Development and implementation of SJTs for health sciences selection – preliminary findings and current challenges

Research Team:

Prof. Margaret Hay

Dr Irene Lichtwark

Mr Sam Henry

SJT Workshop, University of Sydney, July 2017

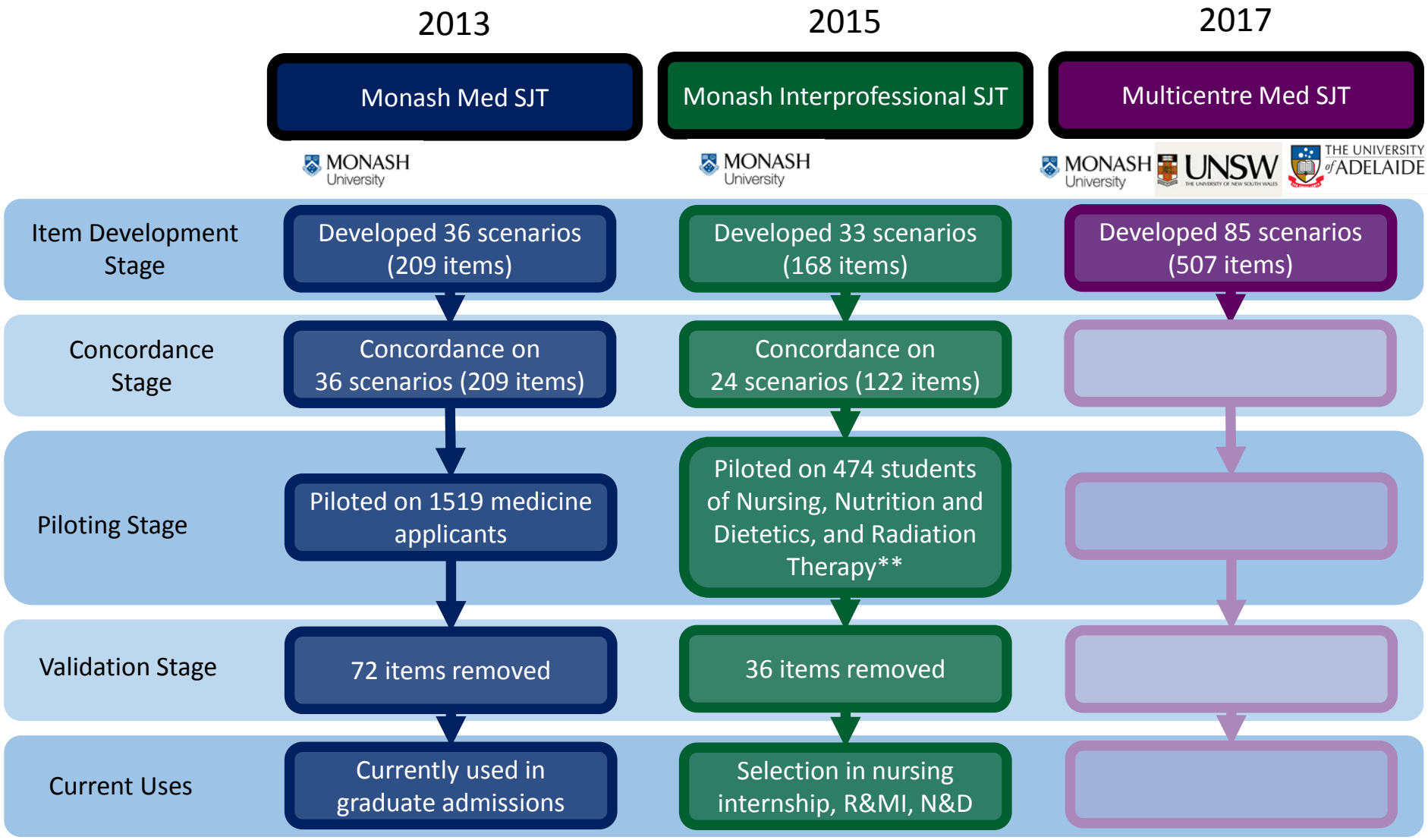
Overview

- SJT development and implementation at Monash
- The psychometric qualities of the different SJTs
- Scoring challenges for the SJT

