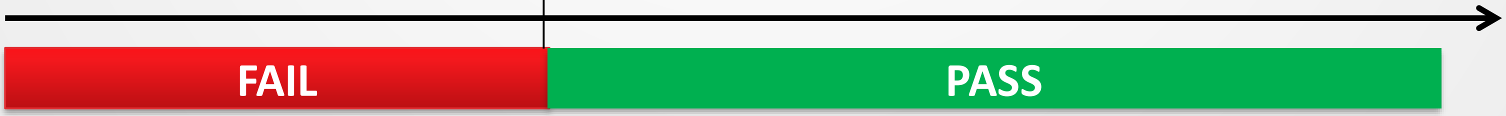
Errors
Forgivable, non-forgivable

"The borderline **graduate of the emergency medicine program** should demonstrate **adequate knowledge for safe clinical judgment**, **decision making and management**, be able to **work with moderate supervision**, equipped with **acceptable life saving skills and technical ability**, and **conduct themselves professionally**."

Skills

Attitude

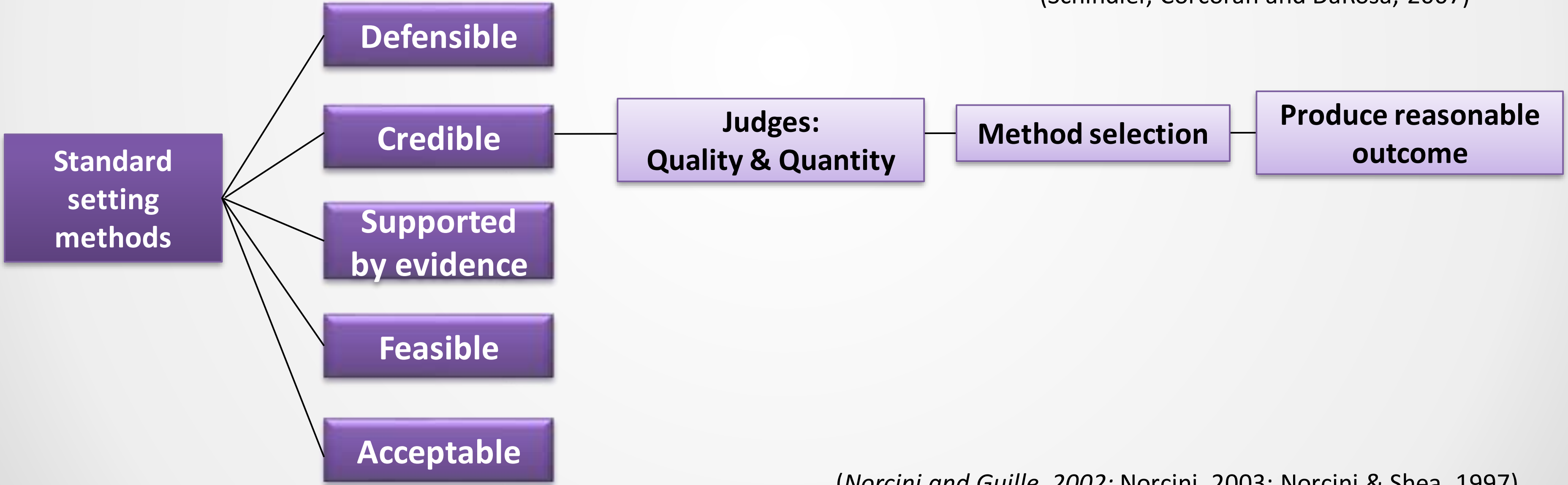
(SCCEM, 10 Nov 2018, A Workshop on Standard Setting A & E Workshop, UM, Kuala Lumpur)



Standard is arbitrary.

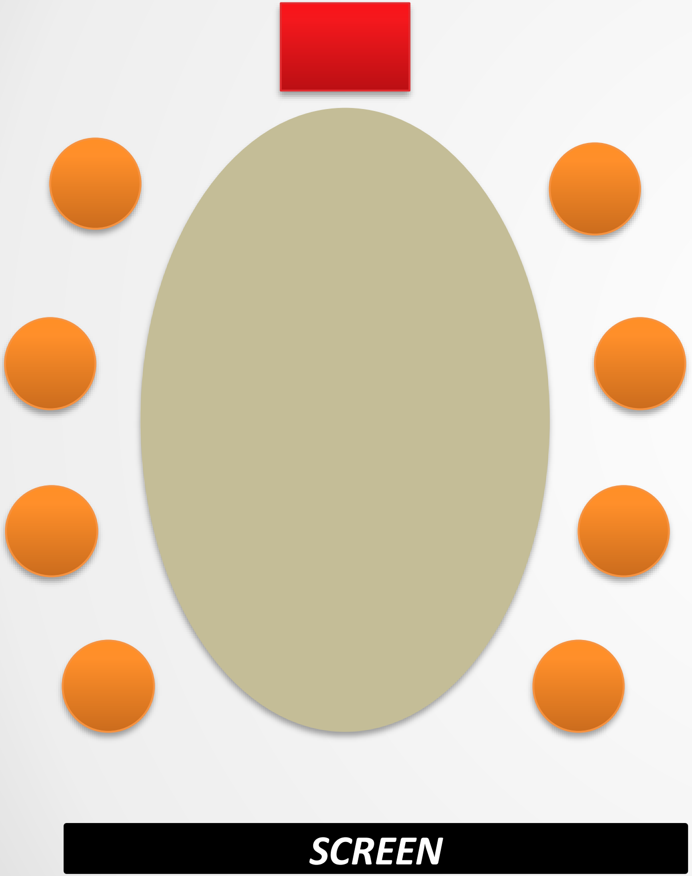
"…… even the most rigorous standard-setting method, followed meticulously, will be somewhat arbitrary however, they should be **credible**."

(Schindler, Corcoran and DaRosa, 2007)



(*Norcini and Guille, 2002;* Norcini, 2003; Norcini & Shea, 1997)

STANDARD SETTING: Judges



Subject matter experts

Know target population

Understand task and assessment tool

Fair-minded

Willing to follow directions

Give full attention to the process

Demographically diverse to avoid bias

6 considered minimum

(Norcini and Guille, 2002)

STANDARD SETTING: Method Selection

	MCQs	Essays	Performance based	Portfolios
Angoff family	🟩	🟩	🟩	
Ebel	🟩	🟩	🟩	
Nedelsky	🟩			
Bookmark	🟩	🟩		
Borderline group/ regression			🟩	
Contrasting group			🟩	
Hofstee	🟩	🟩	🟩	🟩

STANDARD SETTING: Post-Setting

HISTORICAL STANDARDS

Compare with historical standards or external measure

PASS RATES

The pass rates have reasonable relationships with other markers of competence

STAKEHOLDERS

Consider stakeholder opinion and the results related to future performance

(Norcini & Shea, 1997; Norcini, 2003)

FOCUS

#1 OVERVIEW

An overview on standard setting – what, why, when and how?

#2 PROCESS

Elaboration on the process of different standard setting methods

#3 PRACTICE

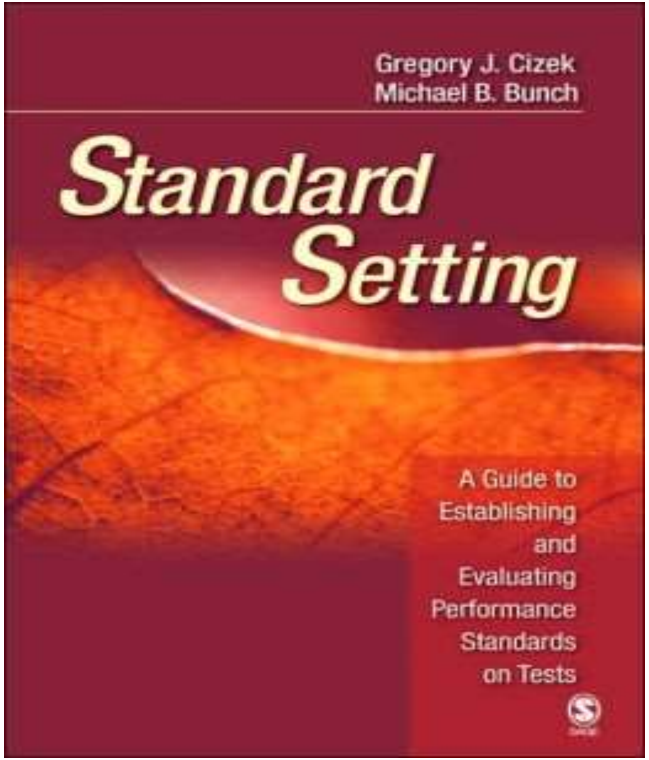
Hands-on experience on performing standard setting methods

#4 INSIGHTS

Gaining new insights on appropriate method to set standard for different examination formats

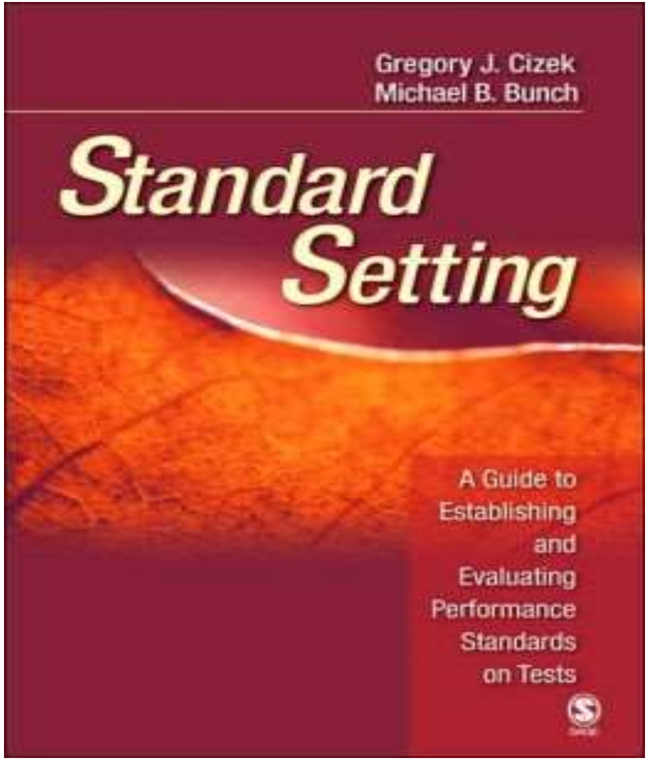
Methods for Setting Standard

- Absolute methods: Test-Items
  - – Angoff (Angoff, 1971)
  - – Ebel (Ebel, 1972)
  - – Nedelsky (Nedelsky, 1954)
- Absolute methods: Test-Takers
  - – Borderline (Livingston & Zieky, 1982)
  - – Contrasting groups (Berk, 1976)
- Compromise methods
  - – Hofstee (Hofstee,1983)


Gregory J. Cizek
Michael B. Bunch
Standard Setting
A Guide to Establishing and Evaluating Performance Standards on Tests

Methods for Setting Standard

- Absolute methods: Test-Items
  - Angoff (Angoff, 1971)
  - Ebel (Ebel, 1972)
  - Nedelsky (Nedelsky, 1954)
- Absolute methods: Test-Takers
  - Borderline (Livingston & Zieky, 1982)
  - Contrasting groups (Berk, 1976)
- Compromise methods
  - Hofstee (Hofstee,1983)


Gregory J. Cizek
Michael B. Bunch
Standard Setting
A Guide to Establishing and Evaluating Performance Standards on Tests

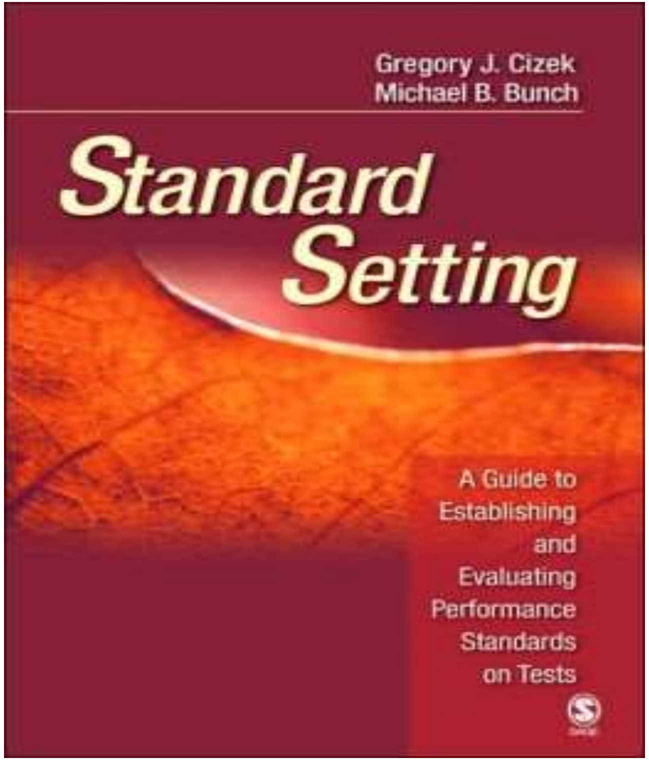
Standard Setting in Action

ABSOLUTE METHODS – TEST-ITEM

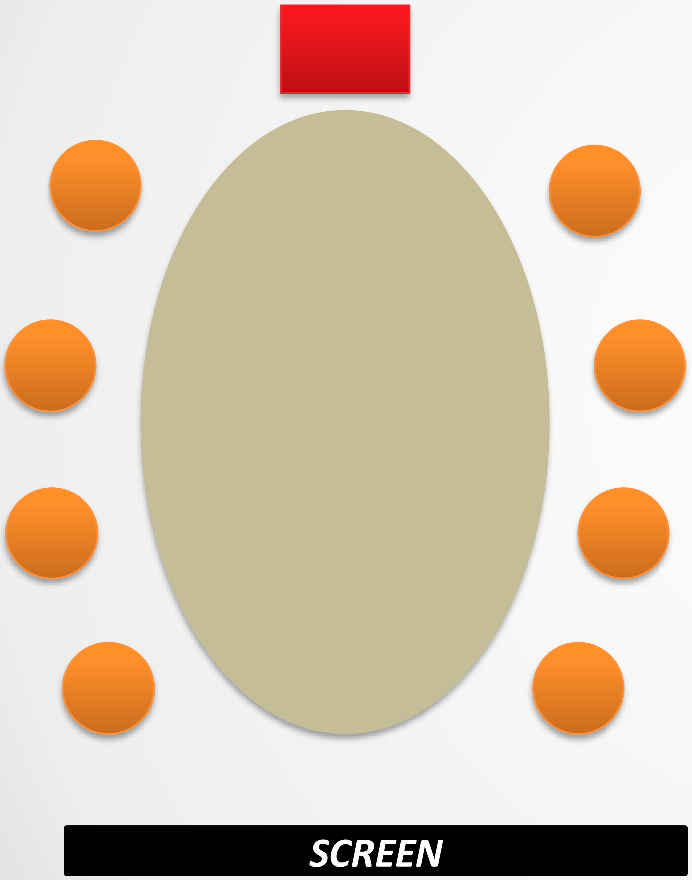
- Process
- Practice
- Insights

Methods for Setting Standard

- Absolute methods: Test-Items
  - Angoff (Angoff, 1971)
  - Ebel (Ebel, 1972)
  - Nedelsky (Nedelsky, 1954)
- Absolute methods: Test-Takers
  - Borderline (Livingston & Zieky, 1982)
  - Contrasting groups (Berk, 1976)
- Compromise methods
  - Hofstee (Hofstee,1983)

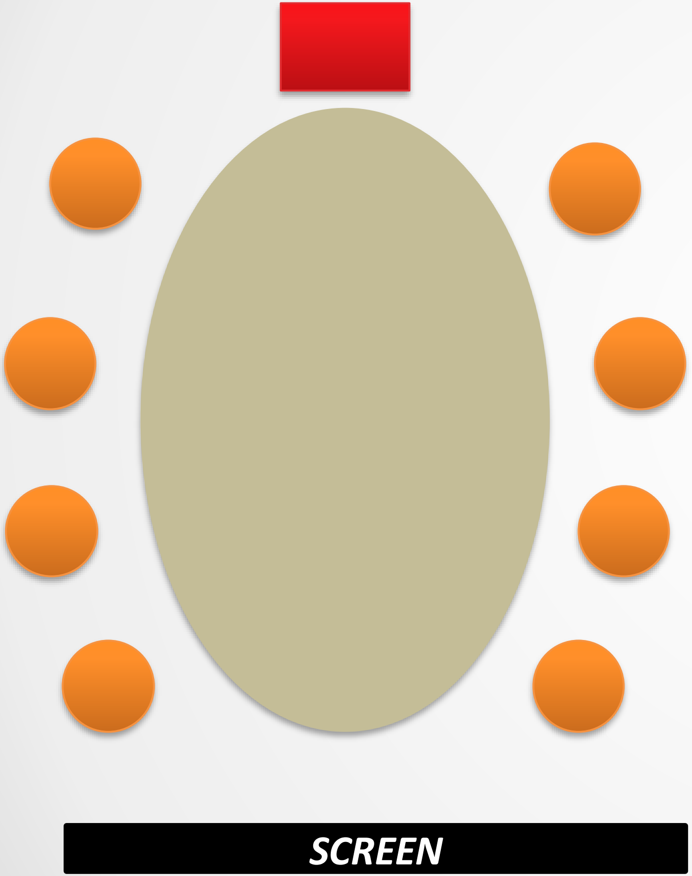

Gregory J. Cizek
Michael B. Bunch
Standard Setting
A Guide to Establishing and Evaluating Performance Standards on Tests

STANDARD SETTING: Angoff



- Angoff is the most common method used for setting standard.
- Types of Angoff:
  - Direct Angoff (Angoff, 1971)
  - Extended Angoff (Hambleton & Plake, 1995)
  - Modified Angoff (Cizek, 1996)
  - Three-level Angoff (Yudkowsky, Downing & Popescu, 2008).
- We treat them as Angoff's family

STANDARD SETTING: Angoff - PRE



Select the judges

Discuss

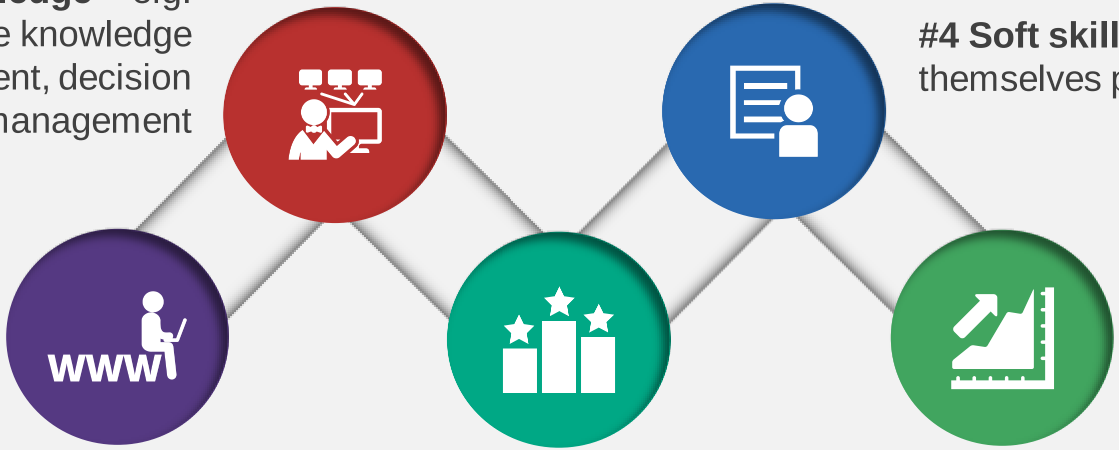
a. Purpose of the assessment
b. Nature of examinees
c. Components of adequate/inadequate knowledge

Select the methods – train judges

Define borderline standard

(Cizek, 2006; Angoff, 1971)

STANDARD SETTING: Define Borderline



#1 Setting – e.g. graduate of the ophthalmology program

# 2 Knowledge – e.g. demonstrate adequate knowledge for safe clinical judgment, decision making and management

#3 Skills – e.g. be able to work with moderate supervision, equipped with acceptable technical ability

#4 Soft skills - e.g. conduct themselves professionally

#5 Errors (considering the forgivable or unforgivable) – e.g. safe clinical judgment, decision making and management

(Mills, Melican & Ahluwalia, 1991)

STANDARD SETTING: Borderline Standards

Setting

Errors

Forgivable, non-forgivable

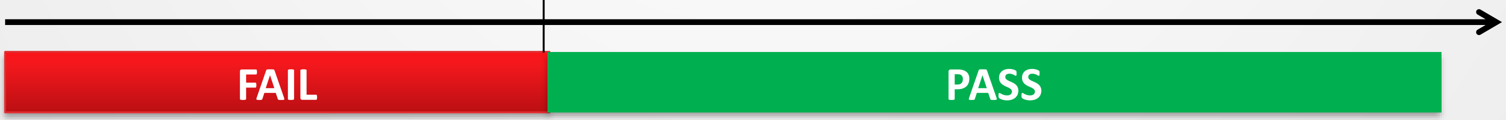
"The borderline **graduate of undergraduate medical program** should demonstrate **adequate fundamental knowledge for safe clinical judgment** and **decision making**, be able to **work under supervision**, competent in **basic clinical skills**, and **conduct themselves professionally**."

