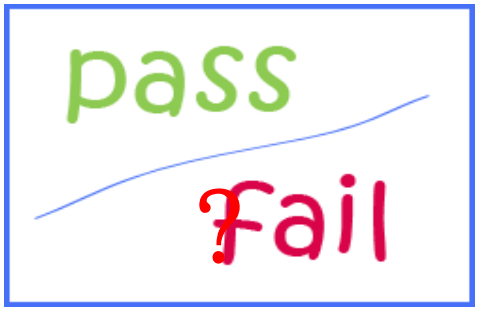
Making defensible Pass/Fail decisions by using student ability and item difficulty

Never Stand Still

Boaz Shulruf [1], Philip Jones [1], Rolf Turner[2]

[1] Office of Medical Education, University of New South Wales, Sydney, Australia
[2] Department of Statistics, University of Auckland, Auckland, New Zealand



ISPA Tel Aviv 25 January 2016

b.shulruf@unsw.edu.au

The Challenges

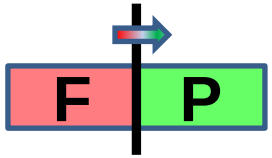
How to make a pass/fail decisions over borderline grades that are:

- Objective and reliable
- Fair and acceptable
- Statistically robust
- Not a norm based (does not force minimum % of failure)
- Minimally susceptible to post examination judgement
- Relevant to the learned curriculum
- Relevant to assessed student population
- Consistent across cohorts and populations
- Not susceptible to extreme scores
- Pass/Fail decision criteria could be known prior the exams
- Simple to use and feasible

b.shulruf@unsw.edu.au

Examinee's Ability & Item Easiness

- Assuming there is an examination with M items and N examinees
- On each item one could get either Pass or Fail
- Assuming all items are unidimensional, then based on IRT:
- Examinee's Ability is: $Ln_{(Ab)}\left(\frac{Pr(\text{correct items})}{Pr(\text{incorrect item}}\right)$
- and item Easiness is: $Ln_{(Eas)}\left(\frac{Pr(\text{correct items})}{Pr(\text{incorrect item}}\right)$



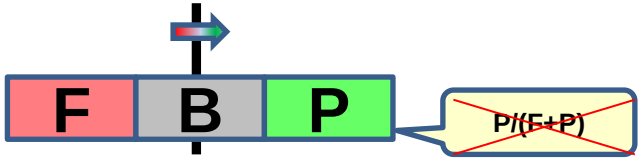
F=number of Fails; **P**=number of Passes

- If **Ln(Ab)=Ln(Eas)**, then the probability of an examinee giving a correct answer to an item is similar to the probability of that item being correctly answered by a randomly selected examinee.
- If **Ln(Ab)>Ln(Eas)**, then the probability of an examinee giving a correct answer to an item is greater than the probability of that item being correctly answered by a randomly selected examinee (and vice versa if Ln(Ab)<Ln(Eas).
- Thus, if there is uncertainty whether an answer to an item is correct (e.g. a Borderline answer), the difference between examinee's Ability and item Easiness can be used to make Pass/Fail decision.
- Since the focus is on 'qualitative' comparison between two indices both on the same scale, there is no need to apply Logit transformation. Comparison of the probabilities (proportions) of Passes would suffice. i.e. **P/(F+P)**

b.shulruf@unsw.edu.au

The principles of the OBM

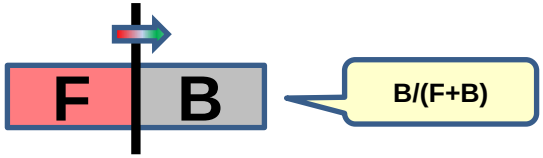
- How can we estimate item Easiness and examinee's Ability if there are three grades: Pass, Borderline and Fail?



F=number of Fails; **P**=number of Passes;
B=number of Borderlines

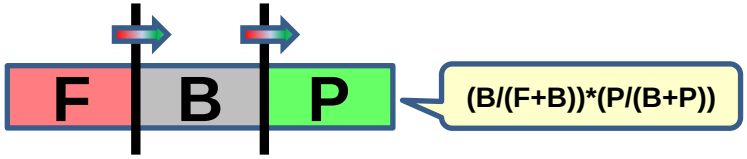
- We can measure examinee's Ability and item Easiness by using two separate indices rather than one:

- Ability of achieving **Borderline** rather than **Fail**
- Easiness of achieving **Borderline** rather than **Fail**