Funding Acknowledgement:
- UMAT Consortium
- Monash University MNHS

The SJT Development Research Program

- Aim: To develop Situational Judgement tests (SJT) for selection into health professional courses



**355 UG/Graduate Nurses, 107 Nutrition and Dietetics students, 12 Radiation Therapy students

10 Steps of SJT development

Step	Description of task
1	Determine test specifications, including domains
2	SJT development workshops with discipline representatives (DRs)
3	Scenario/item review
4	Concordance workshop with different DRs
5	Test construction, development of scoring matrix, transcript into Qualtrics
6	Piloting
7	Psychometric analysis of pilot data, exclusion of low performing items
8	Reliability testing
9	Continuing development (annual SJT development and concordance workshop)
10	Annual SJT review (psychometric analysis of existing and new SJTs after each admission cycle)

SJT domains

Monash Med

- Integrity and ethical reasoning
- Empathy
- Collaboration

Interprofessional SJT

- Integrity and ethical reasoning
- Empathy
- Collaboration
- Resilience and Adaptability

Multi centre Med SJT

- Resilience and Adaptability
- Collaboration including Leadership and Followership

SJT ranking example

Sarah, a junior doctor, is looking after Mr Kucera who has previously been treated for prostate cancer. Preliminary investigations are strongly suggestive of a recurrence. As she finishes taking blood from a neighbouring patient, Mr Kucera leans across and says "tell me honestly, is my cancer back?"

Please rank order the following actions by **Sarah** in response to this situation
(1= Most appropriate; 5= Least appropriate)

Move each item by clicking on the item. Hold the mouse button and move the item into your chosen rank. The rank number will appear once you hover over the items.

- Explain to Mr Kucera that it is likely that his cancer has come back
- Reassure Mr Kucera that he will be fine
- Explain to Mr Kucera that she does not have all the test results, but will speak to him as soon as she does
- Inform Mr Kucera that she will chase up the results of his tests and ask one of her senior colleagues to discuss them with him **4**
- Invite Mr Kucera to join her and a senior nurse in a quiet room, get a colleague to hold her pager then explore his fears

SJT Rating example

The following is an example of a *'rate' appropriateness* scenario. This requires you to rate the appropriateness of each the response options on a scale of 1=very appropriate to 4=very inappropriate. Each item is rated independently of the other items. As such, each rating can be used more than once or not at all. For example, you may rate all items as a 1 if you wish.

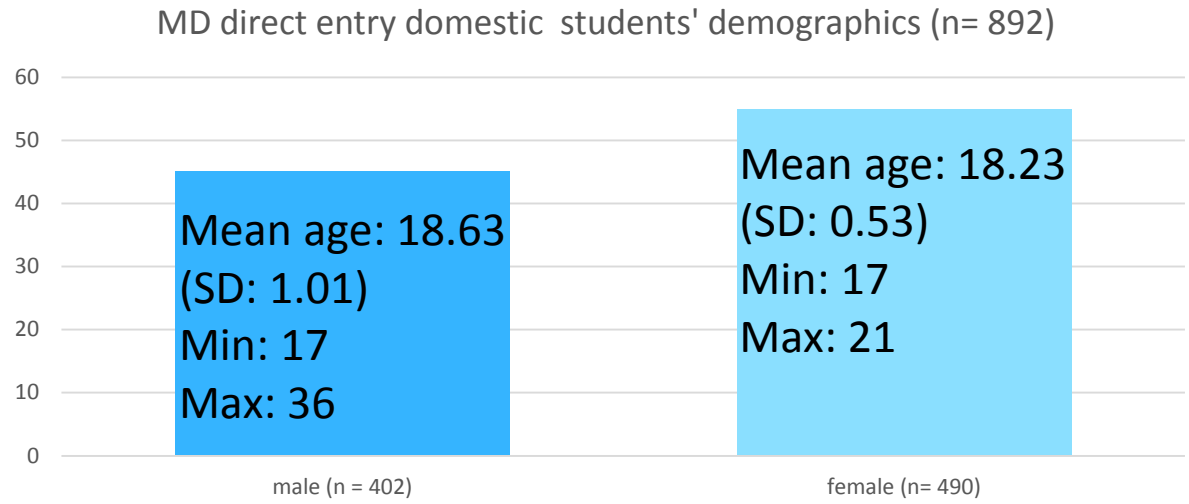
A medical student, Rio, is leading a group of fellow students in a project they have been assigned to work on together. Jerome approaches Rio with a complaint about one of the other students in the group, Elizabeth. Jerome insists that Elizabeth is not dedicating enough time and effort to the group project. Jerome explains that Elizabeth leaves meetings early and her notes are not very comprehensive when she is asked to research topics. He asks if Rio can do something about this as he is leader of the group.