Knowledge

Skills

Attitude

(UNIMAS, 17 April 2018, Standard Setting Workshop)



FAIL

PASS

STANDARD SETTING: Borderline Standards

Setting

“The borderline **graduate of the ophthalmology program** should demonstrate **adequate knowledge for safe clinical judgment**, **decision making and management**, be able to **work with moderate supervision**, equipped with **acceptable technical ability**, and **conduct themselves professionally**.”

Knowledge

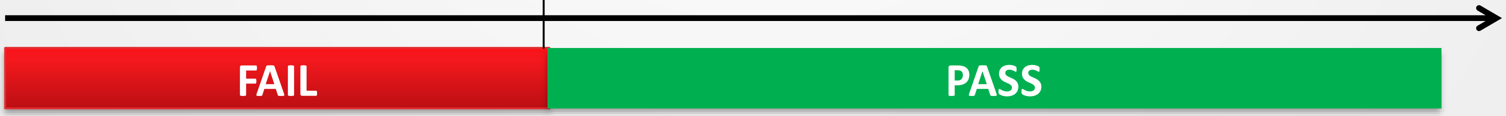
Skills

Attitude

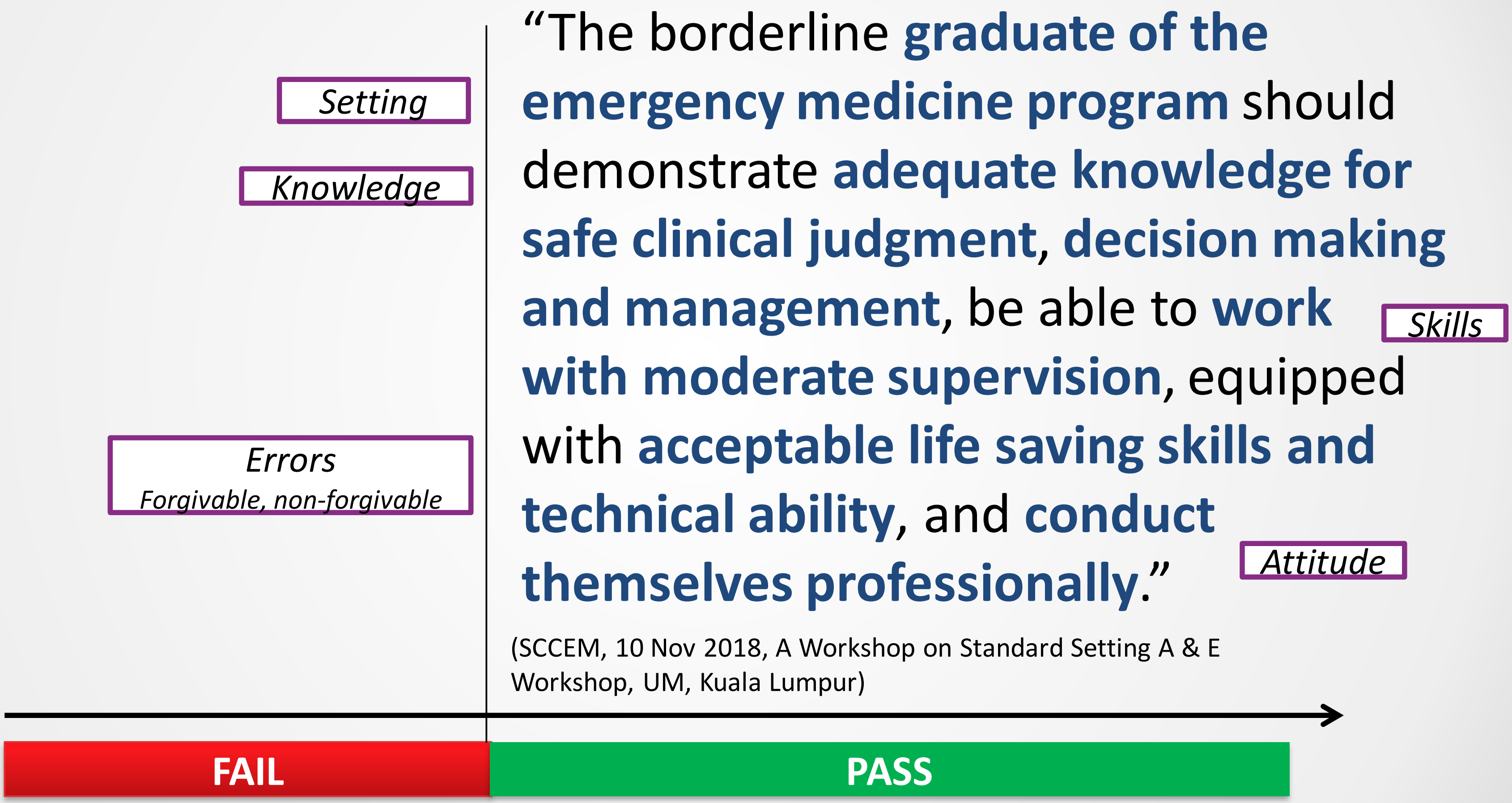
Errors

Forgivable, non-forgivable

(MUCCO, 20-22 Aug 2014, A Workshop on Examination Questions Preparation, Kuala Lumpur)



STANDARD SETTING: Borderline Standards



"The borderline **graduate of the emergency medicine program** should demonstrate **adequate knowledge for safe clinical judgment**, **decision making and management**, be able to **work with moderate supervision**, equipped with **acceptable life saving skills and technical ability**, and **conduct themselves professionally**."

(SCCEM, 10 Nov 2018, A Workshop on Standard Setting A & E Workshop, UM, Kuala Lumpur)

STANDARD SETTING: Borderline Standards

Setting

Knowledge

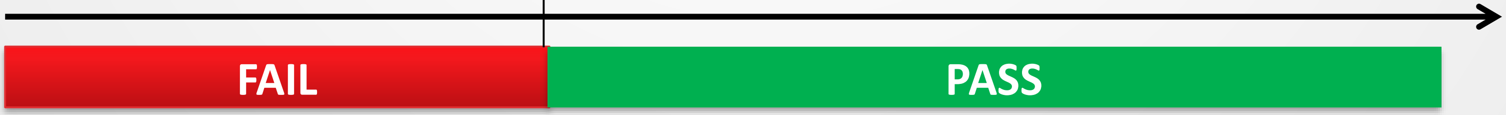
Errors
Forgivable, non-forgivable

"The borderline **graduate of the anaesthesiology program** should demonstrate **adequate knowledge for safe clinical judgment**, **decision making and management**, be able to **work with minimal supervision**, equipped with **acceptable life saving skills and technical ability**, and **conduct themselves professionally**."

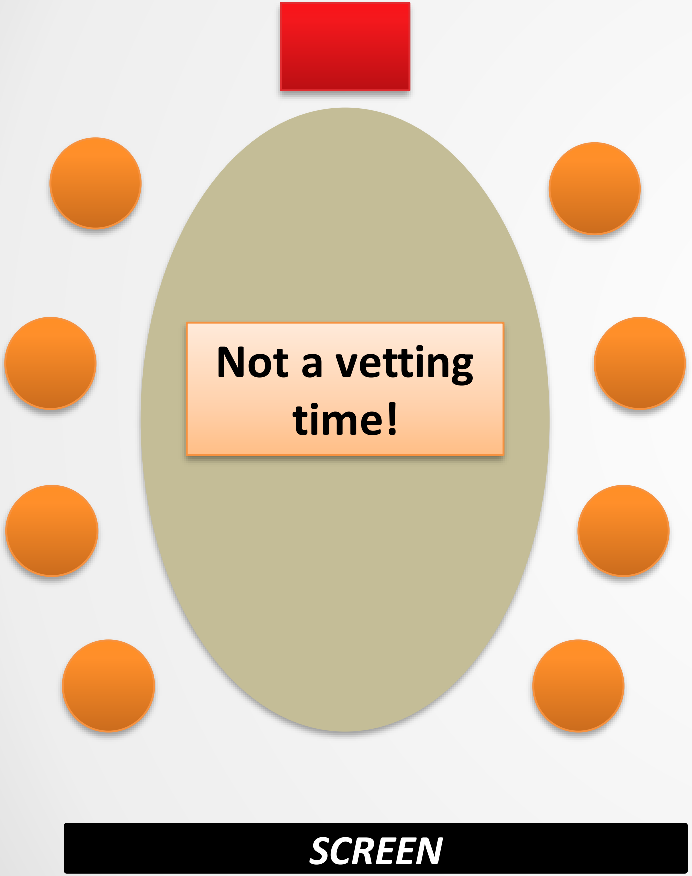
Skills

Attitude

(8 Jan 2022, A Workshop on Standard Setting (Anaesthesiology) Workshop, UPM, Selangor)



STANDARD SETTING: Angoff - DURING



Read through question 1

Judges: Individually, estimate proportion of borderline examinees will correctly answer question 1

Moderator: Record ratings

Moderator: Discuss ratings

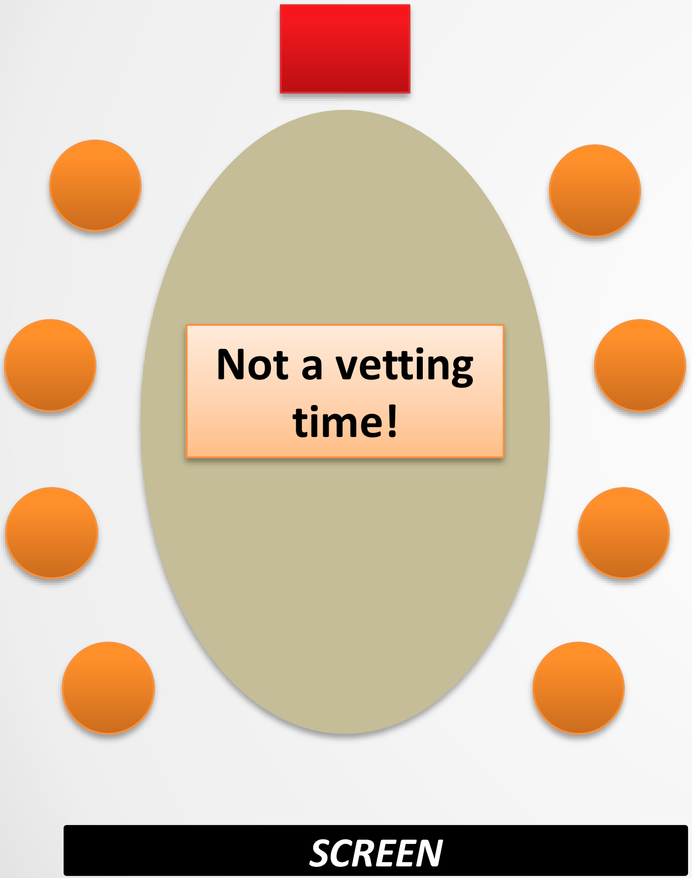
Moderator: Get 2nd ratings after discussion

Calculate mean

Repeat for next questions

(Cizek, 2006; Angoff, 1971)

STANDARD SETTING: Modified Angoff - DURING



Read through question 1

Judges: Individually, estimate the mark that can be obtained by borderline examinees for question 1

Moderator: Record ratings

Moderator: Discuss ratings

Moderator: Get 2nd ratings after discussion

Calculate mean

Repeat for next questions

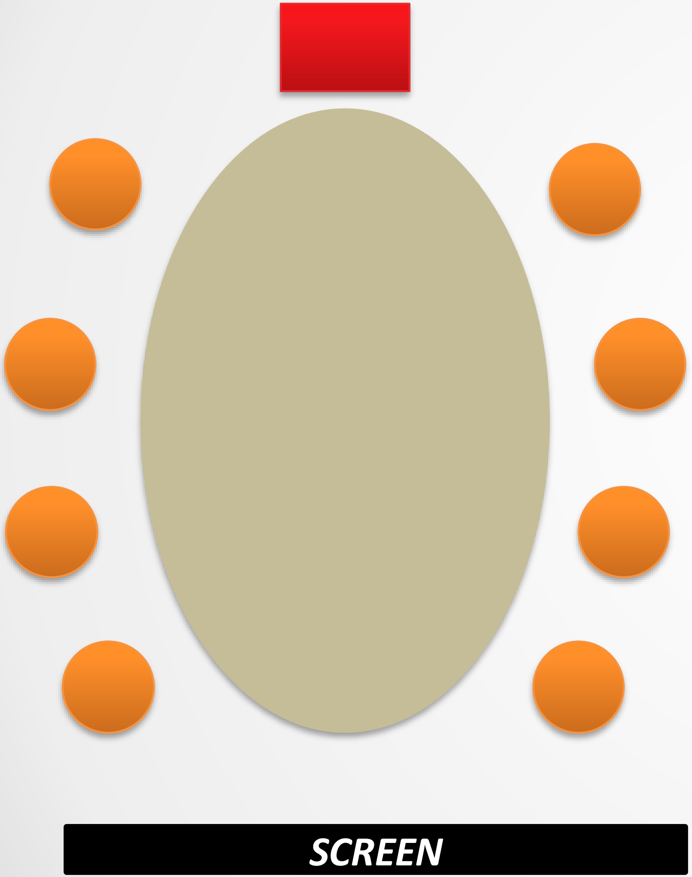
(Cizek, 2006; Angoff, 1971)

	Q1	Q2	Q3	Q4	Q5	Mean
JUDGE 1	60					
	60					
JUDGE 2	50					
	60					
JUDGE 3	**90**					
	60					
JUDGE 4	60					
	50					
JUDGE 5	60					
	60					
JUDGE 6	**40**					
	60					
Mean 1st	**60**					
Mean 2nd	**58.3**					

Cut-off score 1st round

Cut-off score 2nd round

STANDARD SETTING: Angoff - POST



Evaluate the process

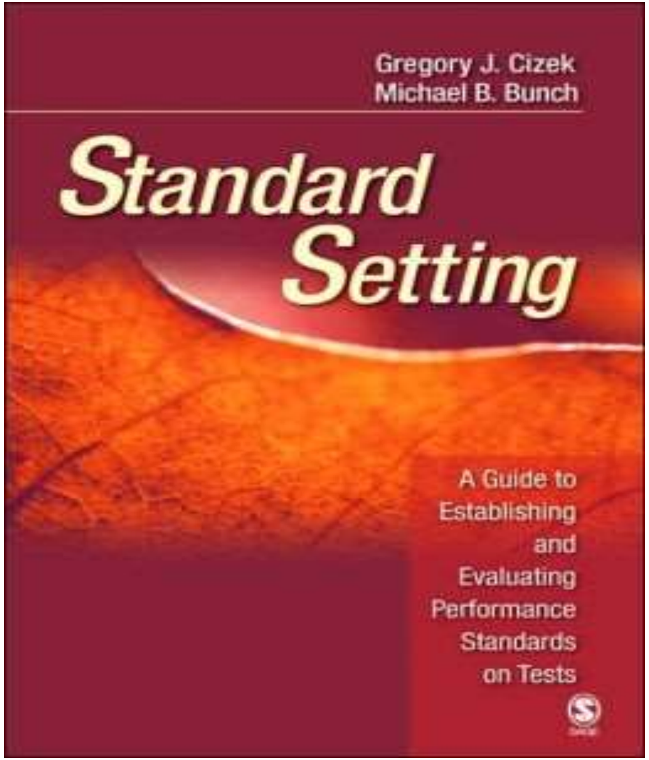
- Judges confidence in the process
- Resulting cut off scores

Documentation

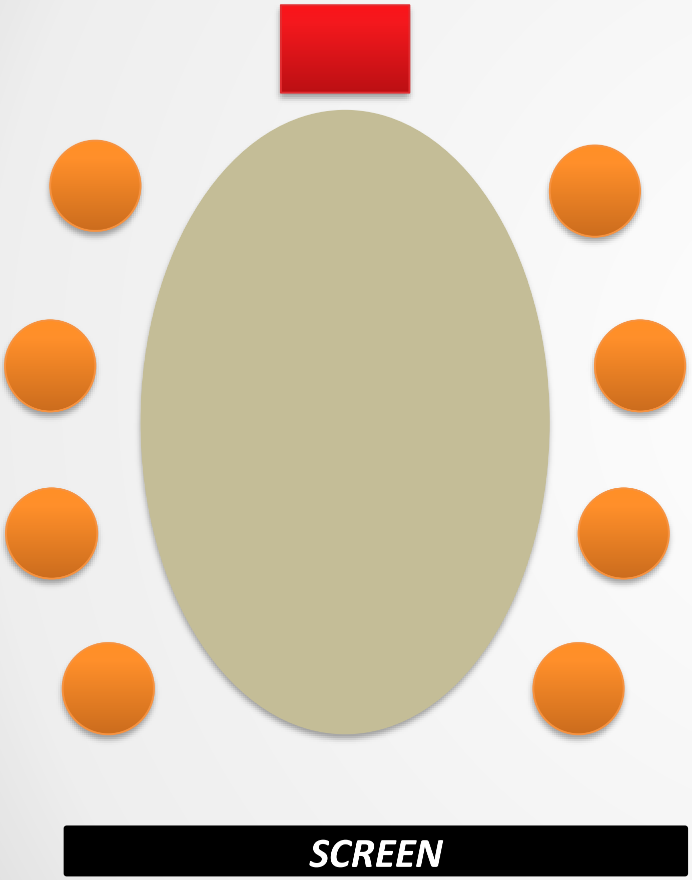
(Cizek, 2006; Angoff, 1971)

Methods for Setting Standard

- Absolute methods: Test-Items
  - Angoff (Angoff, 1971)
  - Ebel (Ebel, 1972)
  - Nedelsky (Nedelsky, 1954)
- Absolute methods: Test-Takers
  - Borderline (Livingston & Zieky, 1982)
  - Contrasting groups (Berk, 1976)
- Compromise methods
  - Hofstee (Hofstee,1983)


Gregory J. Cizek
Michael B. Bunch
Standard Setting
A Guide to Establishing and Evaluating Performance Standards on Tests

STANDARD SETTING: Ebel - PRE



Select the judges

Discuss

a. Purpose of the assessment

b. Nature of examinees

c. Components of adequate/inadequate knowledge

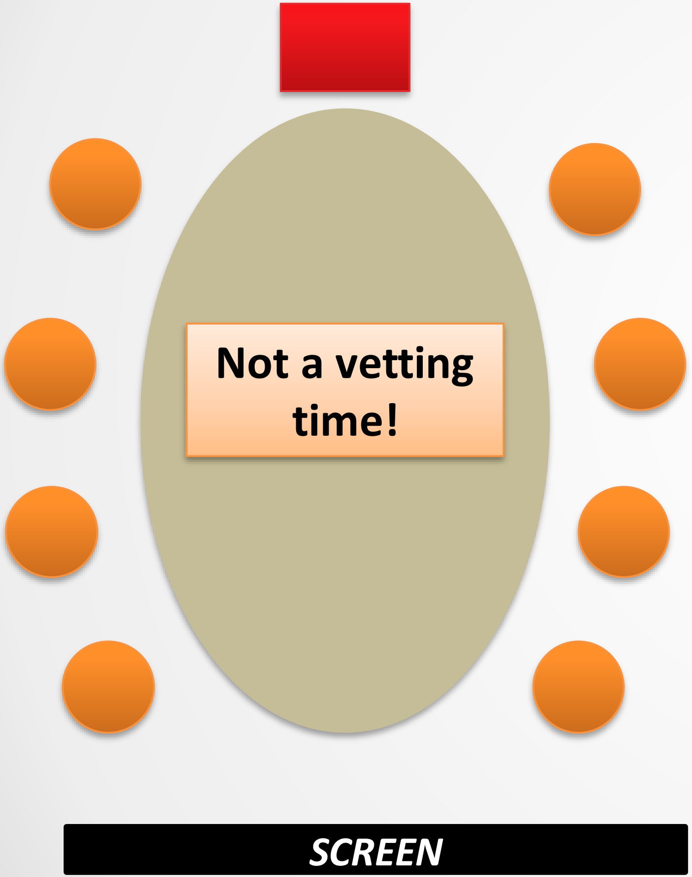
Select the methods – train judges

Define borderline standard

Build a classification table for item based on a category scheme (like difficulty and importance)

(Cizek, 2006; Ebel, 1972)

STANDARD SETTING: Ebel - DURING



Read through each question that was assigned to the respective categories in the classification table.

