

Medical Education Journal Club

ACCREDITATION STATEMENT

Florida International University Herbert Wertheim College of Medicine is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians.

CREDIT DESIGNATION STATEMENT

Florida International University Herbert Wertheim College of Medicine designates this live activity for a maximum of 1.0 *AMA PRA Category 1 Credit(s)*[™]. Physicians should claim only the credit commensurate with the extent of their participation in the activity.



Disclosure Information

Activity Directors / Planners / Reviewers / Faculty

Name	Role	Disclosure / Resolution
Carla S. Lupi, MD	Activity Director/Planner/ Speaker	Dr. Lupi reports no relevant financial relationships.
Vivian Obeso, MD	Planner/Speaker	Dr. Obeso reports no relevant financial relationships.
Christian Castro	Planner	Mr. Castro reports no relevant financial relationships.
Melissa Ward-Peterson, MPH	Planner	Ms. Ward-Peterson reports no relevant financial relationships.



Herbert Wertheim
College of Medicine

Learning Objectives

- List the search results for one medical education database consulted in the design of a course or teaching session.
- Conduct a critical appraisal of an article in the medical education research.
- Identify the applicability of research results to one's own course or teaching session.

Medical Education Journal Club

- Establish a forum for faculty to share and discuss recent literature in medical education
- Use best evidence in medical education literature to evaluate and advance current practices in our educational program
- Establish a culture that promotes curricular innovation and change in an evidence-based manner
- Stimulate educational scholarship

Objectives for Today's Session

- Discuss the importance of assessing insight and foresight in medical trainees
- Review one set of modifications to the MCQ format for assessing insight and foresight
- Identify the challenges of validating an assessment of insight and foresight



Today's Article

Determining measures of insight and foresight from responses to multiple choice questions

MIKE TWEED, MARK THOMPSON-FAWCETT, PETER SCHWARTZ & TIM J. WILKINSON

University of Otago, New Zealand

Abstract

Background: Clinical decisions require insight and foresight. When these are lacking, overconfidence and error can occur.

Aims: Multiple choice questions (MCQ) responses were used to estimate insight, as determined by confidence in responses, and foresight, as determined by the safety of incorrect responses.

Methods: An MCQ response system was developed which reflected confidence in, and safeness of, responses. An optional examination incorporating this system was offered to medical students at one university.

Results: A total of 372 students completed the study. There was evidence of insight. The proportion of responses that were correct

Tweed M, Thompson-Fawcett M, Schwartz P, Wilkinson TJ. Determining measures of insight and foresight from responses to multiple choice questions. *Med Teach*. 2013;35(2):127-33.



Herbert Wertheim
College of Medicine

Insight? Foresight? MCQs – are you kidding?

BACKGROUND AND BASIC DEFINITIONS

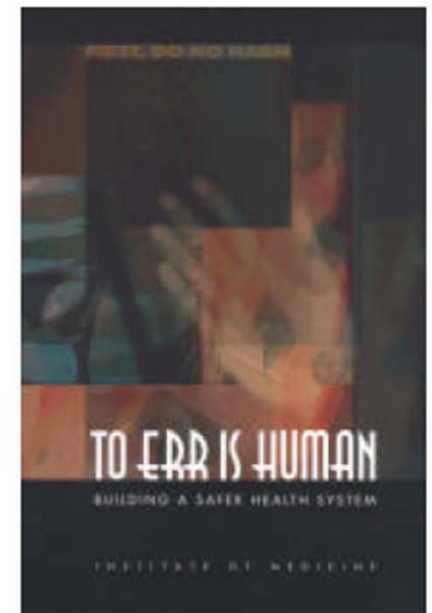
November 1999

INSTITUTE OF MEDICINE

Shaping the Future for Health

TO ERR IS HUMAN: BUILDING A SAFER HEALTH SYSTEM

Health care in the United States is not as safe as it should be--and can be. At least 44,000 people, and perhaps as many as 98,000 people, die in hospitals each year as a result of medical errors that could have been prevented, according to estimates from two major studies. Even using



Diagnostic Error = Big Problem



- 28.6% of paid medmal claims
- 68.8% outpatient vs. 31.2% inpatient
- 2x rate of death or disability when compared to other error categories
- 15% in patients w/new problem

Newman-Toker, DE. BMJ Quality and Safety, 2013.



Educating on Diagnostic Error @ FIUCOM

LEARNING

- OF Period 3
- CCM Period 2
- Self-awareness - several

Opportunities:

Metacognition - PB

Value of Excellence – PB

DxRx Cases; HCQI?

Other toe dips?