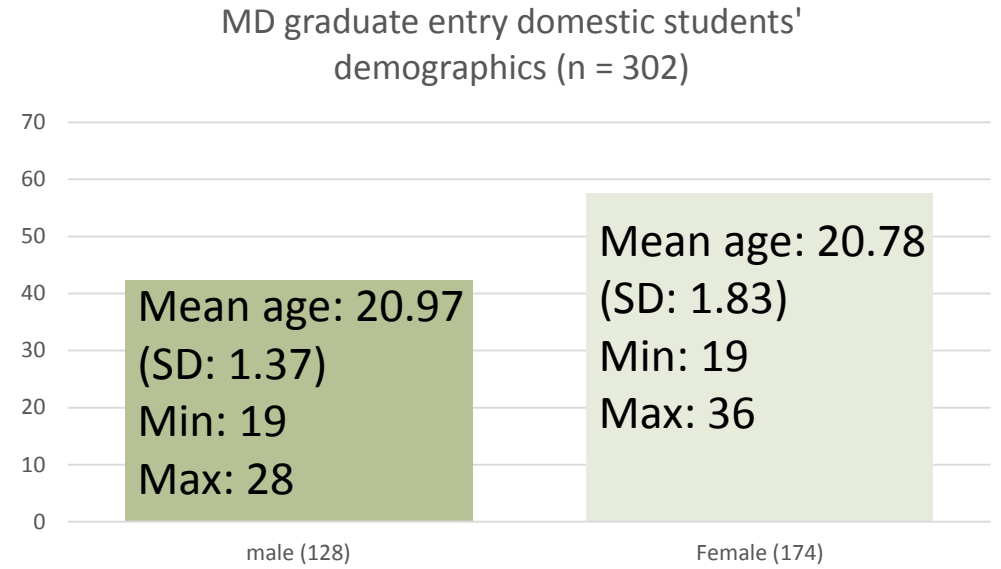
How appropriate are each of the following responses by **Rio** in this situation

Scale 1= very appropriate, 2= appropriate, 3= inappropriate, 4 = very inappropriate

	1	2	3	4
Take no action unless further complaints are received from other members of the group	○	○	○	○
Suggest to Jerome that he raises his concerns with Elizabeth directly	○	○	○	○
Report Elizabeth's lack of dedication to the academic tutor	○	○	○	○
Ask other members of the project group whether they have similar concerns about Elizabeth's behaviour	○	○	○	○
Speak to Elizabeth directly to raise concerns about the time and effort she is dedicating to the group project	○	○	○	○

MD SJT Validation Sample Demographics





MD SJT Validation Sample enrolled / not enrolled

Group	total SJT	enrolled	not enrolled
2015 interviewed applicant	113	40	73
2015 Enrolled Year 1 student	61	61	0
2016 interviewed applicant	335	159	176
2016 enrolled year 1 student	30	29	1
2017 interviewed applicant	352	173	179
Total	891	**462**	**429**