Judges make judgment about percentages of items in each category that borderline examinees answered correctly

Moderator: Record ratings

Calculate mean

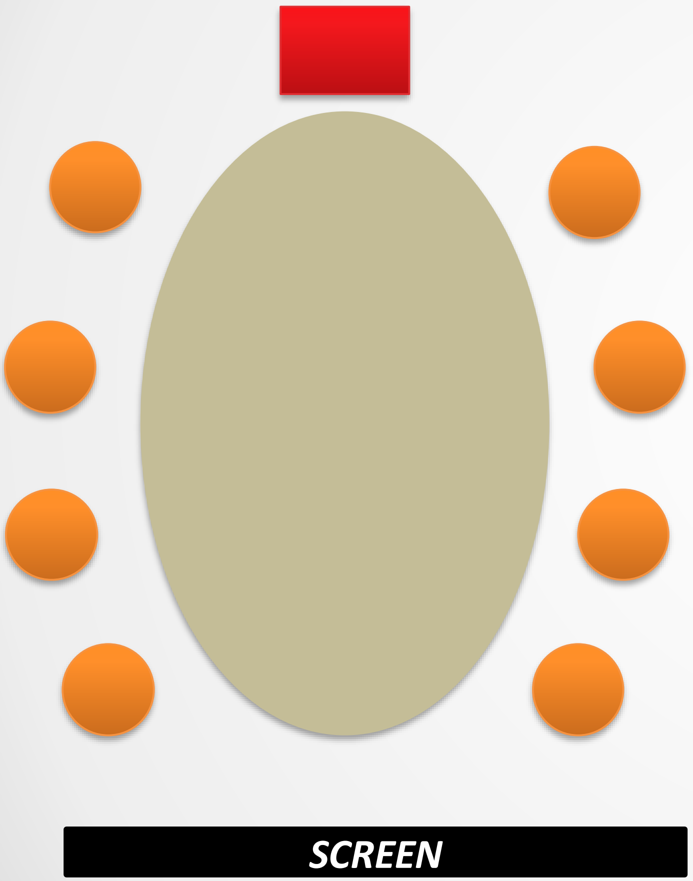
Repeat for next questions

(Cizek, 2006; Ebel, 1972)

Category	% Right	No. of Questions	Score
Essential			
Easy	95	3	2.85
Hard	80	2	1.60
Important			
Easy	90	3	2.70
Hard	75	4	3.00
Acceptable			
Easy	80	2	1.60
Hard	50	3	1.50
		17	12.25

Cut-off score

STANDARD SETTING: Ebel - POST



Evaluate the process

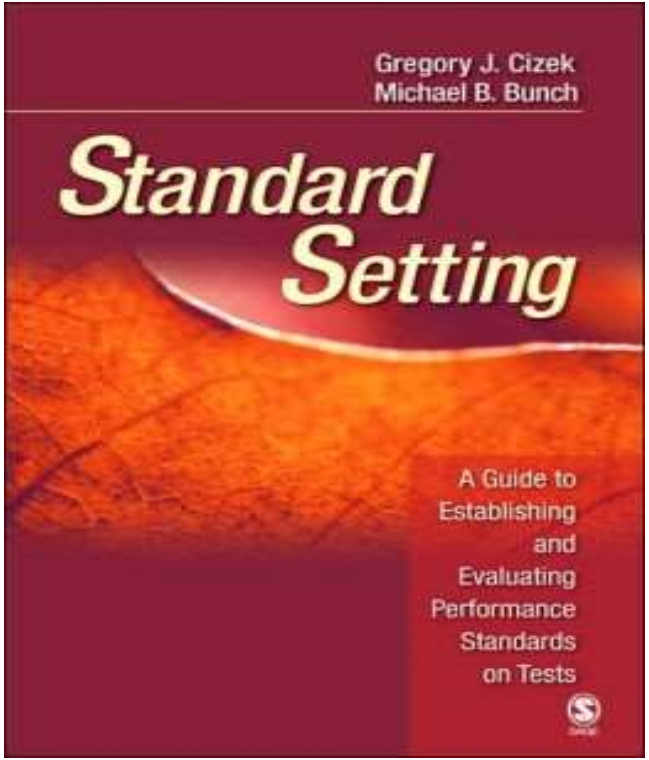
- **Judges confidence in the process**
- **Resulting cut off scores**

Documentation

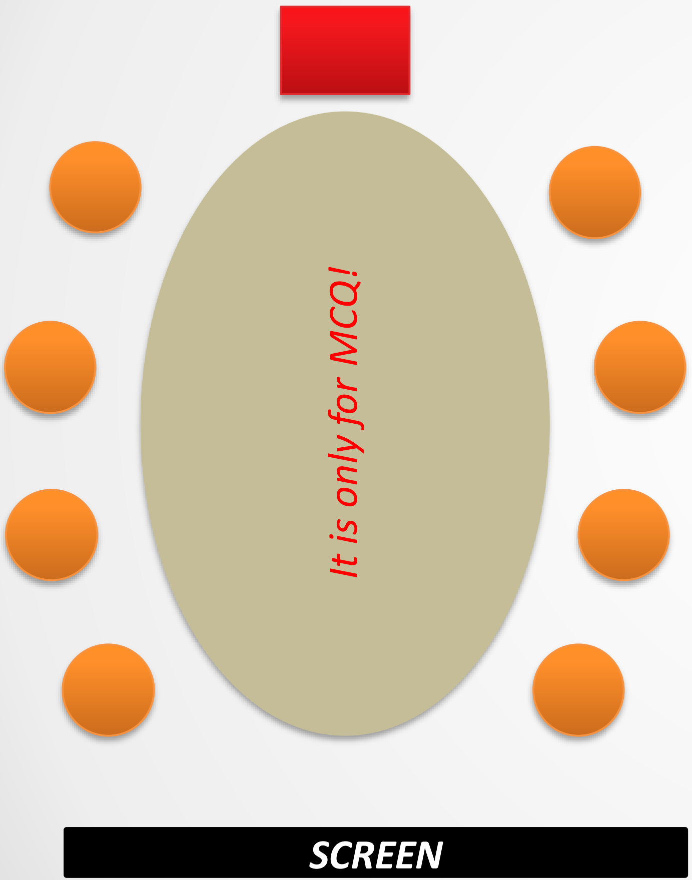
(Cizek, 2006; Ebel, 1972)

Methods for Setting Standard

- Absolute methods: Test-Items
  - Angoff (Angoff, 1971)
  - Ebel (Ebel, 1972)
  - Nedelsky (Nedelsky, 1954)
- Absolute methods: Test-Takers
  - Borderline (Livingston & Zieky, 1982)
  - Contrasting groups (Berk, 1976)
- Compromise methods
  - Hofstee (Hofstee,1983)


Gregory J. Cizek
Michael B. Bunch
Standard Setting
A Guide to Establishing and Evaluating Performance Standards on Tests

STANDARD SETTING: Nedelsky - PRE



Select the judges

Discuss

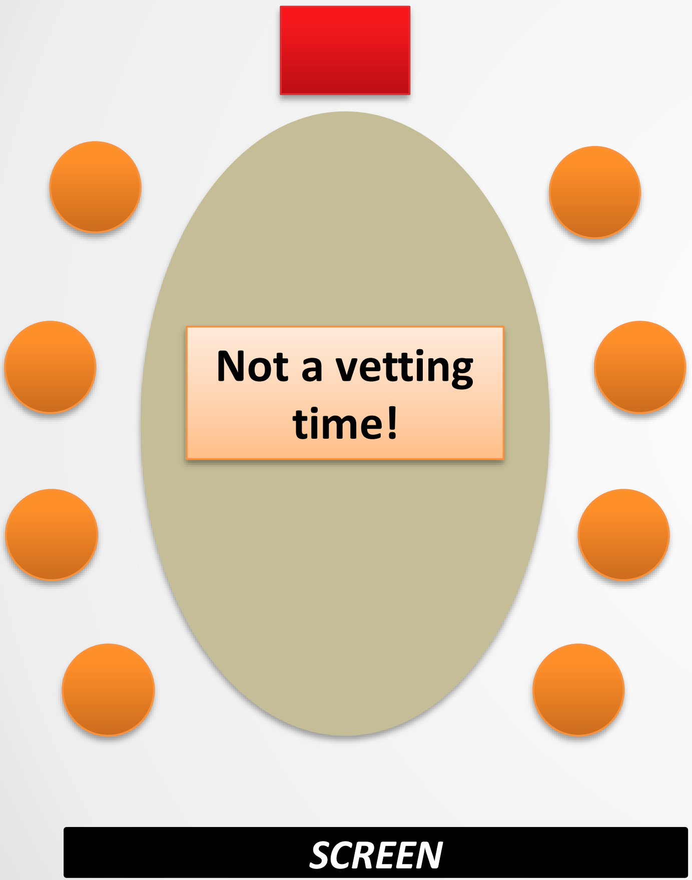
a. Purpose of the assessment
b. Nature of examinees
c. Components of adequate/inadequate knowledge

Select the methods – train judges

Define borderline standard

(Cizek, 2006; Nedelsky, 1954)

STANDARD SETTING: Nedelsky - DURING



Not a vetting time!

SCREEN

Read through each question

Judges: Working individually, judges mark the wrong answers the borderline students would be able to eliminate.

Moderator: Record ratings

Moderator: Discuss and change ratings

Repeat for next questions

Calculate passing score

(Cizek, 2006; Nedelsky, 1954)

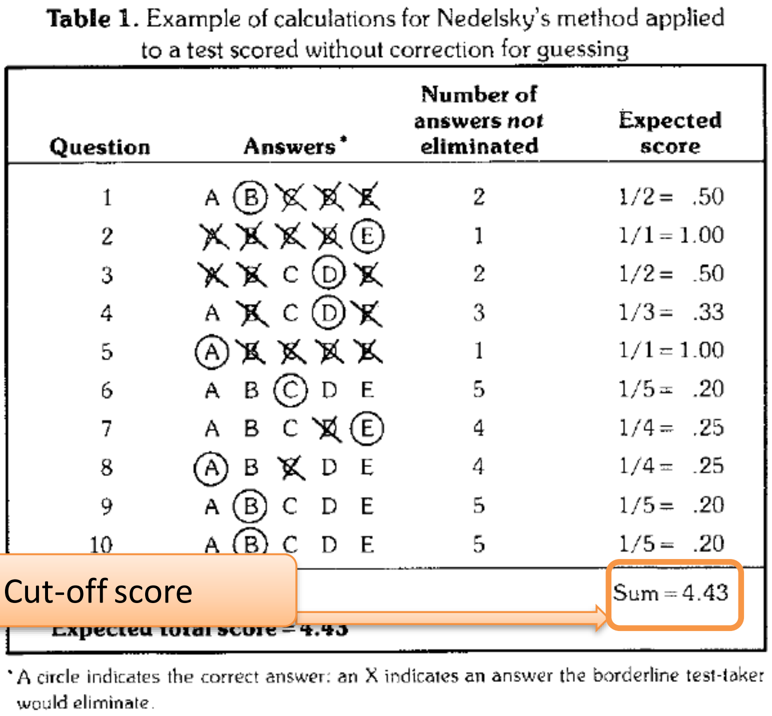


Table 1. Example of calculations for Nedelsky's method applied to a test scored without correction for guessing

Question	Answers*	Number of answers *not* eliminated	Expected score
1	A Ⓑ ~~C~~ ~~D~~ ~~E~~	2	1/2 = .50
2	~~A~~ ~~B~~ ~~C~~ ~~D~~ Ⓔ	1	1/1 = 1.00
3	~~A~~ ~~B~~ C Ⓓ ~~E~~	2	1/2 = .50
4	A ~~B~~ C Ⓓ ~~E~~	3	1/3 = .33
5	Ⓐ ~~B~~ ~~C~~ ~~D~~ ~~E~~	1	1/1 = 1.00
6	A B Ⓒ D E	5	1/5 = .20
7	A B C ~~D~~ Ⓔ	4	1/4 = .25
8	Ⓐ B ~~C~~ D E	4	1/4 = .25
9	A Ⓑ C D E	5	1/5 = .20
10	A Ⓑ C D E	5	1/5 = .20
			Sum = 4.43

Expected total score = 4.43

*A circle indicates the correct answer; an X indicates an answer the borderline test-taker would eliminate.

Cut-off score

- Three methods of calculating passing score:
  - Mean
  - Median
  - Trimmed mean

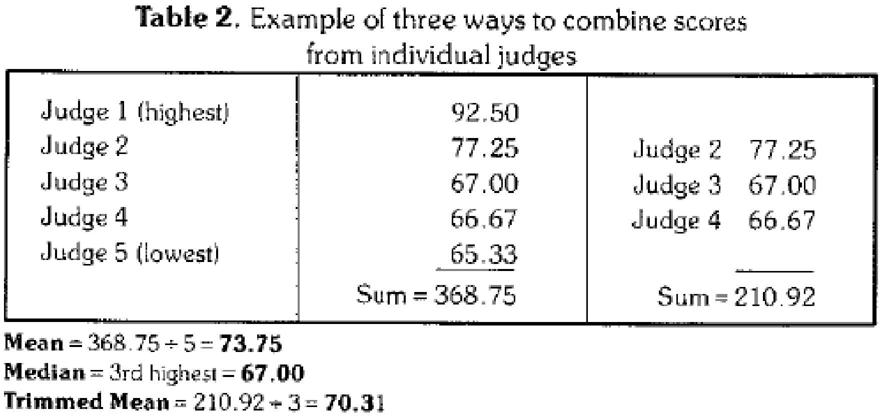
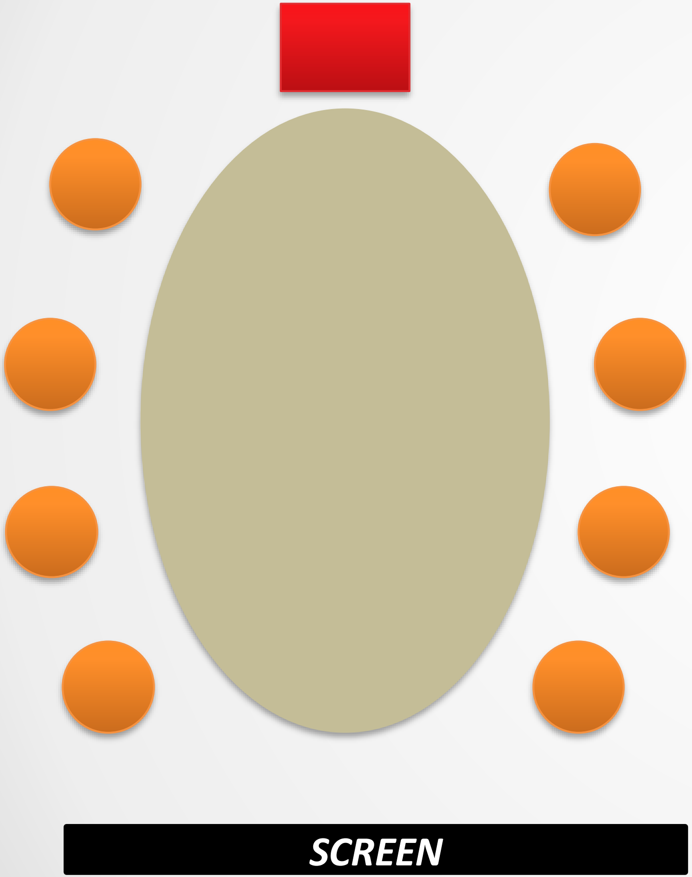


Table 2. Example of three ways to combine scores from individual judges

Judge 1 (highest)	92.50	
Judge 2	77.25	Judge 2 77.25
Judge 3	67.00	Judge 3 67.00
Judge 4	66.67	Judge 4 66.67
Judge 5 (lowest)	65.33	
	Sum = 368.75	Sum = 210.92

Mean = 368.75 ÷ 5 = **73.75**
Median = 3rd highest = **67.00**
Trimmed Mean = 210.92 ÷ 3 = **70.31**

STANDARD SETTING: Nedelsky - POST



Evaluate the process

- Judges confidence in the process
- Resulting cut off scores

Documentation

(Cizek, 2006; Nedelsky, 1954)

ABSOLUTE METHOD – TEST-ITEM: Advantages

They are used frequently in high stakes examination

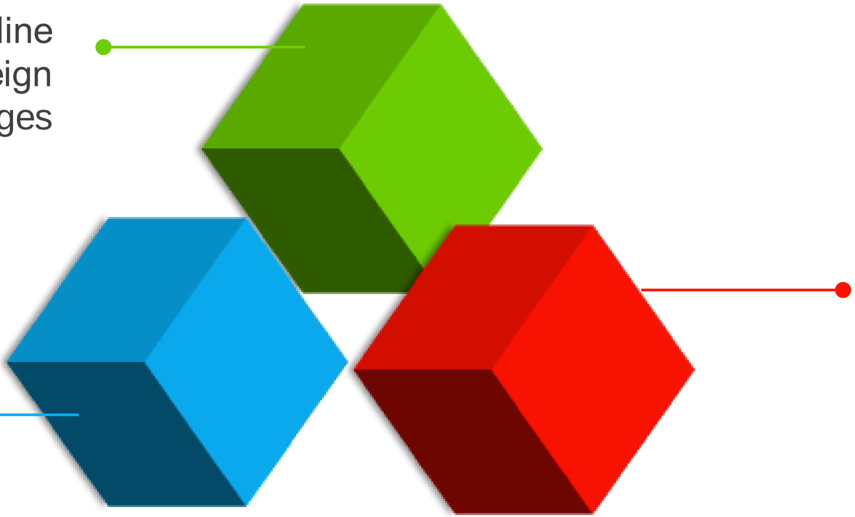
They are relatively easy to use

There is a considerable body of published work to support their use

They focus on item content

(Norcini and Guille, 2002)

ABSOLUTE METHOD – TEST-ITEM: Disadvantages



The concept of a ‘borderline group’ is sometimes foreign to judges

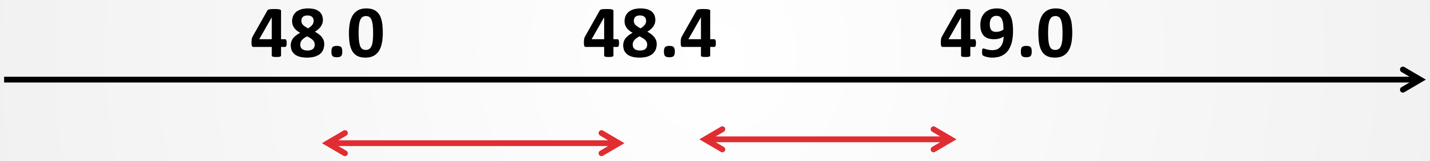
The methods can be tedious

Judges sometimes feel they are ‘pulling numbers out of air’

(Norcini and Guille, 2002)

STANDARD SETTING: ISSUES

ROUNDING?