- Ability of achieving **Pass** rather than **Borderline**
- Easiness of achieving **Pass** rather than **Borderline**

- The **indices** are then combined together by multiplication to give the product:



OBM Pass Index = (B/(F+B))*(P/(B+P))

b.shulruf@unsw.edu.au

UNSW AUSTRALIA

The principles of the OBM2

- $Ability_{(comb)}$ and $Easiness_{(comb)}$ are on the same scale hence comparable

- For each Borderline item it is possible to calculate $Ability_{OBM}$ and $Easiness_{OBM}$
- If the $Ability_{OBM} > Easiness_{OBM}$, then the probability of an examinee giving a correct answer is greater than the probability of that item being correctly answered by a randomly selected examinee (and vice versa if $Ability_{OBM} < Easiness_{OBM}$).

Examples:

- The distribution of grades (all examinees) of a particular item was F=14, B=11, P=175
- **Item Easiness:** (B/(F+B))*(P/(B+P))= (11/(14+11))*(174/(11+175))=**.41**

- Examinee A summary grades of 9 criteria on one OSCE station F=1, B=3, P=5
- **Examinee A Ability:** (B/(F+B))*(P/(B+P))= (3/(1+3))*(5/(3+5))=**.47** **.47>.41 Pass**

- Examinee B summary grades of 9 criteria on one OSCE station F=2, B=3, P=4
- **Examinee B Ability:** (B/(F+B))*(P/(B+P))= (3/(2+3))*(4/(3+4))=**.34** **.34<.41 Fail**

b.shulruf@unsw.edu.au

Comparing the conventional vs. OBM2 method

- The Phase 1 clinical skills examination comprises six standardised stations
- For each station a student can get one of the following grades:
  - Fail, Borderline, Pass, Exceptional
  - Each grade is also converted to a numerical score (with Borderline representing 50%)
- Criteria for passing the examination:
  - students must pass at least three stations
  - Overall (mean) score for the six stations must be ≥50%

- The quality of the OBM2 was estimated by comparison of the overall clinical examination grades in the Phase 2 clinical examination with the overall outcomes of the Phase 1 clinical examination as calculated in two ways:

- Passing by the original method
- Modifying each Borderline grade to either Pass of Fail using the OBM2 and then making the Pass Fail decision using the original criteria

- Outcomes of the two methods were compared with results of clinical examination Phase2
- Overall data of 1136 students were included in the analysis,

b.shulruf@unsw.edu.au

Comparing the conventional vs. OBM2 method

		Phase 2		
		Pass	Fail	
Original decision	Pass	1056	71	Original decision
	Fail	4	5	
OBM2	Pass	945	50	OBM2
	Fail	115	26	

Index	Original/actual	OBM2
Accuracy (Overall Fraction Correct)	0.93	0.84
Sensitivity	1.00	0.88
Specificity	0.07	0.35

Results and interpretations

- The OBM2 provided more stringent outcomes than the original decision model (i.e. no standard setting method applied)
- Progression to Phase 2 was made based on the original hence it is expected to be more accurate
- OBM2 better identified students at risks

b.shulruf@unsw.edu.au

Conclusions

- The OBM2 attributes:
  - simple
  - statistically robust
  - not affected by extreme scores
  - based on commonly accepted concepts of item difficulty (easiness) and student ability
  - Does not require post examination judgement

- The OBM2 is a defensible tool to assist pass/fail decisions over Borderline grades.

b.shulruf@unsw.edu.au

References and further readings on the OBM

- Shulruf, B., Booth, R., Baker, H., Bagg, R., & Barrow, M. (in press). Using the objective borderline method (OBM) to support board of examiners' decisions in a medical programme. *Journal of Further and Higher Education*
- Shulruf, B., Jones, P., & Turner, R. (2015). Using student ability and item difficulty for standard setting. *Higher Education Studies, 5*(4), 106-118. doi: 10.5539/hes.v5n4p
- Shulruf, B., Poole, P., Jones, P., & Wilkinson, T. (2014). The objective borderline method (OBM): A probabilistic method for standard setting *Assessment and Evaluation in Higher Education* doi: 10.1080/02602938.2014.918088
- Shulruf, B., Turner, R., Poole, P., & Wilkinson, T. (2013). The objective borderline method (OBM): A probability-based model for setting up an objective pass/fail cut-off score for borderline grades in medical education programmes. *Advances in Health Sciences Education, 18*(2), 231-244. doi: 10.1007/s10459-012-9367-y

b.shulruf@unsw.edu.au

UNSW AUSTRALIA