ASSESSING

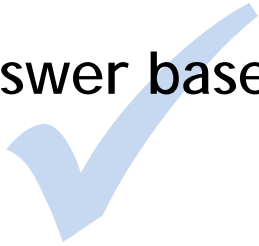


FIU

**Herbert Wertheim
College of Medicine**

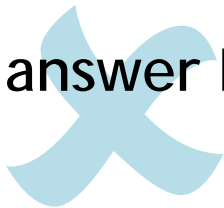
Answer to MCQs: How does the examinee get there?

Right Answer based on...



- Solid knowledge and reasoning
- A guess (scoring system encourages risk/test wising)
- Incorrect reasoning

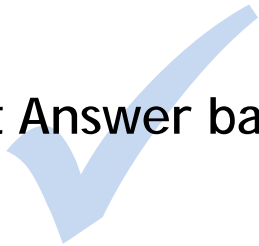
Wrong answer based on...



- Incorrect reasoning
- A guess (scoring system encourages risk/test wising)

Possible sources for answers to an MCQ

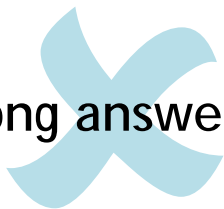
Right Answer based on...



- Solid knowledge and reasoning
- A guess (scoring system encourages risk/test wising)
- Inappropriate reasoning

Is this real life?

Wrong answer based on...



- Incorrect reasoning
- A guess (scoring system encourages risk/test wising)

Possible sources for answers to questions

Right Answer based on...

- Solid knowledge and reasoning
- A guess (scoring system encourages risk/test wising)
- Incorrect reasoning

Wrong answer based on...

- Incorrect reasoning
- A guess (scoring system encourages risk/test wising)