48.0 **48.4** **49.0**

False positive

False negative
Cost

STANDARD SETTING: Negative Marking

Negative marking in MTF →

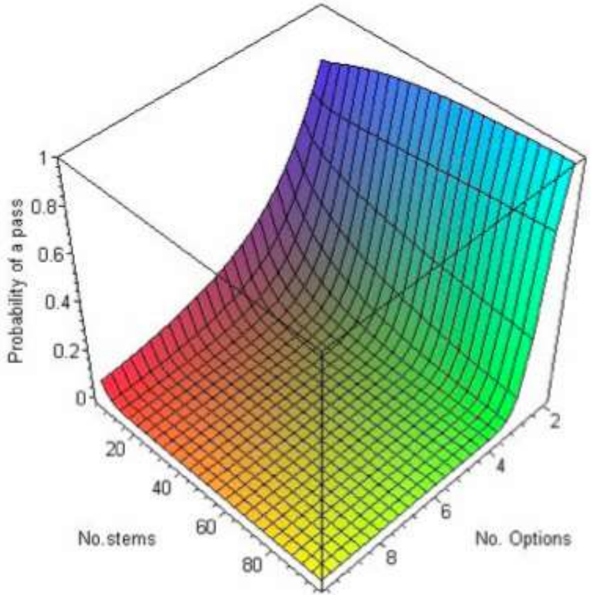
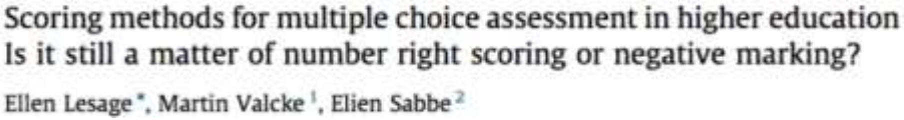


Figure 1 The probability of passing using a pure guessing strategy

(Holt, 2006)

Doesn’t solve guessing problem.
(Bar-Hillel et al., 2005; Betts et al., 2009)

But add in more uncertainty - risk taking behaviour
(Budescu & Bar-Hillel, 1993; Choppin, 1988; Fowell & Jolly, 2000; Hammond et al., 1998; Kurz, 1999; Moss, 2001; Prihoda et al., 2006)



Scoring methods for multiple choice assessment in higher education
Is it still a matter of number right scoring or negative marking?

Ellen Lesage [*], Martin Valcke [1], Elien Sabbe [2]

Suggestions:

1. To replace negative marking with standard setting.
2. Guessing effect can be reduced with good item construction.
3. To replace MTF with SBA
4. Increase sampling in assessment

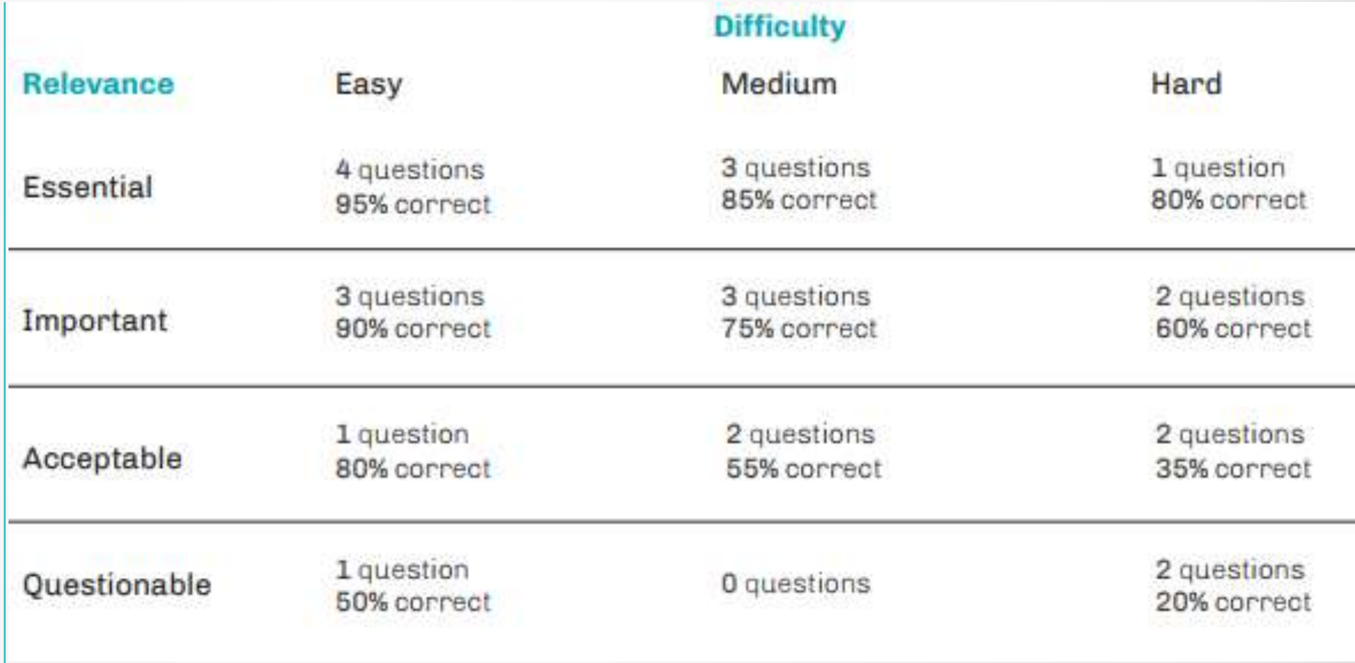
STANDARD SETTING: ISSUES

SAMPLING MIXTURE?

ANGOFF

Sample proportionately based on blueprint

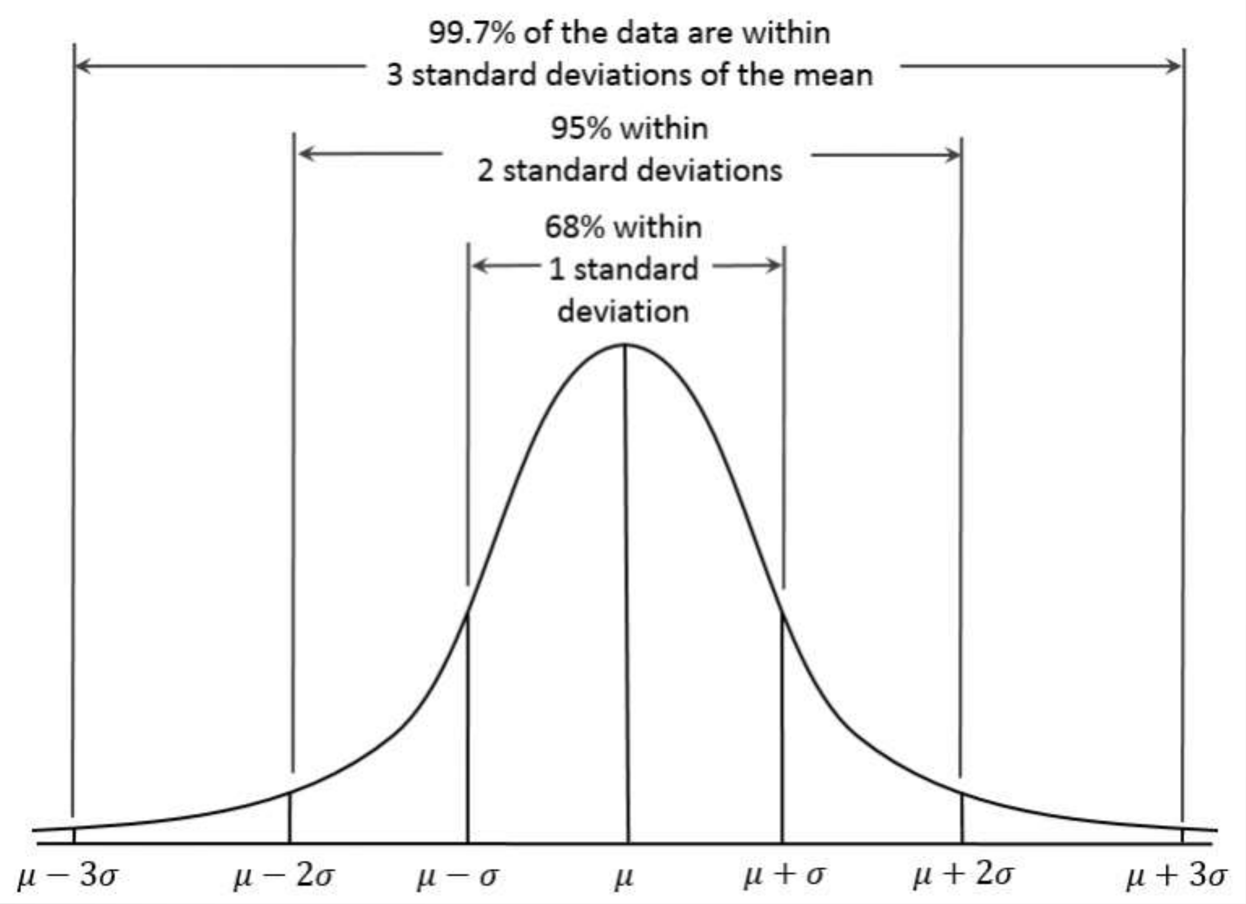
Must Know - _%
Should Know - _%
Nice to Know - _%



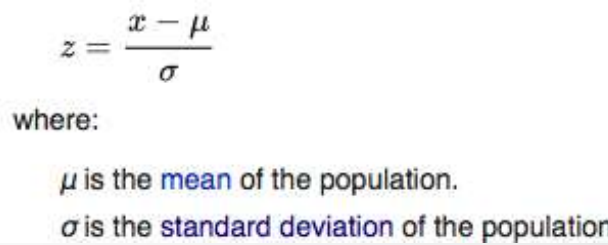
		Difficulty	
Relevance	Easy	Medium	Hard
Essential	4 questions 95% correct	3 questions 85% correct	1 question 80% correct
Important	3 questions 90% correct	3 questions 75% correct	2 questions 60% correct
Acceptable	1 question 80% correct	2 questions 55% correct	2 questions 35% correct
Questionable	1 question 50% correct	0 questions	2 questions 20% correct

EBEL METHOD – Based on item relevance and difficulty (but less used as compared to Angoff's)

STANDARD SETTING: CONVERSION TO '50%'



1st Step: Calculate Z Score



$$z = \frac{x - \mu}{\sigma}$$

where:

μ is the mean of the population.

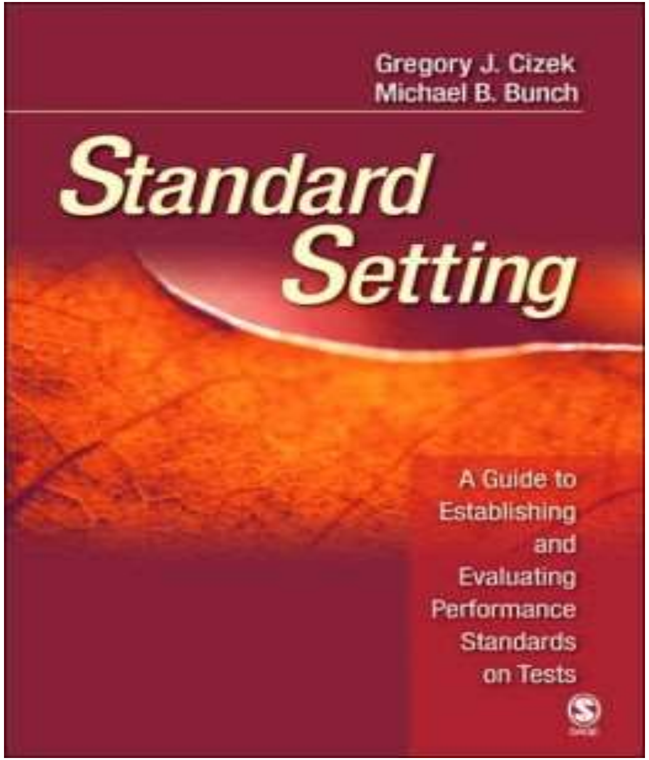
σ is the standard deviation of the populatior

2nd Step: Calculate Standardized Score

= (Z-score X Standard Deviation) + Desired Passing Score

Methods for Setting Standard

- Absolute methods: Test-Items
  - – Angoff (Angoff, 1971)
  - – Ebel (Ebel, 1972)
  - – Nedelsky (Nedelsky, 1954)
- Absolute methods: Test-Takers
  - – Borderline (Livingston & Zieky, 1982)
  - – Contrasting groups (Berk, 1976)
- Compromise methods
  - – Hofstee (Hofstee,1983)

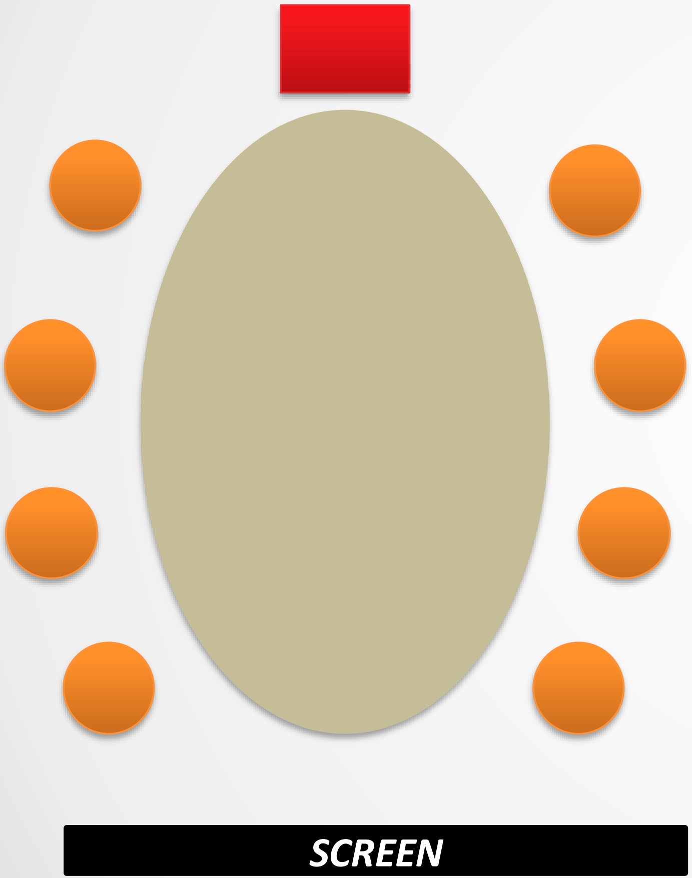

Gregory J. Cizek
Michael B. Bunch
Standard Setting
A Guide to Establishing and Evaluating Performance Standards on Tests

Standard Setting in Action

ABSOLUTE METHODS – TEST-TAKER

- Process
- Practice
- Insights

STANDARD SETTING: Borderline Method - PRE



Select the judges

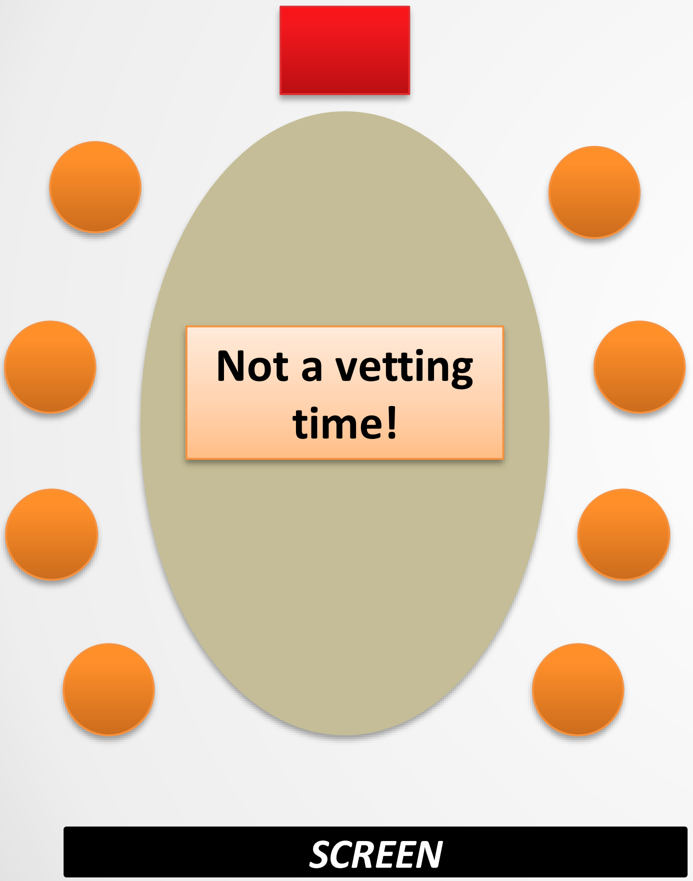
Discuss
a. Purpose of the assessment
b. Nature of examinees
c. Components of adequate/inadequate performance

Inform the borderline standard

Orientate the judges to the test, station or case to the checklist item scores

(Cizek, 2006; Hambleton, 1998)

STANDARD SETTING: Borderline Method - DURING



Observe performance of candidate

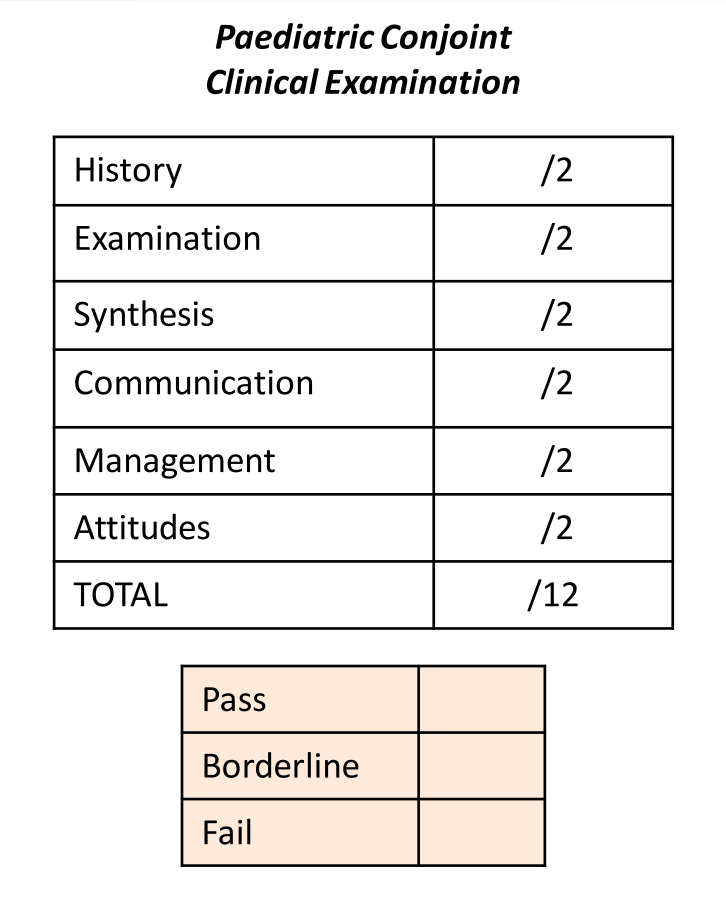
Judges:
#1 provide a global rating of the overall performance each examinee on three-point scale: Fail, Borderline, Pass
#2 The performance is also scored using a multiple-item checklist