Mean rating SJT score (SD)		Mean ranking SJT score (SD)		Mean total SJT score (SD)	
enrolled	not enrolled	enrolled	not enrolled	enrolled	not enrolled
188.95 (21.57)	190.26 (18.46)	245.50 (10.60)	**243.55*** (11.63)	434.44 (26.60)	433.81 (25.46)

Independent sample t test: Rating score $F(889) = .954$, $p = .330$, Ranking score $F(917) = .678$, ***p* =0.028**, SJT total score $F(889) = 1.154$, $p = .009$.

MD undergrad and graduate entry SJT Validation Sample: Course data Enrolled Students 2015 to 2017

		Assessment data obtained	
Entry Year	N	Year 1	Year 2
2015	99	99	97
2016	190	190	Too early
2017	173	Too early	Too early
Total	462	289	97

		BMS Assessment data obtained (pre MD course)		
Entry Year	N	Year 1	Year 2	BMS final mark
2015 for 2017	129			127
2016 for 2017	56			51
2016 for 2018	124	124	124	Too early
Total	309			178

Direct Entry MD SJT Psychometrics (n=892)

	Cronbach's alpha	Mean	Median	Std. Dev	Min	Max	Max possible
SJT Total (35 Scenarios)	**.87**	435	439	26.242	238	494	541
Rating (17 Scenarios)	**.86**	189.89	192	19.093	48	227	245
Ranking (18 Scenarios)	**.77**	245.11	246	12.17	154	276	296

Direct Entry SJT MD Psychometrics (n=892)

	Mean	SD	Minimum	Maximum
Difficulty[a]	.49	.20	.08	.97
Discriminating Power[b]	.19	.09	.04	.41
Corrected Item-Total Correlations[c]	.20	.10	.01	.49

a: The ideal difficulty for items in this test is .625
b: Items with Discriminating power above .13 are considered reasonable (Cohen & Swerdlik, 2005).
c: An average corrected item-total correlation between .20 and .40 "represents an optimal level of item specificity" (Piedmont & Hyland, 1993).

Construct Validity: Correlations between SJT, UMAT, ATAR and MMI (n=892)

	UMAT 1	UMAT 2	UMAT 3	UMAT Total	ATAR	MMI Total	Rating Total	Ranking Total	SJT Total
UMAT 1	1	.088**	.281**	.729**	.155**	0.015	-0.062	-0.010	-0.053
UMAT 2		1	-0.052	.517**	0.040	.164**	0.054	.137**	.101**
UMAT 3			1	.659**	.143**	-0.045	-.070*	-.097**	-.096**
UMAT Total				1	.178**	0.065	-0.044	0.011	-0.030
ATAR					1	.203**	0.003	0.028	0.014
MMI Total						1	.079*	.195**	.144**
Rating Total							1	.332**	.915**
Ranking Total								1	.684**
SJT Total									1

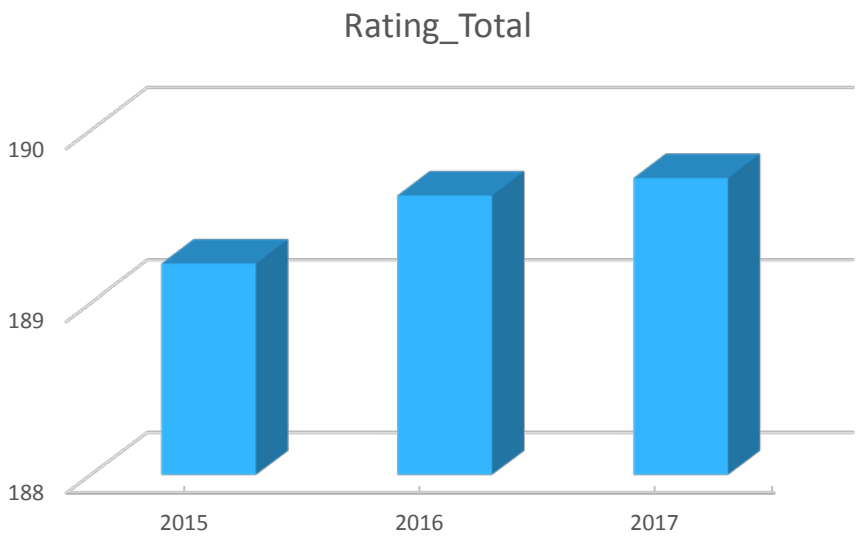
** Correlation is significant at the 0.01 level (2-tailed).