No option for consultation or confirmation

Insight

Understanding of self and one's deficiencies

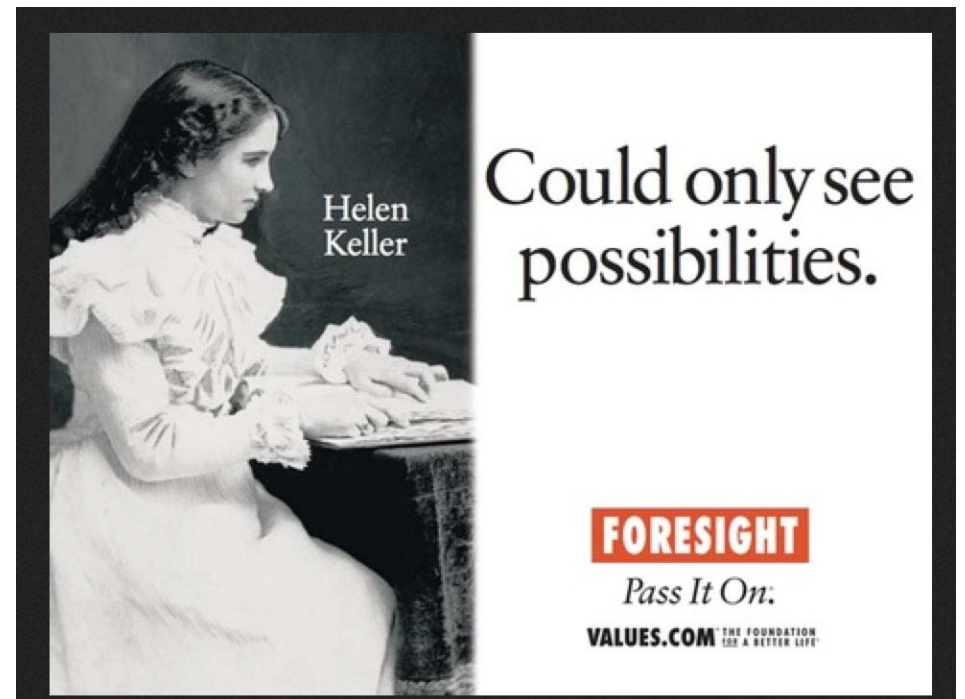
What's the likelihood that I am correct



Foresight

Foreseeability for duty of care

What are the risks and benefits of my choice

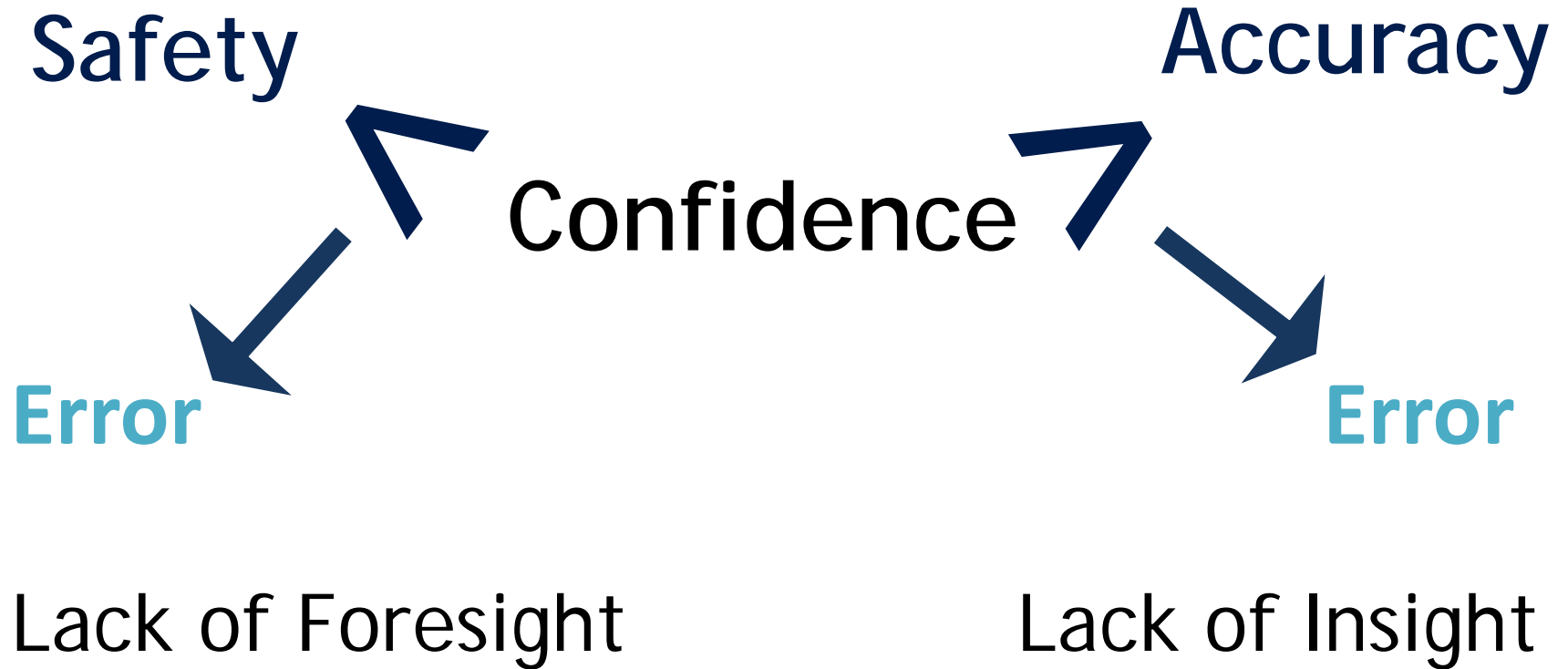


Confidence > Accuracy



Error

Insight



Research Questions

- Can we measure students' levels of insight?
- Can we measure students' levels of foresight?
- How do these measures of insight and foresight relate to students' levels of experience?
- How do these measure of insight and foresight relate to students' levels of knowledge?

METHODS

Methods

- Subjects: 372 student volunteers yr 2-5
- New Zealand 6 yr medical school
- Extended matching (9 option) MCQs from school data bank modified with 10th option:
 - – “I do not know”
- 12 questions per student
- Instructions to students:
 - *use “do not know” (rather than not answering)
 - *rate your confidence level for each question

Methods: What about the MCQs?

- Extended matching, from a question bank had been used for one year only
- Reviewed by “academic staff”
- Not all questions had unsafe responses
- Safety determinations: expert clinician panel, the median of all the expert opinions

Methods: Confidence Rating Scale

	Certainty			
	None	Low	Moderate	High
Related to knowledge	The candidate has no idea of correct response and any response would be a guess. <i>Instructed to answer 'Do Not Know'</i>	The candidate has no clear idea of correct response but has some knowledge on the subject. Any response would be based on limited information.	The candidate has a reasonable idea of correct response on a basis of moderate knowledge on the subject. Any response would be based on sufficient information	The candidate is certain of correct response on a basis of detailed knowledge on the subject. Any response would not be a guess.
Related to practice	The candidate would need to consult a colleague or references prior to considering any response.	The candidate would need to consult a colleague or references but would be able to give a response first.	The candidate would need to consult a colleague or references to confirm the correctness of the response.	The candidate would have no need to consult a colleague or reference.



Methods: MCQ example

A 70 year old man presents with chest ache and breathlessness that came on for the first time today. This occurred whilst walking. The ache is located at the front of his chest and is persistent. It has never occurred before. He is normally well and takes no medication. There is no family history of cardiac, pulmonary or thromboembolic disease.