Moderator: Record ratings

Repeat for next stations

(Cizek, 2006; Hambleton, 1998)

STANDARD SETTING: Borderline Method - DURING



Paediatric Conjoint Clinical Examination

History	/2
Examination	/2
Synthesis	/2
Communication	/2
Management	/2
Attitudes	/2
TOTAL	/12

Pass	
Borderline	
Fail	

Collate the marks of candidates rated as borderline

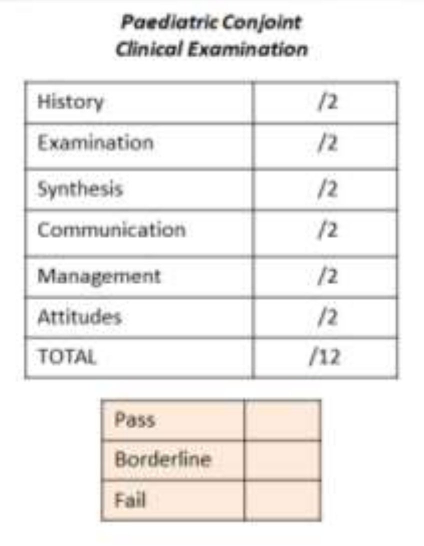
Mean or median of the borderline cohort is taken as STATION PASSING SCORE

In Conjunctive Strategy – the candidates must exceed the STATION PASSING SCORE to pass

In Compensatory Strategy –the station passing score is summed up across station to form OVERALL PASSING SCORE

STANDARD SETTING: Borderline Method - DURING

STATION 1



Paediatric Conjoint Clinical Examination

History	/2
Examination	/2
Synthesis	/2
Communication	/2
Management	/2
Attitudes	/2
TOTAL	/12

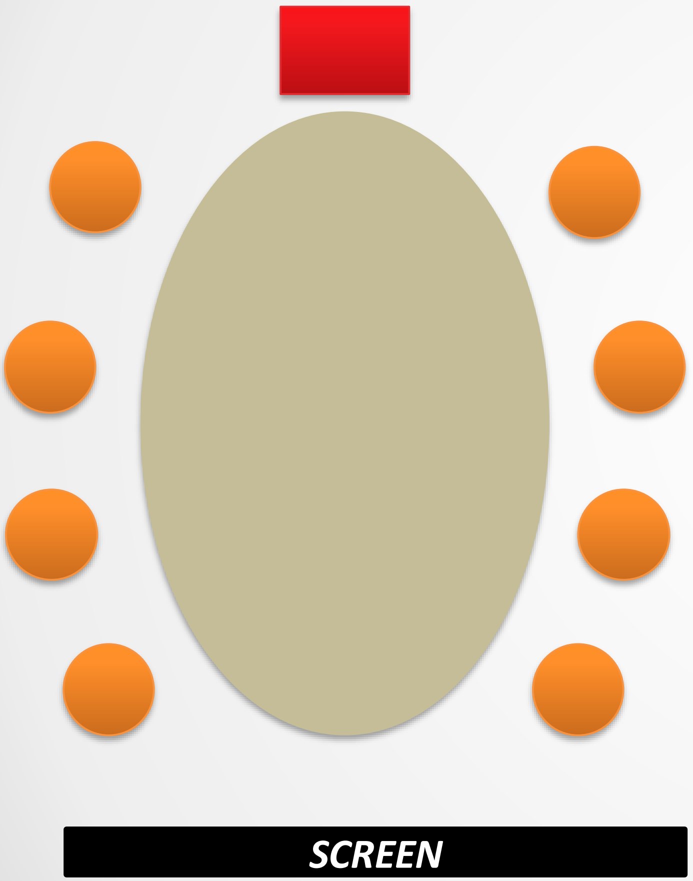
Pass	
Borderline	
Fail	

EXAMINER 1

	Marks given	Pass	Borderline	Fail
Student 1	12	/		
Student 2	8		/	
Student 3	10		/	
Student 4	6			/
Student 5	10		/	
Student 6	10	/		
Student 7	8		/	
Student 8	8			/

Passing score for this station = (8 + 10 + 10 +8)/ 4
= 9

STANDARD SETTING: Borderline Method - POST



Evaluate the process

- **Judges confidence in the process**
- **Resulting cut off scores**

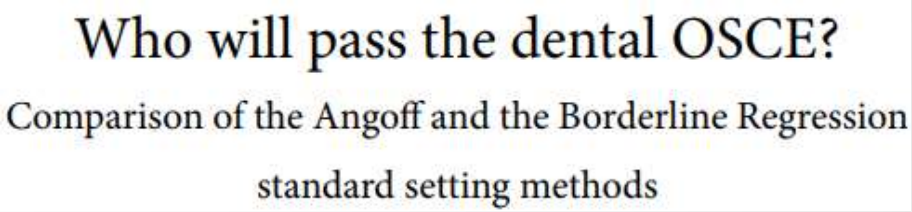
Documentation

(Cizek, 2006; Hambleton, 1998)

STANDARD SETTING: Borderline Method

1. Simple, save time

2. More acceptable passing scores than Angoff's (Klein et al, 2008)



Who will pass the dental OSCE?

Comparison of the Angoff and the Borderline Regression standard setting methods

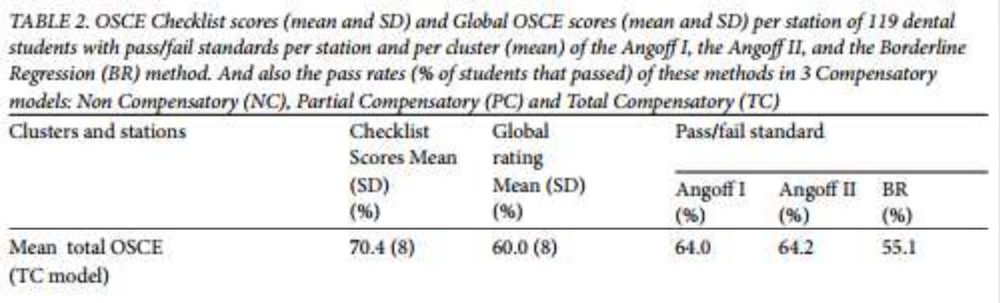


TABLE 2. OSCE Checklist scores (mean and SD) and Global OSCE scores (mean and SD) per station of 119 dental students with pass/fail standards per station and per cluster (mean) of the Angoff I, the Angoff II, and the Borderline Regression (BR) method. And also the pass rates (% of students that passed) of these methods in 3 Compensatory models: Non Compensatory (NC), Partial Compensatory (PC) and Total Compensatory (TC)

Clusters and stations	Checklist Scores Mean (SD) (%)	Global rating Mean (SD) (%)	Pass/fail standard		
			Angoff I (%)	Angoff II (%)	BR (%)
Mean total OSCE (TC model)	70.4 (8)	60.0 (8)	64.0	64.2	55.1

3. May be influenced by the examinees "non-examination factors" (gender, university, etc)(Cizek, 2007)

4. Does not utilize all data. What if no one or too few in borderline?

(Wood, Hunphrey-Murto, Norman, 2006)

STANDARD SETTING: Borderline Regression Method

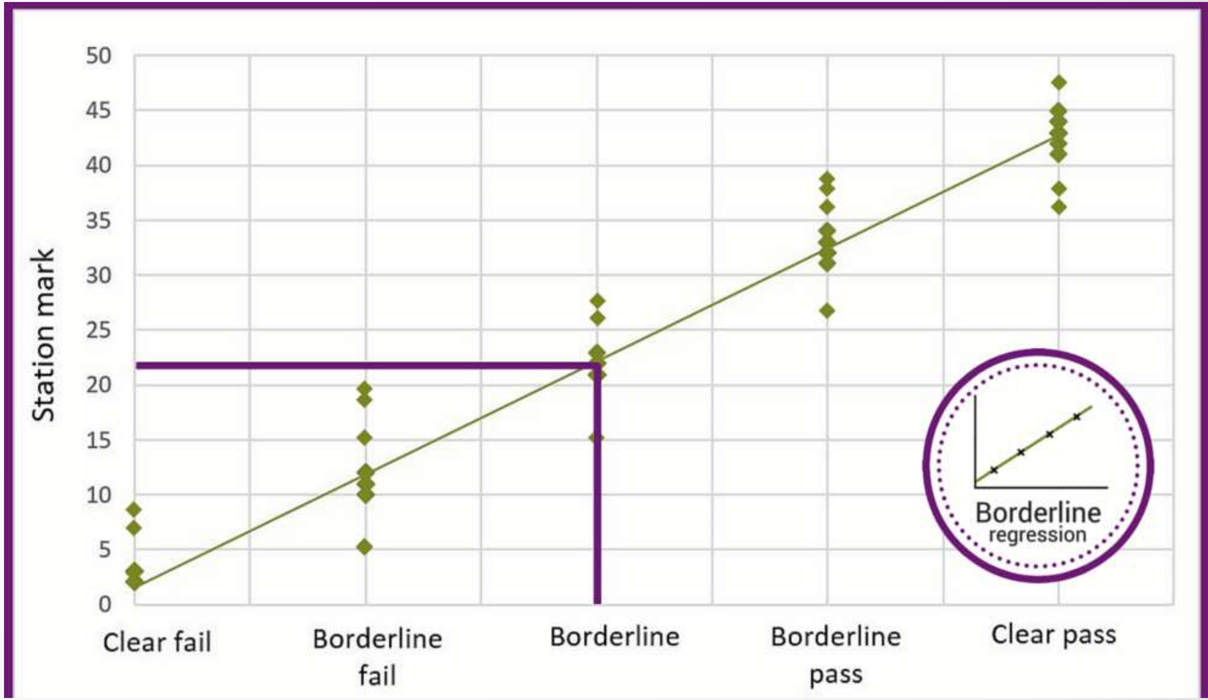
1. All data (Fail, Borderline, Pass) from Station 1 are entered.

2. Run Linear Regression

3. The point where regression line intersects borderline = station passing score

4. Repeat for other stations

5. Sum all stations passing score = PASSING SCORE



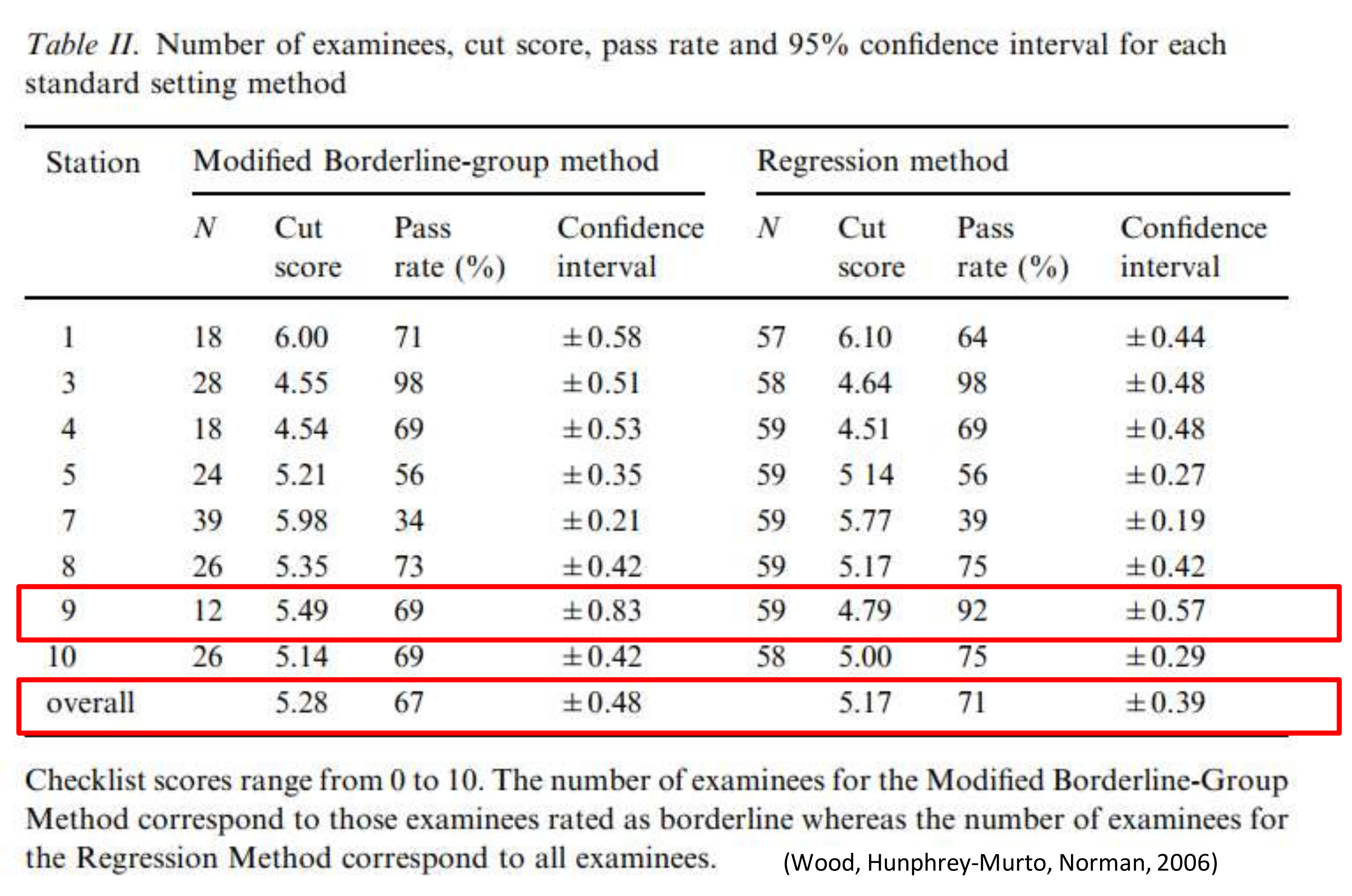


Table II. Number of examinees, cut score, pass rate and 95% confidence interval for each standard setting method

Station	Modified Borderline-group method				Regression method			
	N	Cut score	Pass rate (%)	Confidence interval	*N*	Cut score	Pass rate (%)	Confidence interval
1	18	6.00	71	±0.58	57	6.10	64	±0.44
3	28	4.55	98	±0.51	58	4.64	98	±0.48
4	18	4.54	69	±0.53	59	4.51	69	±0.48
5	24	5.21	56	±0.35	59	5 14	56	±0.27
7	39	5.98	34	±0.21	59	5.77	39	±0.19
8	26	5.35	73	±0.42	59	5.17	75	±0.42
9	12	5.49	69	±0.83	59	4.79	92	±0.57
10	26	5.14	69	±0.42	58	5.00	75	±0.29
overall		5.28	67	±0.48		5.17	71	±0.39

Checklist scores range from 0 to 10. The number of examinees for the Modified Borderline-Group Method correspond to those examinees rated as borderline whereas the number of examinees for the Regression Method correspond to all examinees.

(Wood, Hunphrey-Murto, Norman, 2006)

If we look at station 9, where borderline candidate is not many, the passing score was significantly higher

However, if the mean of both method yield comparable passing score

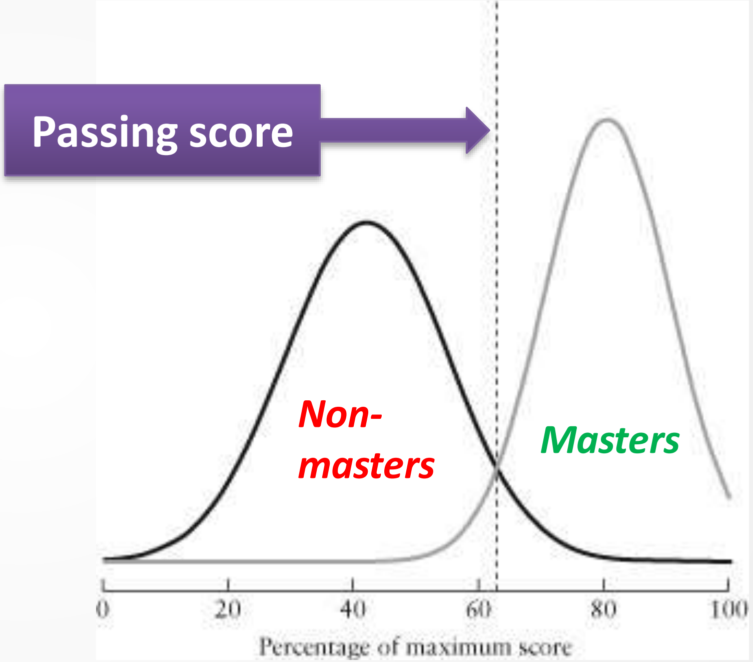
STANDARD SETTING: Contrasting Group Method

Paediatric Conjoint Clinical Examination

History	/2
Examination	/2
Synthesis	/2
Communication	/2
Management	/2
Attitudes	/2
TOTAL	/12

Masters	
Non-masters	

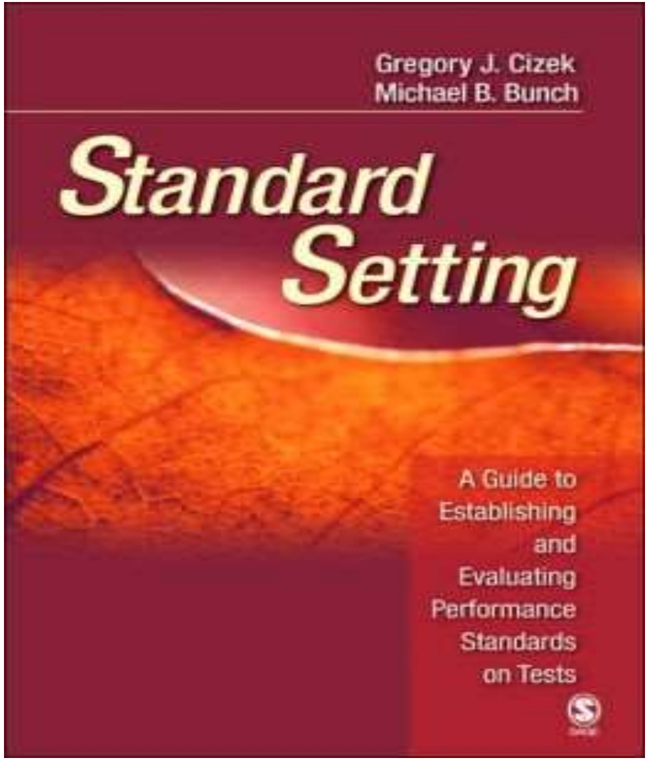
Judgment made on real and across candidates performances



But can we always categorize candidates into masters and non masters?