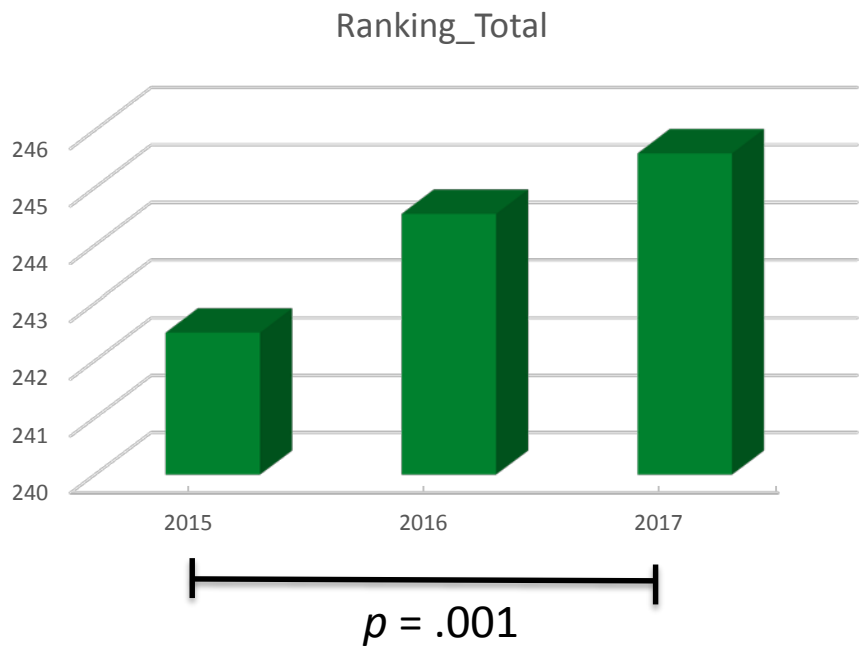
*Correlation is significant at the 0.05 level (2-tailed).

Predictive validity: Correlations between SJT, Y1 and Y2 mid and end of semester and OSCE marks (n=892)

	SJT Rating	SJT Ranking	SJT_Total	Y1 1011 MST raw	Y1 1011 EOS raw	Y1 1022 MST raw	Y1 1022 EOY raw	Y1 Total raw	Y1 1022 OSCE raw	Y2 2031 MST raw	Y2 2031 EOS raw	Y2 2042 MST raw	Y2 2042 EOY raw	Y2 2042 OSCE raw	Y2 Total raw
SJT Rating	1	.332**	.915**	0.000	0.057	0.049	0.057	-0.011	0.037	0.078	-0.089	-0.013	0.107	0.046	0.011
SJT Ranking		1	.684**	0.059	0.084	0.017	0.020	0.073	.132*	0.168	-0.006	0.113	0.138	0.109	0.121
SJT_Total			1	0.027	0.081	0.044	0.052	0.027	0.088	0.142	-0.066	0.049	0.147	0.089	0.070
Y1_1011_EOS_raw					1	.593**	.636**	.749**	.325**	.620**	.669**	.660**	.621**	.475**	.737**
Y1_1022_MST_raw						1	.732**	.844**	.475**	.711**	.621**	.717**	.652**	.414**	.733**
Y1_1022_EOY_raw							1	.848**	.386**	.796**	.782**	.794**	.714**	.443**	.792**
Y1_Total_raw								1	.628**	.800**	.772**	.767**	.763**	.584**	.857**
Y1_1022_OSCE_raw									1	.278**	.373**	.315**	.438**	.577**	.505**
Y2_2031_MST_raw										1	.690**	.771**	.668**	.391**	.752**
Y2_2031_EOS_raw											1	.739**	.760**	.530**	.861**
Y2_2042_MST_raw												1	.750**	.446**	.829**
Y2_2042_EOY_raw													1	.631**	.881**
Y2_2042_OSCE_raw														1	.715**
Y2_Total_raw															1