On examination oxygen saturations are 96% breathing 40% oxygen but drop quickly when breathing air. Pulse 120 per minute and regular. Blood pressure 80/60 with no abnormal pulsus paradoxus. Cardiac auscultation confirms tachycardia and no murmurs. Chest auscultation is clear.

ABG, ECG and CXR are done. Results provided.

From the options provided, what is the most likely diagnosis? Choose one response only.

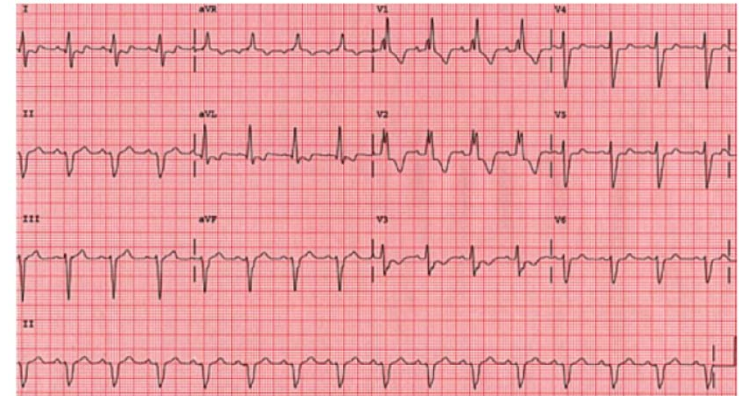
1. Acute asthma
2. Acute coronary syndrome
3. Dissection thoracic aortic aneurysm
4. Hyperventilation syndrome
5. Oesophageal spasm
6. Pericardial effusion
7. Pneumothorax
8. Pulmonary embolism
9. Viral pneumonia
10. I do not know and would need to consult a colleague or reference

If you did not choose option 10, how certain are you in this choice?

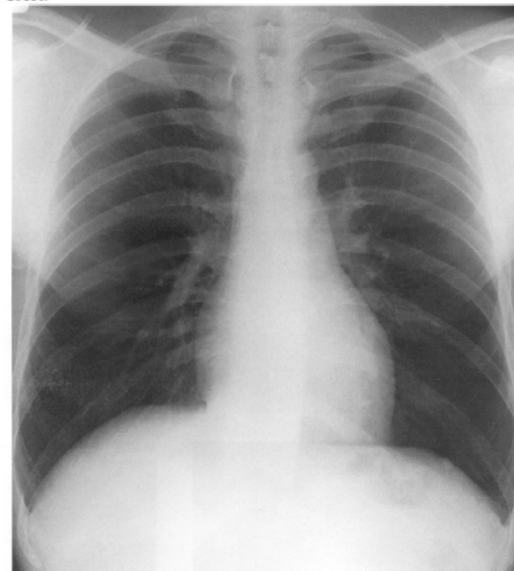
Low Moderate High

Arterial blood gas breathing 40% oxygen:
pH 7.44 (reference 7.35-7.45)
PaCO₂ 28 mmHg (reference 35-45)
PaO₂ 120 mmHg. (reference 80-110)

ECG:



CXR:



Methods:
Validity concerns at this point?

Option	Expert 1	Expert 2
1. Acute Asthma	Moderately Unsafe	Highly Unsafe
2. Acute Coronary Syndrome	Not Unsafe	Highly Unsafe
3. Dissecting aortic aneurysm	Not Unsafe	Highly Unsafe
4. Hyperventilation Syndrome	Highly Unsafe	Highly Unsafe
5. Oesophageal Spasm	Highly Unsafe	Highly Unsafe
6. Pericardial Effusion	Low Unsafe	Highly Unsafe
7. Pneumothorax	Moderately Unsafe	Highly Unsafe
8. Pulmonary Embolism	Not Unsafe	Not Unsafe
9. Viral Pneumonia	Not Unsafe	Highly Unsafe
10. I do not know and would need to consult a colleague or reference		

RESULTS WITH DISCUSSION

Subjects: 372 Volunteers

Year	Percent of Class
2	37.5
3	35.8
4	22.7
5	56.3

Answered a total of 4464 questions!