Methods for Setting Standard

- Absolute methods: Test-Items
  - – Angoff (Angoff, 1971)
  - – Ebel (Ebel, 1972)
  - – Nedelsky (Nedelsky, 1954)
- Absolute methods: Test-Takers
  - – Borderline (Livingston & Zieky, 1982)
  - – Contrasting groups (Berk, 1976)
- Compromise methods
  - – Hofstee (Hofstee,1983)

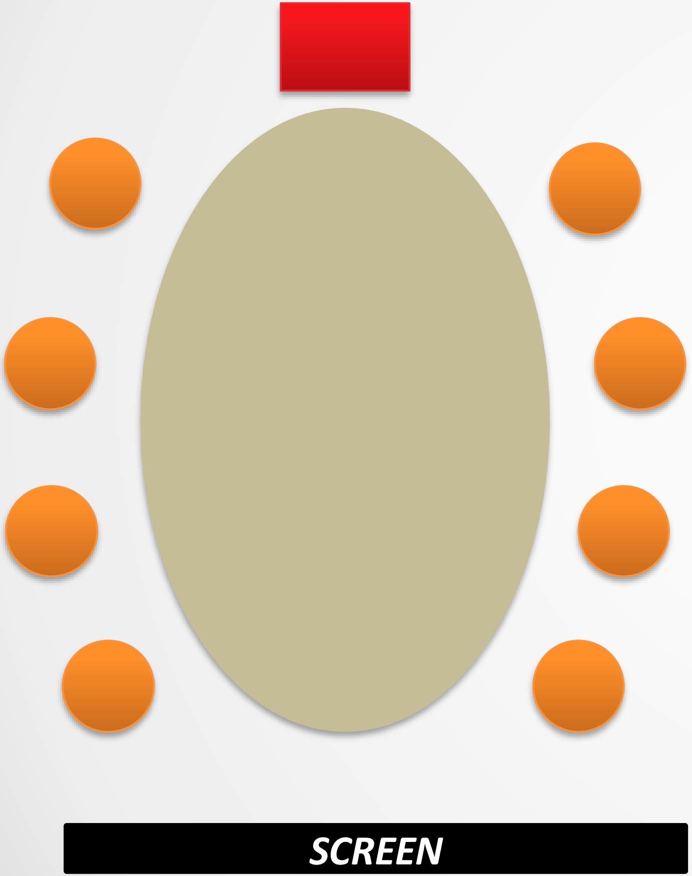

Gregory J. Cizek
Michael B. Bunch
Standard Setting
A Guide to Establishing and Evaluating Performance Standards on Tests

Standard Setting in Action

COMPROMISE METHOD

- Process
- Practice
- Insights

STANDARD SETTING: Hofstee - PRE



Select the judges

Discuss

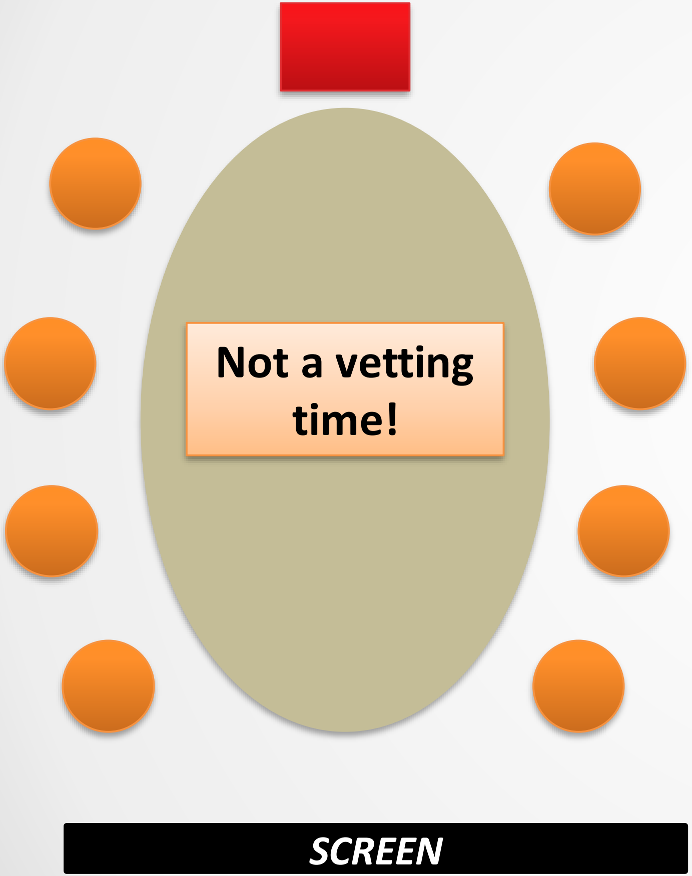
a. Purpose of the assessment
b. Nature of examinees
c. Components of adequate/inadequate knowledge

Select the methods – train judges

Review the test in detail

(Cizek, 2006; Hofstee,1983)

STANDARD SETTING: Hofstee - DURING

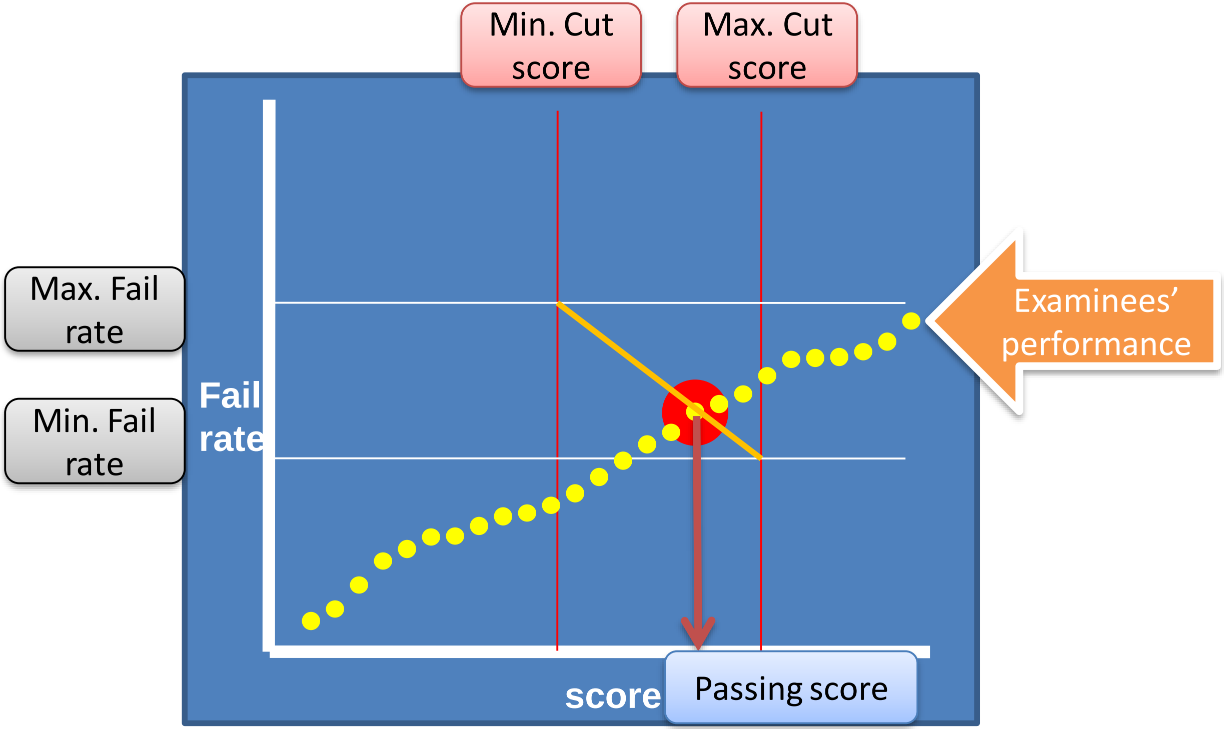


Ask the judges to answer 4 questions:

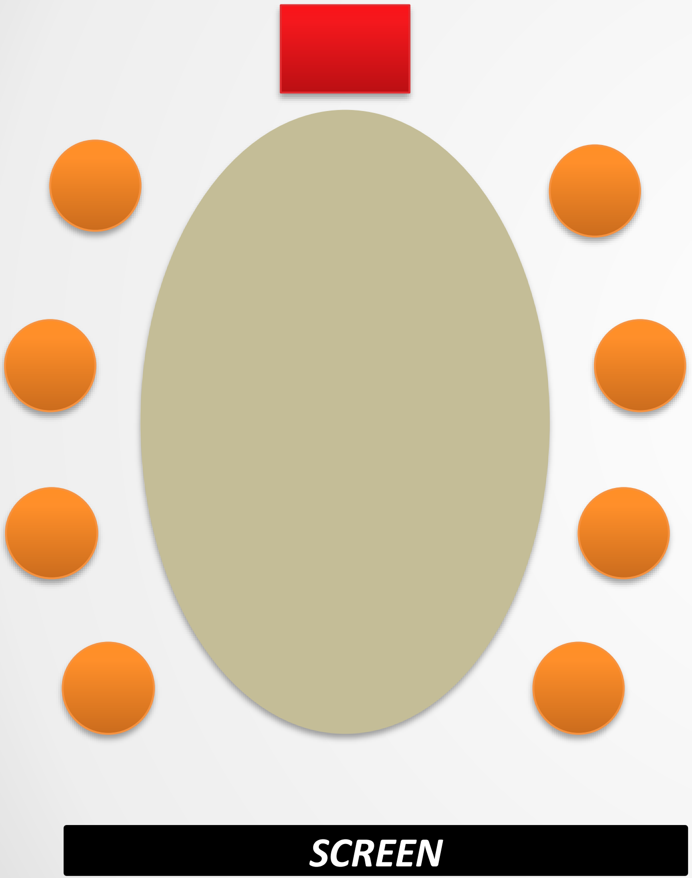
- ✓ What is the minimum acceptable cut score?
- ✓ What is the maximum acceptable cut score?
- ✓ What is the minimum acceptable fail rate?
- ✓ What is the maximum acceptable fail rate?

After the test is given, graph the distribution of scores and select the cut score.

(Cizek, 2006; Hofstee,1983)


Min. Cut score
Max. Cut score
Max. Fail rate
Min. Fail rate
Fail rate
Examinees' performance
score
Passing score

STANDARD SETTING: Hofstee - POST



Evaluate the process

- Judges confidence in the process
- Resulting cut off scores

Documentation

(Cizek, 2006; Hofstee,1983)

STANDARD SETTING: Hofstee - POST

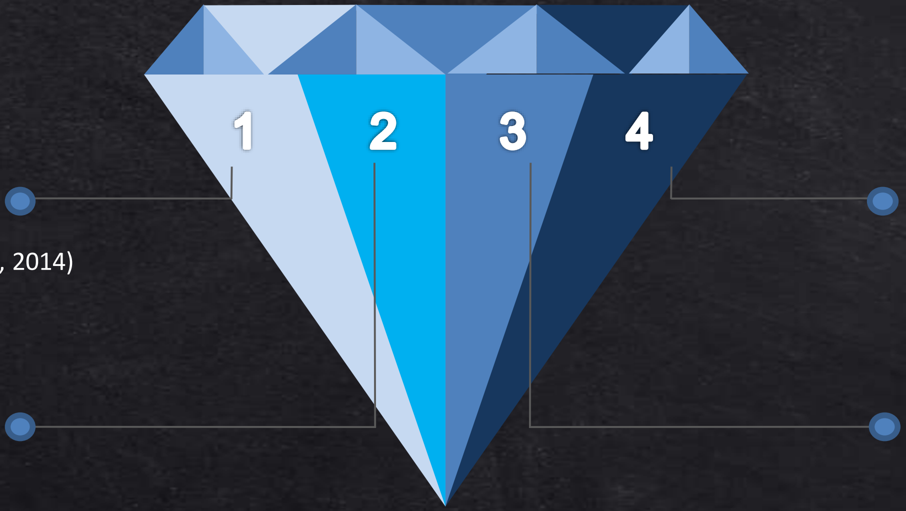
Advantages

- Easy to implement
- Educators are comfortable with the decision

Disadvantages

- The cut score may not be in the area defined by the judges’ estimates.
- The method is not the first choice in a high stakes testing situation.

FINAL NOTES



Method Dependent

The resulting standards are method dependent

(AMEE Guide 37, 2008; AMEE Guide 85, 2014)

Learning Process

No most accurate score or gold standard

(AMEE Guide 37, 2008; AMEE Guide 85, 2014)

Credible Panel

Panels must be those familiar with students, assessment and content

(Cizek, 2007; AMEE Guide 85, 2014)

Methodology

Choose method depending on purpose, evidence and resources

(AMEE Guide 85, 2014)

Thank You

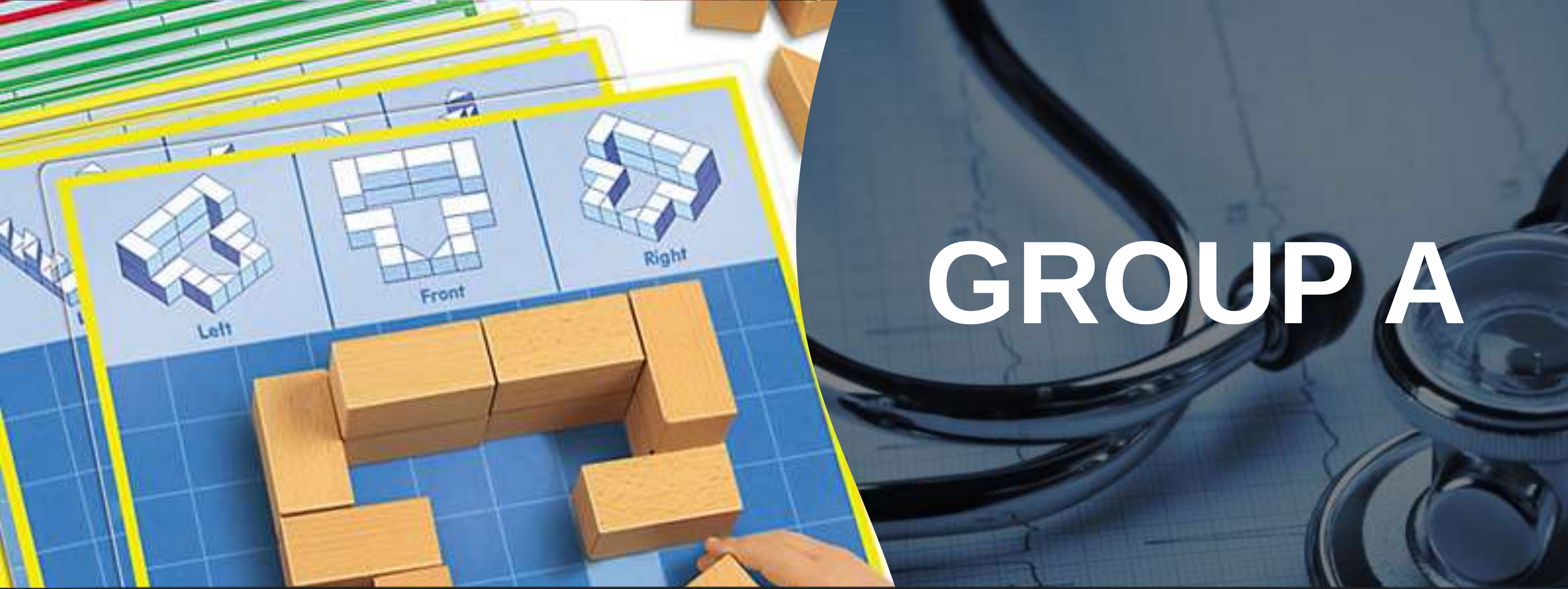
MUHAMAD SAIFUL BAHRI YUSOFF, MD, MSC, PHD

Department of Medical Education, School of Medical Sciences, Universiti Sains Malaysia, email: msaiful_bahri@usm.my.

https://www.researchgate.net/profile/Muhamad_Saiful_Bahri_Yusoff

Further reading

- Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.), *Educational measurement (2nd ed., pp. 508-600). Washington, DC: American Council on* Education.
- Berk RA., Determination of Optional Cutting Scores in Criterion-Referenced Measurement *The Journal of Experimental Education* Vol. 45, No. 2 (Winter, 1976), pp. 4-9
- Cizek, G.J. and Bunch, M.B. (2007). Standard setting: A guide to establishing and evaluating performance standards. Thousand Oaks, CA: Sage Publications. (Practical)
- Ebel RL. Essentials of Educational Measurement. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1972.
- Livingston SA, Zieky MJ. Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests. Princeton, NJ: Educational Testing Service, 1982.

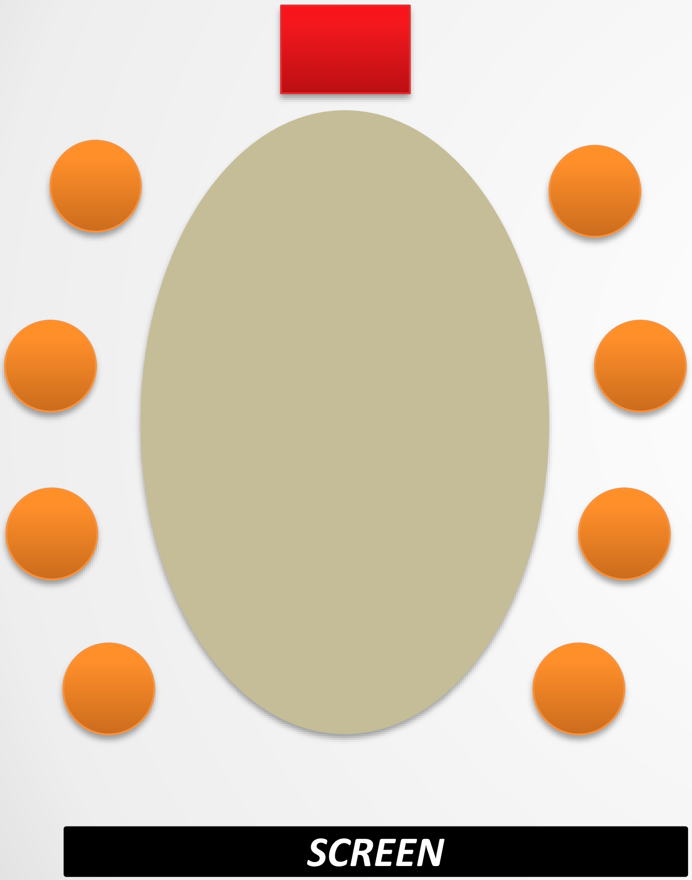


GROUP A

STANDARD SETTING

ANGOFF

STANDARD SETTING: Angoff - PRE



Select the judges

Discuss

a. Purpose of the assessment
b. Nature of examinees
c. Components of adequate/inadequate knowledge

Select the methods – train judges

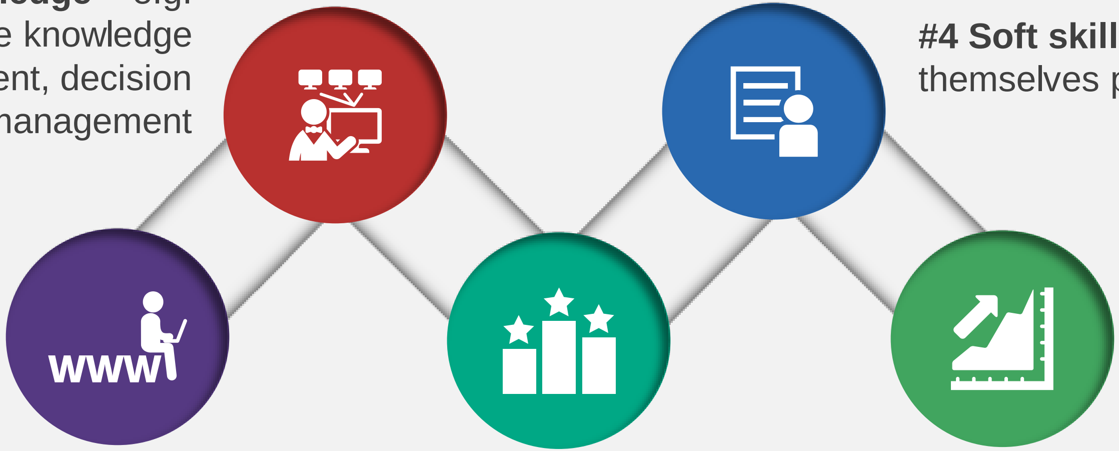
Define borderline standard

(Cizek, 2006; Angoff, 1971)

STANDARD SETTING: Define Borderline

# 2 Knowledge – e.g. demonstrate adequate knowledge for safe clinical judgment, decision making and management

#4 Soft skills - e.g. conduct themselves professionally



#1 Setting – e.g. graduate of the ophthalmology program

#3 Skills – e.g. be able to work with moderate supervision, equipped with acceptable technical ability

#5 Errors (considering the forgivable or unforgivable) – e.g. safe clinical judgment, decision making and management

(Mills, Melican & Ahluwalia, 1991)

STANDARD SETTING: Borderline Standards

Setting

Errors

Forgivable, non-forgivable

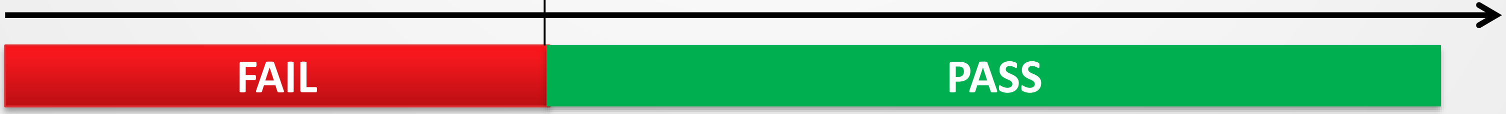
"The borderline **graduate of the ophthalmology program** should demonstrate **adequate knowledge for safe clinical judgment**, **decision making and management**, be able to **work with moderate supervision**, equipped with **acceptable technical ability**, and **conduct themselves professionally**."