Direct entry MD students: Changes of SJT scores across 3 years of testing





Nursing SJT Validation Sample: Demographics

	Male (%)	Female (%)	Missing	Total
	39 (12.0)	270 (82.8)	46 (12.9)	355 (100)
Age group				
20-24	22 (56.4)	167 (61.9)		
25-29	9 (23.1)	65 (24.1)		
30-34	4 (10.3)	18 (6.7)		
35-39	0	12 (4.4)		
40-44	2 (5.1)	4 (1.5)		
45-49	0	2 (0.7)		
50-54	2 (5.1)	0		

326 were graduate nurses, 29 applicants for the Master of Nursing Practice. No sig differences in SJT scores were found between the two groups.

74% spoke English as first language, other first languages included:
Amharic, Arabic, Armenian, Bahasa indinesia, Chinese, Croatian, Dinka, English, Filipino, French, Gujarati, Hebrew, Hindi, Igbo (Nigerian), Indonesian, Khmer, Korean, Lorma, Mandarin, Nepalese, Nepali, Punjabi, Russian, Shona, Singhalese, Tagalog (Filipino), Tamil, Urdu, Vietnamese.

Nursing SJT psychometrics (n=355)

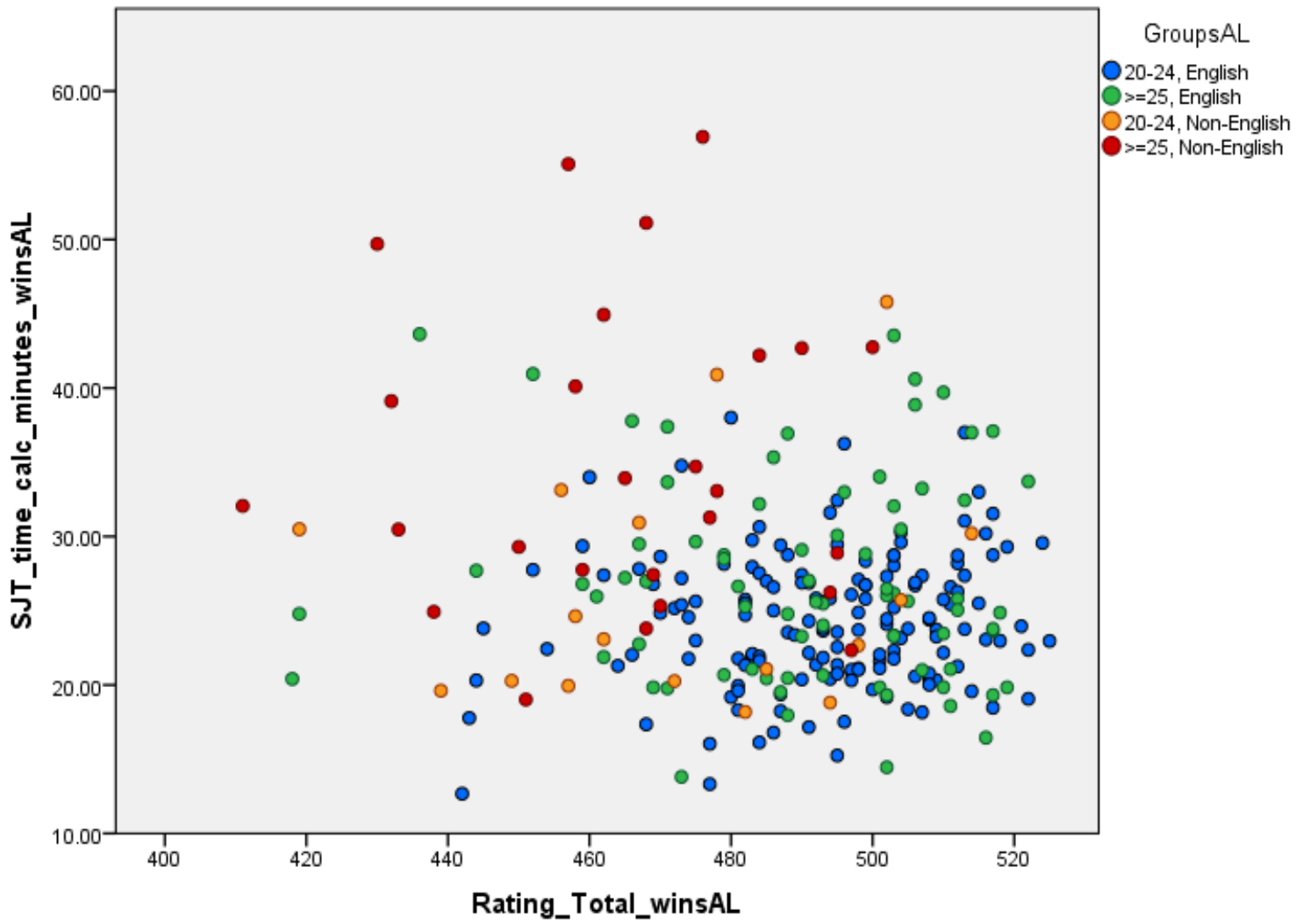
	Mean	SD	Minimum	Maximum
Difficulty [a]	.65	.20	.10	.95
Discriminating Power [b]	.28	.20	.02	.59
Corrected Item Total Correlation [c]	.25	.12	.02	.52

a: The ideal difficulty for items in this test is .625

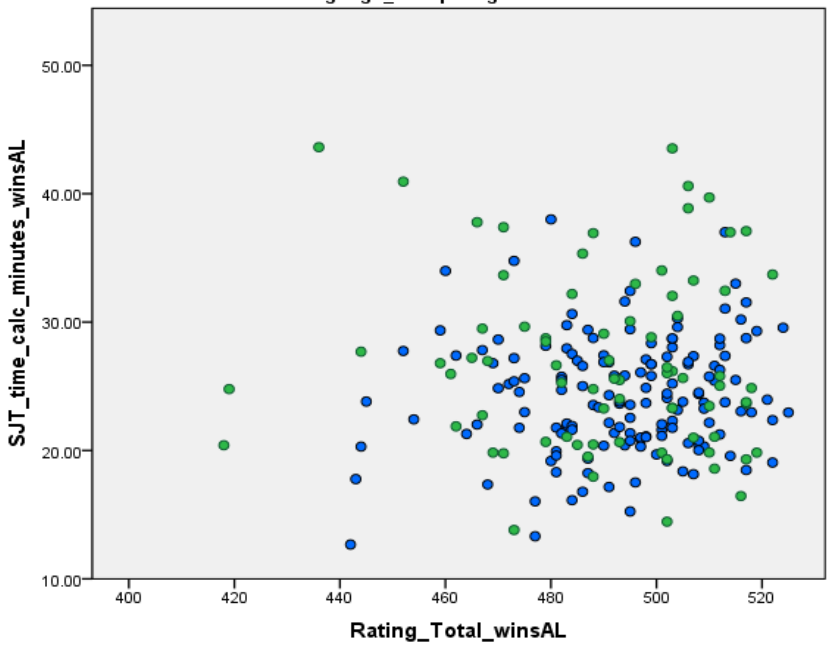
b: Items with Discriminating power above .13 are considered reasonable (Cohen & Swerdlik, 2005).

c: An average corrected item-total correlation between .2 and .4 "represents an optimal level of item specificity" (Piedmont & Hyland, 1993).