From Table 3

Category	N (% of total questions)
Correct	1901 (42%)
Incorrect, not unsafe	915 (20.5%)
Incorrect and unsafe	1054 (23.8%)
Do not know	594 (13.31%)
Total	4464

Table 3 - Insight

Response	Confidence				Total
	None	Low	Moderate	High	
Correct	481	641	779	1901	
	10.77%	14.36%	17.45%	42.58%	
Incorrect, not unsafe	498	338	79	915	
	11.16%	7.57%	1.77%	20.50%	



**Herbert Wertheim
College of Medicine**

Table 3 – this is good for foresight

confidence

Safety	Incorrect and unsafe			
	Low	194 4.34%	159 3.56%	100 2.24%
	Moderate	147 3.29%	69 1.55%	21 0.47%
	High	193 4.32%	134 3.00%	37 0.82%
		35%	27%	16%



Table 3: this is *not* good for foresight -of ALL the incorrect responses....

		confidence		
Safety	Incorrect, not unsafe	498 11.16%	338 7.57%	79 1.77%
	Incorrect and unsafe			
	Low	194 4.34%	159 3.56%	100 2.24%
	Moderate	147 3.29%	69 1.55%	21 0.47%
	High	193 4.32%	134 3.00%	37 0.82%
% total unsafe at all:		52%	52%	68% ***???
% total mod/high unsf:		32%	29%	24%

The most concerning – table 4 unsafe + high confidence

Number of responses	Number of candidates with high confidence in any unsafe responses	Number of candidates with high confidence in highly unsafe responses
0	250	341
1	92	26
2	25	4
3	4	1
4	1	0

Note: There were a small number of candidates who had several highly unsafe responses held with a high degree of confidence.

table 5

Table 5. Effect of experience on insight: % responses correct and % responses incorrect and not unsafe for different levels of confidence and year group.

	Confidence			ANOVA, <i>p</i>
	Low	Moderate	High	
<i>Responses correct (%)</i>				
Year 2	23.87	38.87	53.05	<0.001
Year 3	31.50	47.16	57.49	<0.001
Year 4	37.82	47.33	78.07	<0.001
Year 5	36.05	53.82	84.81	<0.001
ANOVA, <i>p</i>	0.006	0.008	<0.001	
<i>Responses incorrect and not unsafe (%)</i>				
Year 2	34.52	20.11	7.45	<0.001
Year 3	31.40	24.41	11.45	<0.001
Year 4	31.91	23.95	7.87	<0.001
Year 5	36.05	27.32	4.53	<0.001
ANOVA, <i>p</i>	0.681	0.239	0.058	

Table 6 – if you were designing an intervention program, where would you intervene???

	Confidence			ANOVA, <i>p</i>
	Low	Moderate	High	
Year 2	41.62	41.02	39.50	0.934
Year 3	37.10	28.43	31.06	0.128
Year 4	30.27	28.72	14.07	0.006
Year 5	27.90	18.86	10.65	<0.001
ANOVA, <i>p</i>	0.003	<0.001	<0.001	

Table 8

% of incorrect that are unsafe held with				
For students with total correct of	Low confidence	Moderate confidence	High confidence	ANOVA, <i>p</i>
0–3	36.98	17.45	5.97	0.497
4–6	23.12	17.26	9.59	0.059
7–9	16.17	21.94	9.50	<0.001
10–11	17.86	3.57	17.86	0.057
ANOVA, <i>p</i>	<0.001	<0.001	<0.001	

In Summary

Strengths of design

- Robust Ns permitting meaningful statistical evaluation
- Validity through correlation w/ knowledge & experience
- Validity through use of expert panel

Limitations of design

- Safety determination process and data not well described
- No other validation data on individual MCQ items
- Not a summative assessment, so what if?

Norcini's Criteria for Good Assessment

- **Validity**: supports the use of the results for a particular purpose
- **Reproducibility**: yields same scores under similar circumstances
- **Equivalence**: assessment yields equivalent scores or decisions when administered across different institutions or cycles of testing
- **Feasibility**
- **Educational effect**: motivates to prepare in a way that has educational value
- **Catalytic effect**: provides results and feedback that creates, enhances & supports education; drives future learning forward
- **Acceptability**: stakeholders find process and results credible

Norcini J, Anderson B, Bollela V, Burch V. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach. 2011;33(3):206-14.

What else?

- Correlation with OSCE derived data – for the impact of patient interaction, and????
- Reproducibility, and in summative use
- Equivalence – across institutions, etc.
- Educational and Catalytic effects on measures of metacognition, diagnostic accuracy and culture of patient safety



Email response from Dr. Tweed:

- “Also longer term I am going to be looking for collaborations on extending this out beyond our institution if you would be interested.”
- “All panel members make a judgment on all incorrect options. In phase 3, we had a group discussion on the concept of unsafeness, then some practice questions, but then the counting judgments are still independent.”



Herbert Wertheim
College of Medicine

Questions?

Please complete the CME evaluation
to receive credit for attendance.