Knowledge

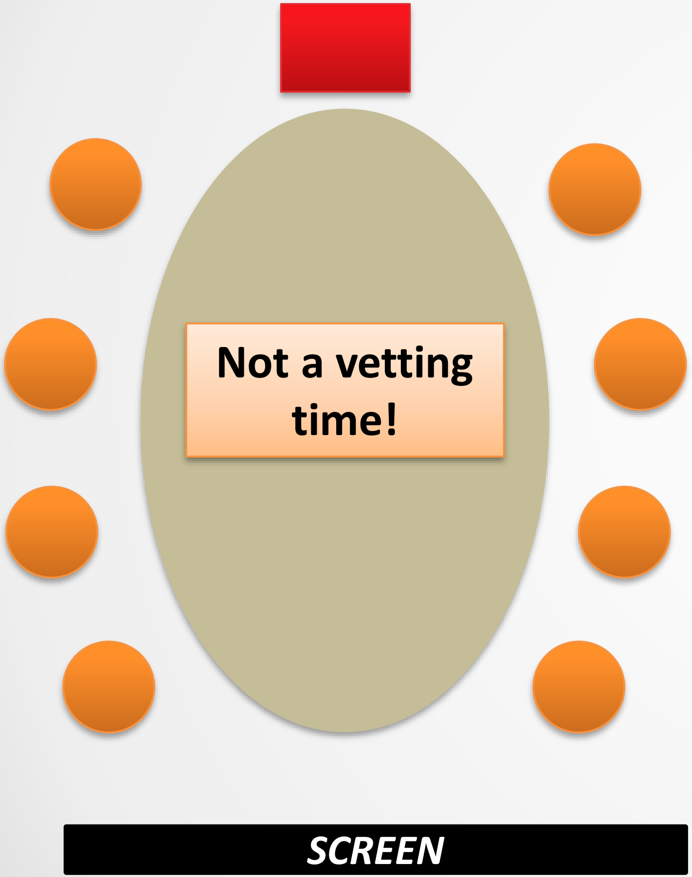
Skills

Attitude

(MUCCO, 20-22 Aug 2014, A Workshop on Examination Questions Preparation, Kuala Lumpur)



STANDARD SETTING: Angoff - DURING



Read through question 1

Judges: Individually, estimate proportion of borderline examinees will correctly answer question 1

Moderator: Record ratings

Moderator: Discuss ratings

Moderator: Get 2nd ratings after discussion

Calculate mean

Repeat for next questions

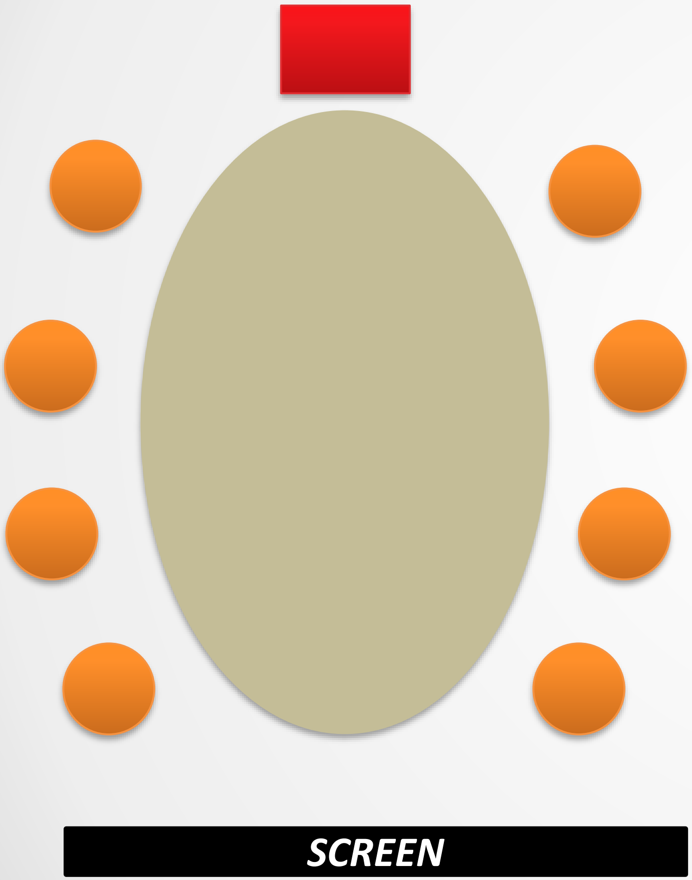
(Cizek, 2006; Angoff, 1971)

	Q1	Q2	Q3	Q4	Q5	Mean
JUDGE 1	60					
	60					
JUDGE 2	50					
	60					
JUDGE 3	**90**					
	60					
JUDGE 4	60					
	50					
JUDGE 5	60					
	60					
JUDGE 6	**40**					
	60					
Mean 1st	**60**					
Mean 2nd	**58.3**					

Cut-off score 1st round

Cut-off score 2nd round

STANDARD SETTING: Angoff - POST

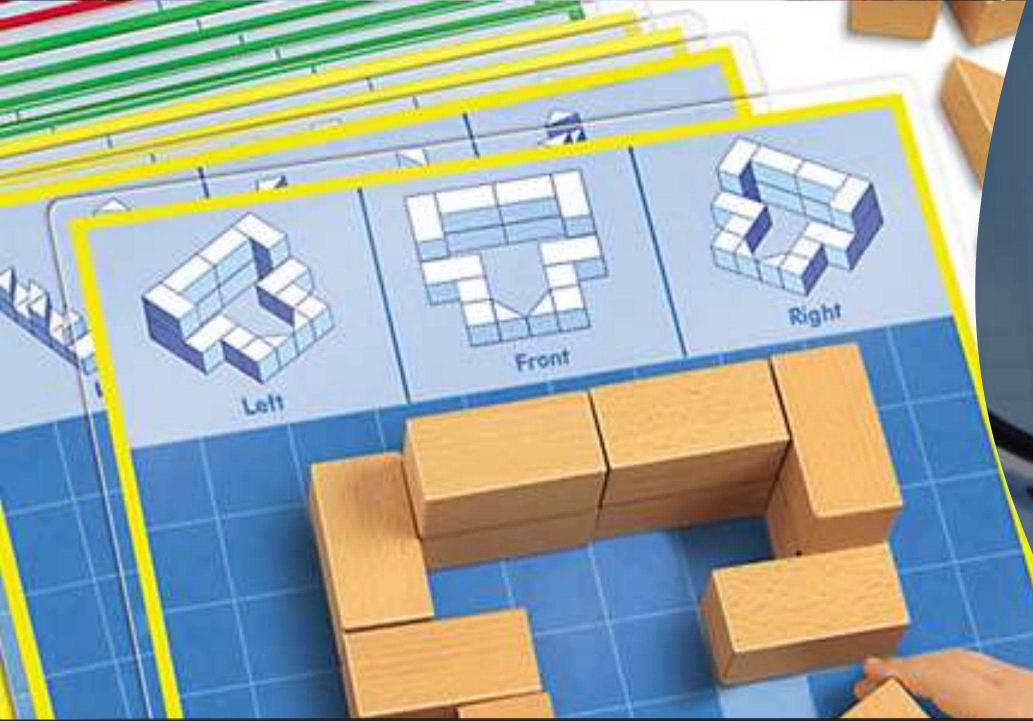


Evaluate the process

- Judges confidence in the process
- Resulting cut off scores

Documentation

(Cizek, 2006; Angoff, 1971)

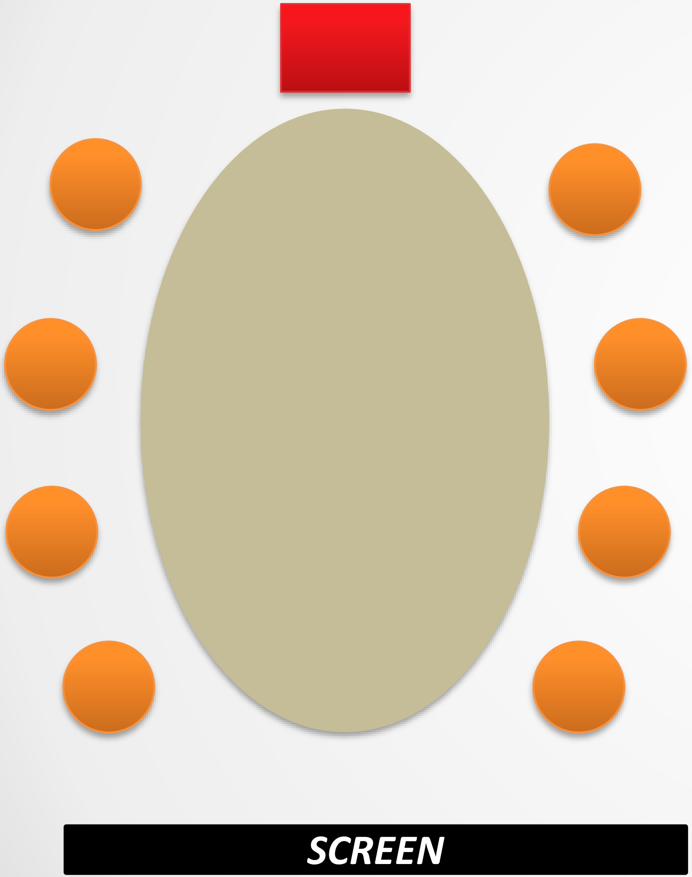


GROUP B

STANDARD SETTING

MODIFIED ANGOFF

STANDARD SETTING: Modified Angoff - PRE



Select the judges

Discuss

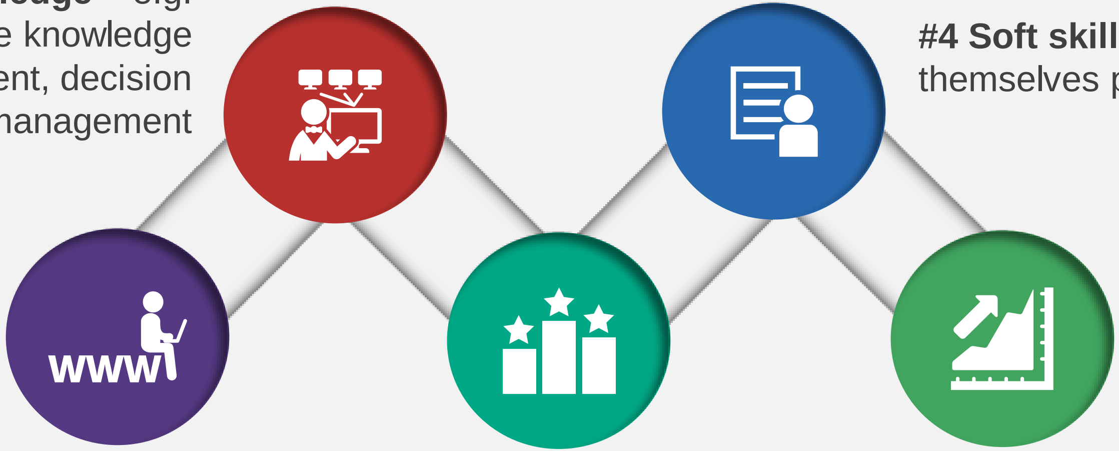
a. Purpose of the assessment
b. Nature of examinees
c. Components of adequate/inadequate knowledge

Select the methods – train judges

Define borderline standard

(Cizek, 2006; Angoff, 1971)

STANDARD SETTING: Define Borderline



# 2 Knowledge – e.g. demonstrate adequate knowledge for safe clinical judgment, decision making and management

#4 Soft skills - e.g. conduct themselves professionally

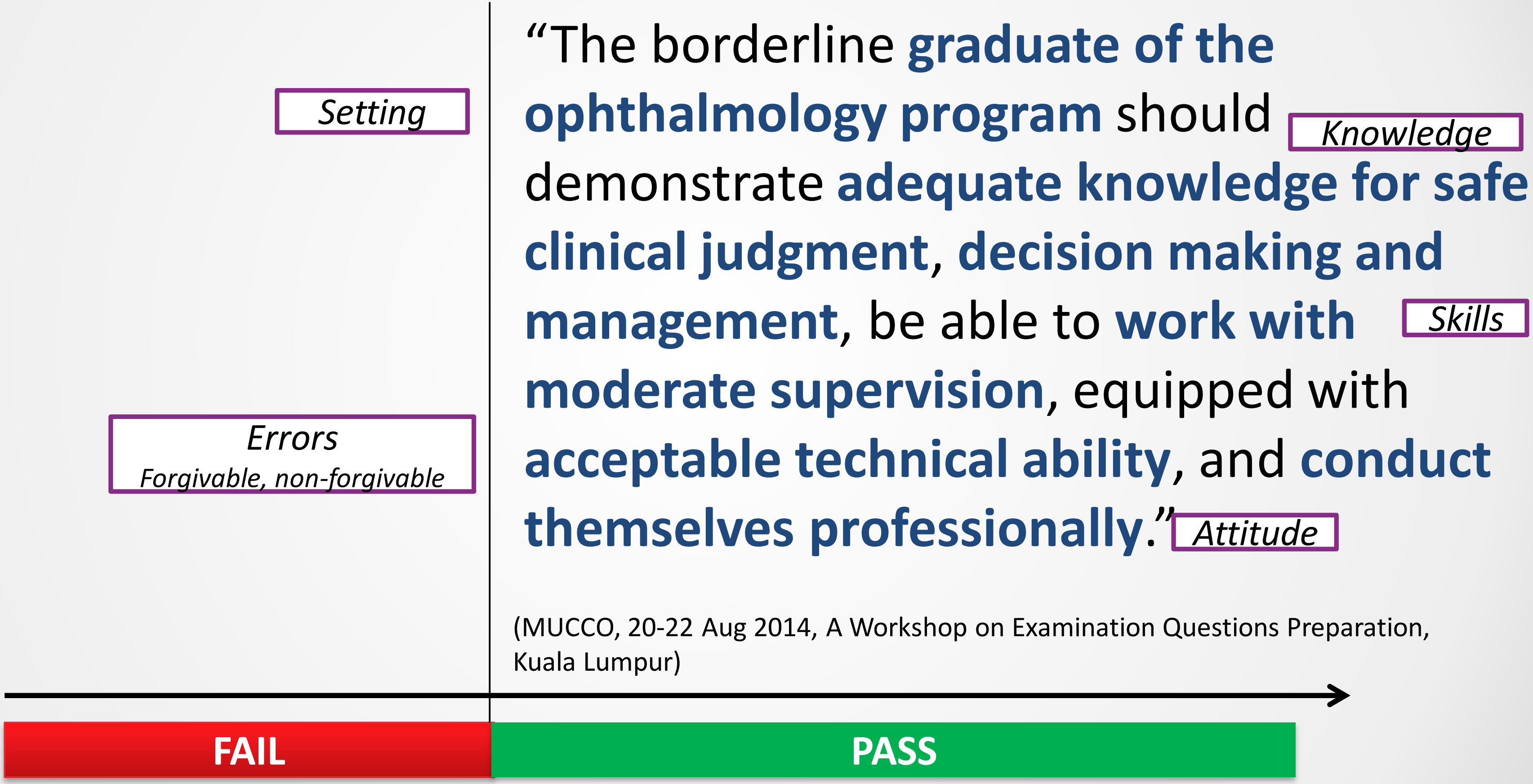
#1 Setting – e.g. graduate of the ophthalmology program

#3 Skills – e.g. be able to work with moderate supervision, equipped with acceptable technical ability

#5 Errors (considering the forgivable or unforgivable) – e.g. safe clinical judgment, decision making and management

(Mills, Melican & Ahluwalia, 1991)

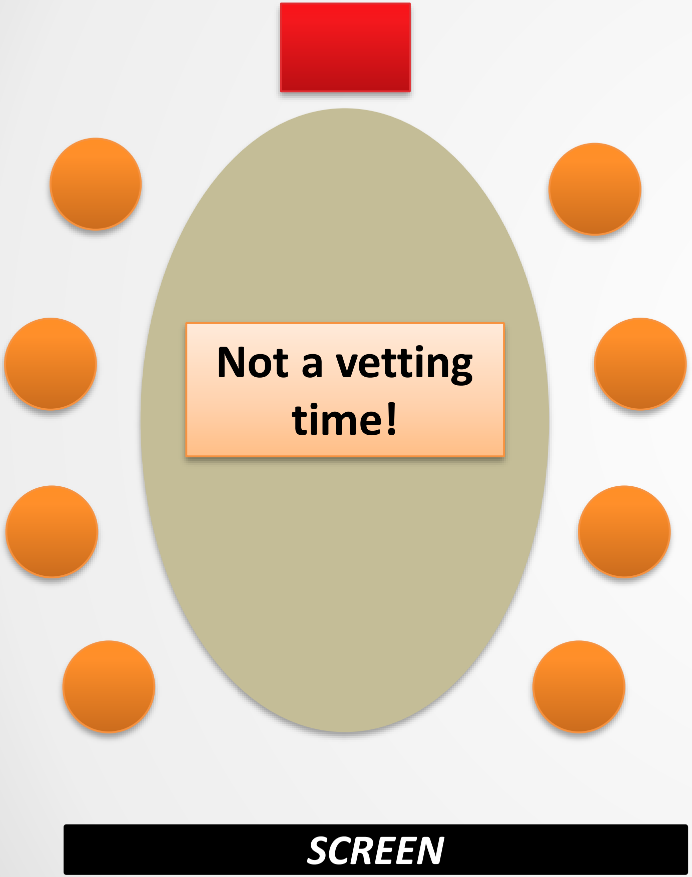
STANDARD SETTING: Borderline Standards



"The borderline **graduate of the ophthalmology program** should demonstrate **adequate knowledge for safe clinical judgment**, **decision making and management**, be able to **work with moderate supervision**, equipped with **acceptable technical ability**, and **conduct themselves professionally**."

(MUCCO, 20-22 Aug 2014, A Workshop on Examination Questions Preparation, Kuala Lumpur)

STANDARD SETTING: Modified Angoff - DURING



Read through question 1

Judges: Individually, estimate the mark that can be obtained by borderline examinees for question 1

Moderator: Record ratings

Moderator: Discuss ratings

Moderator: Get 2nd ratings after discussion

Calculate mean

Repeat for next questions

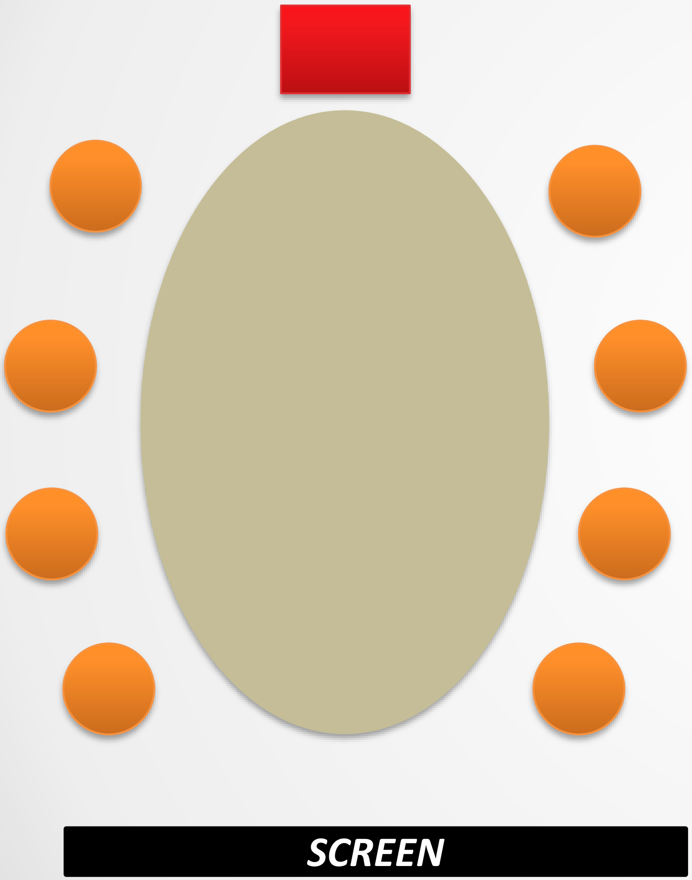
(Cizek, 2006; Angoff, 1971)

	Q1	Q2	Q3	Q4	Q5	Mean
Total Mark	10					
JUDGE 1	6					
	6					
JUDGE 2	5					
	6					
JUDGE 3	**9**					
	6					
JUDGE 4	6					
	5					
JUDGE 5	6					
	6					
JUDGE 6	**4**					
	6					
Mean 1st	**6**					
Mean 2nd	**5.83**					

Cut-off score 1st round

Cut-off score 2nd round

STANDARD SETTING: Modified Angoff - POST

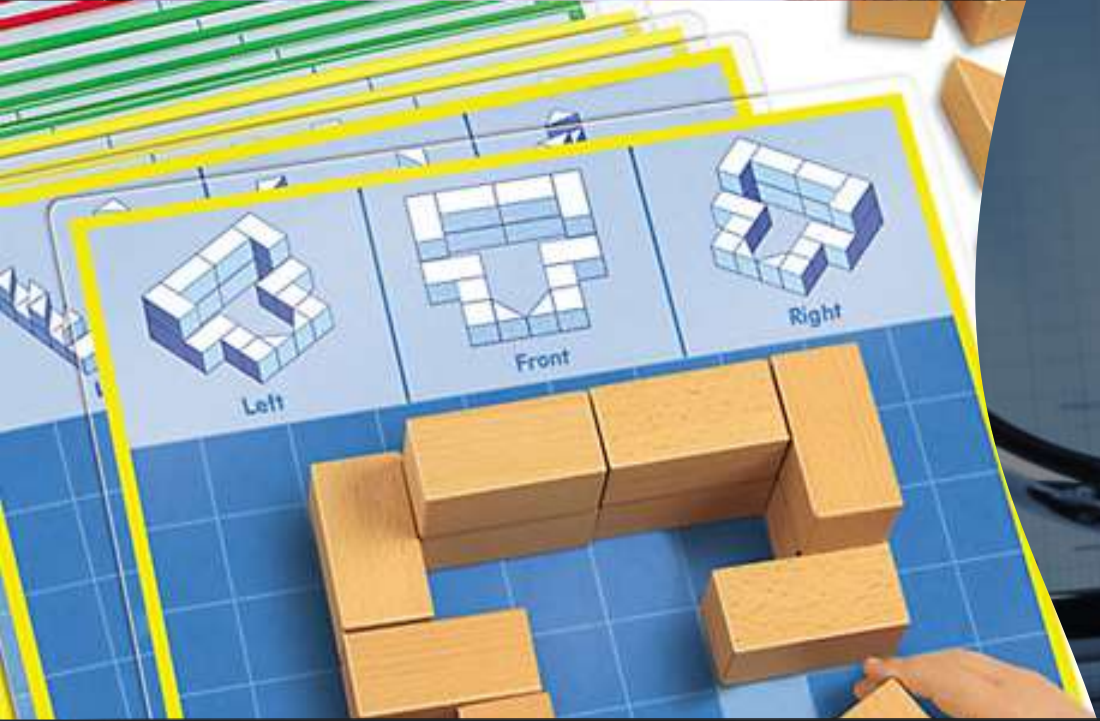


Evaluate the process

- Judges confidence in the process
- Resulting cut off scores

Documentation

(Cizek, 2006; Angoff, 1971)

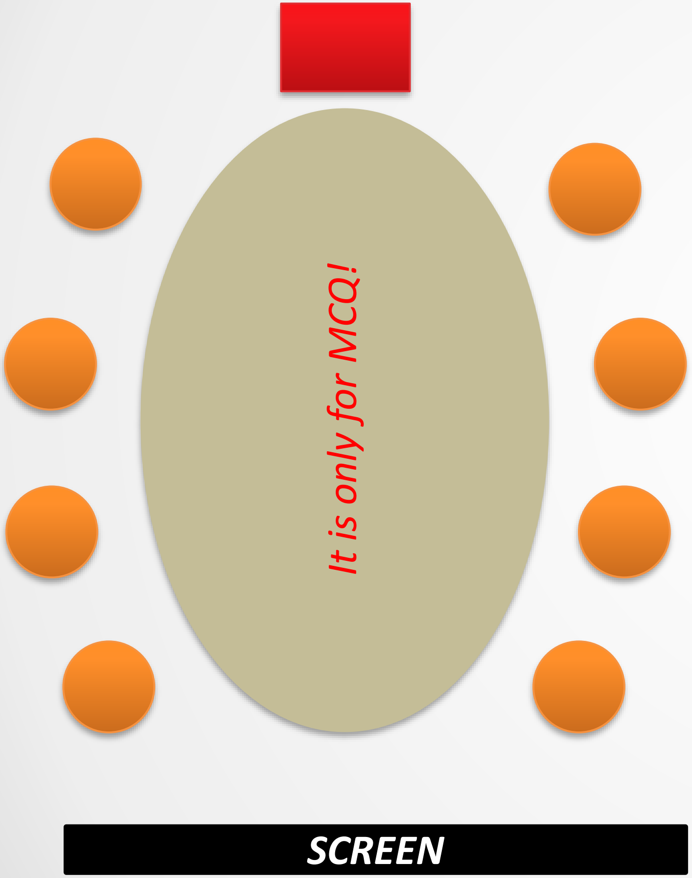


GROUP C

STANDARD SETTING

NEDELSKY

STANDARD SETTING: Nedelsky - PRE



Select the judges

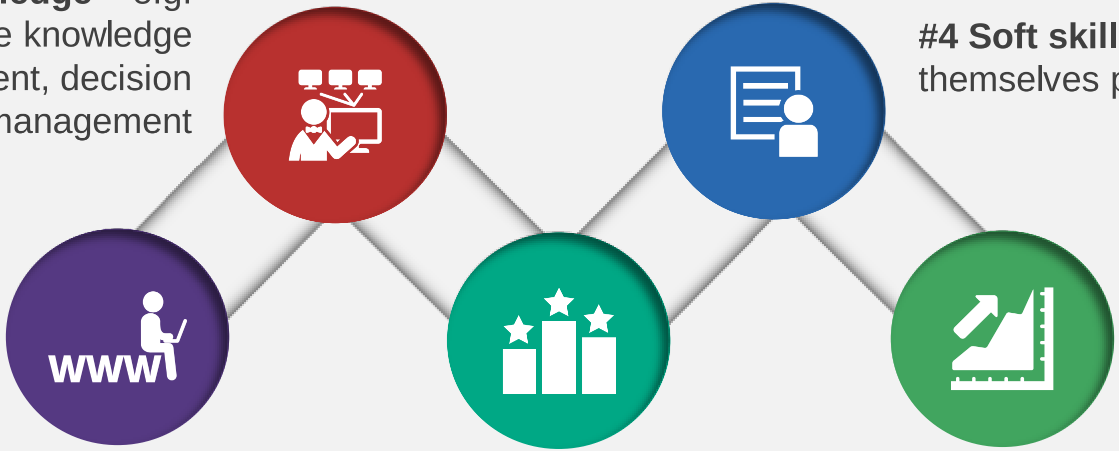
Discuss
a. Purpose of the assessment
b. Nature of examinees
c. Components of adequate/inadequate knowledge

Select the methods – train judges

Define borderline standard

(Cizek, 2006; Nedelsky, 1954)

STANDARD SETTING: Define Borderline



#1 Setting – e.g. graduate of the ophthalmology program

# 2 Knowledge – e.g. demonstrate adequate knowledge for safe clinical judgment, decision making and management

#3 Skills – e.g. be able to work with moderate supervision, equipped with acceptable technical ability

#4 Soft skills - e.g. conduct themselves professionally

#5 Errors (considering the forgivable or unforgivable) – e.g. safe clinical judgment, decision making and management

(Mills, Melican & Ahluwalia, 1991)

STANDARD SETTING: Borderline Standards

Setting

Knowledge

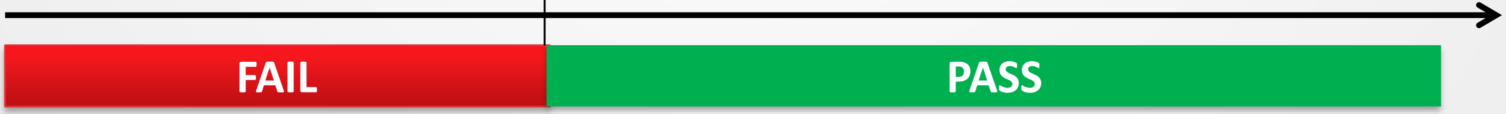
Errors
Forgivable, non-forgivable

"The borderline **graduate of the anaesthesiology program** should demonstrate **adequate knowledge for safe clinical judgment**, **decision making and management**, be able to **work with minimal supervision**, equipped with **acceptable life saving skills and technical ability**, and **conduct themselves professionally**."

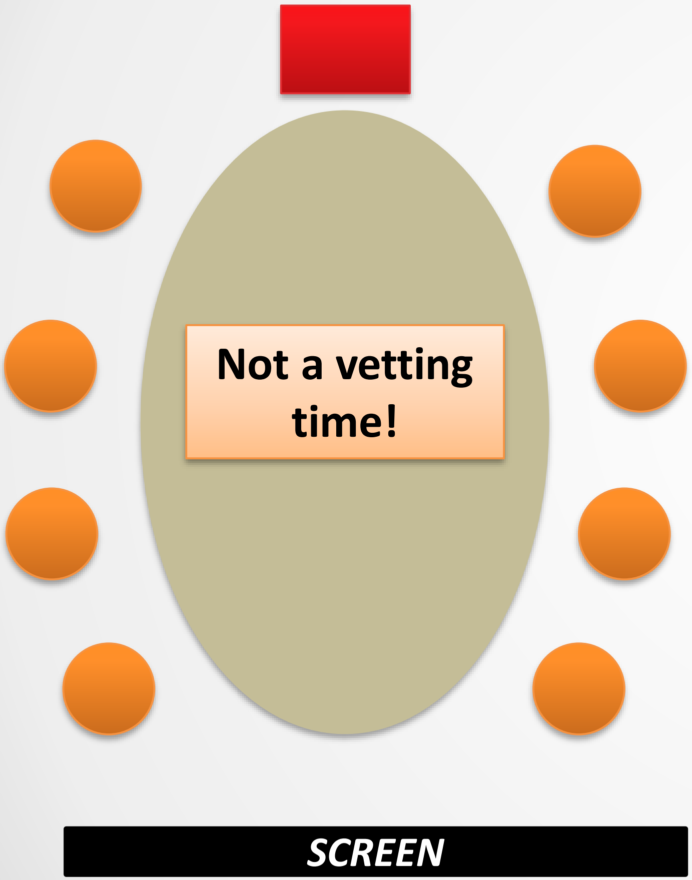
Skills

Attitude

(8 Jan 2022, A Workshop on Standard Setting (Anaesthesiology) Workshop, UPM, Selangor)



STANDARD SETTING: Nedelsky - DURING



Read through each question

Judges: Working individually, judges mark the wrong answers the borderline students would be able to eliminate.

Moderator: Record ratings

Moderator: Discuss and change ratings

Repeat for next questions

Calculate passing score

(Cizek, 2006; Nedelsky, 1954)

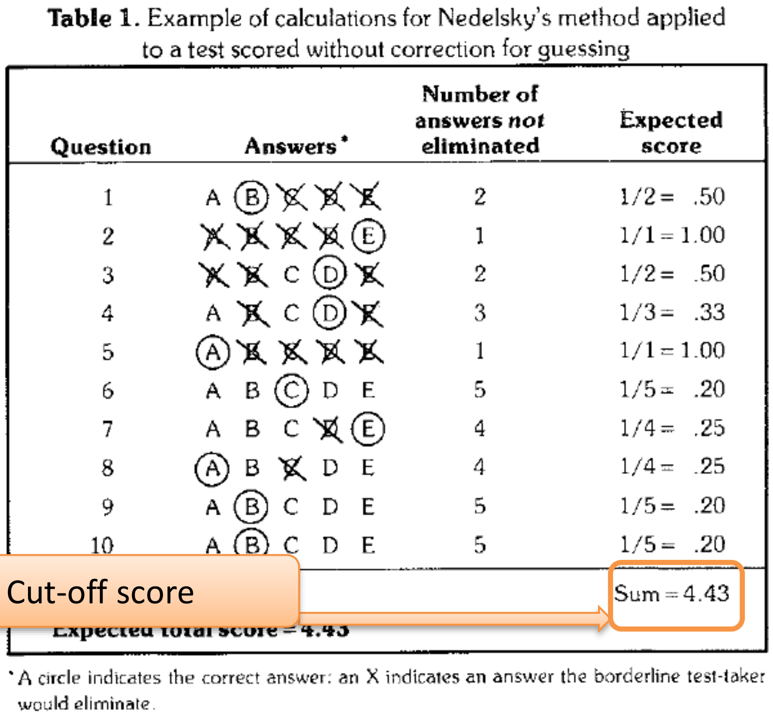


Table 1. Example of calculations for Nedelsky's method applied to a test scored without correction for guessing

Question	Answers*	Number of answers *not* eliminated	Expected score
1	A [✓ B] ~~C~~ ~~D~~ ~~E~~	2	1/2 = .50
2	~~A~~ ~~B~~ ~~C~~ ~~D~~ [✓ E]	1	1/1 = 1.00
3	~~A~~ ~~B~~ C [✓ D] ~~E~~	2	1/2 = .50
4	A ~~B~~ C [✓ D] ~~E~~	3	1/3 = .33
5	[✓ A] ~~B~~ ~~C~~ ~~D~~ ~~E~~	1	1/1 = 1.00
6	A B [✓ C] D E	5	1/5 = .20
7	A B C ~~D~~ [✓ E]	4	1/4 = .25
8	[✓ A] B ~~C~~ D E	4	1/4 = .25
9	A [✓ B] C D E	5	1/5 = .20
10	A [✓ B] C D E	5	1/5 = .20
			Sum = 4.43

Expected total score = 4.43

*A circle indicates the correct answer; an X indicates an answer the borderline test-taker would eliminate.

Cut-off score

- Three methods of calculating passing score:
  - Mean
  - Median
  - Trimmed mean

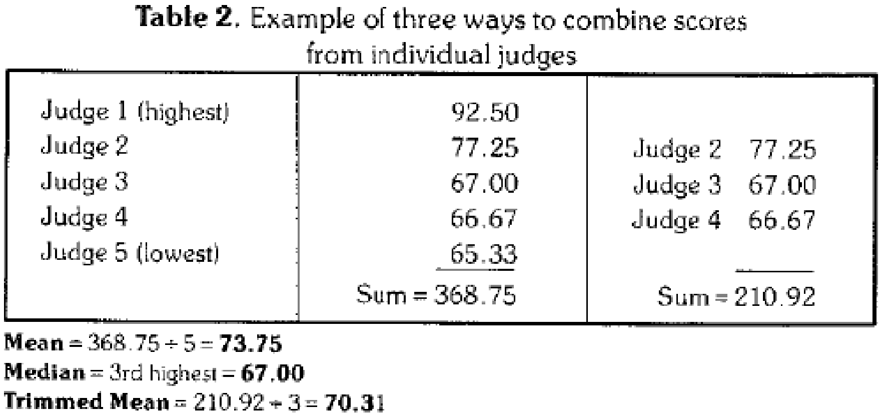
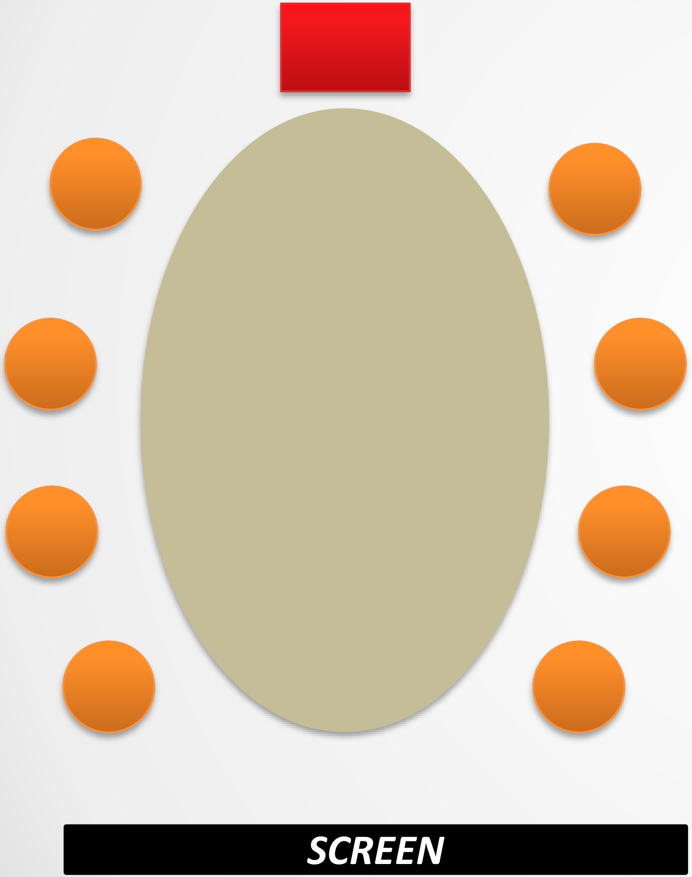


Table 2. Example of three ways to combine scores from individual judges

Judge 1 (highest)	92.50	
Judge 2	77.25	Judge 2 77.25
Judge 3	67.00	Judge 3 67.00
Judge 4	66.67	Judge 4 66.67
Judge 5 (lowest)	65.33	
	Sum = 368.75	Sum = 210.92

Mean = 368.75 ÷ 5 = **73.75**
Median = 3rd highest = **67.00**
Trimmed Mean = 210.92 ÷ 3 = **70.31**

STANDARD SETTING: Nedelsky - POST



Evaluate the process

- Judges confidence in the process
- Resulting cut off scores

Documentation

(Cizek, 2006; Nedelsky, 1954)