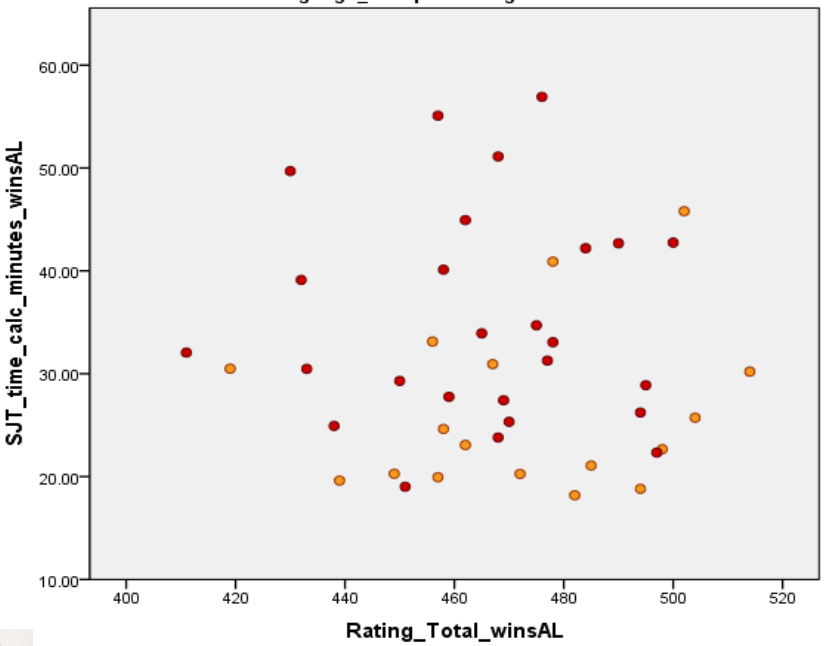
Nursing Validation Sample: ESL and age effects on SJT scores, SJT Total



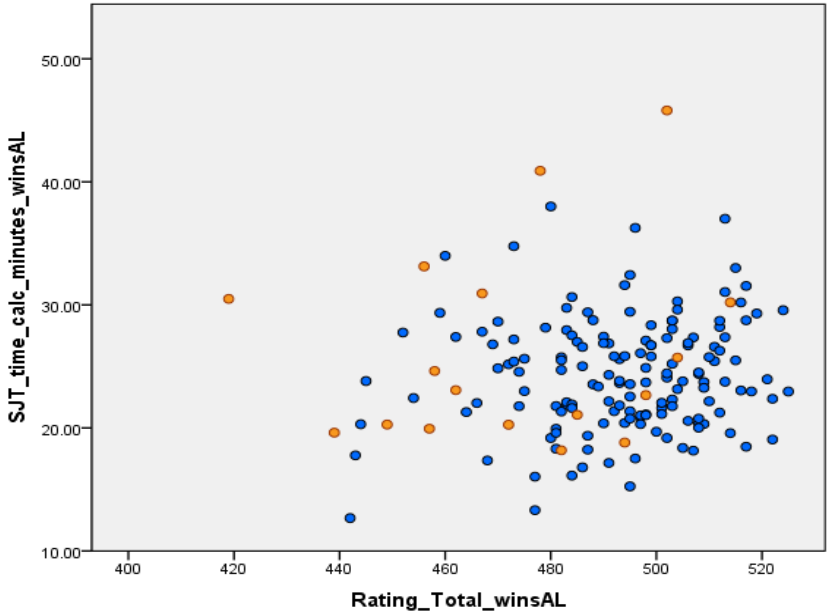
Language_Group: English



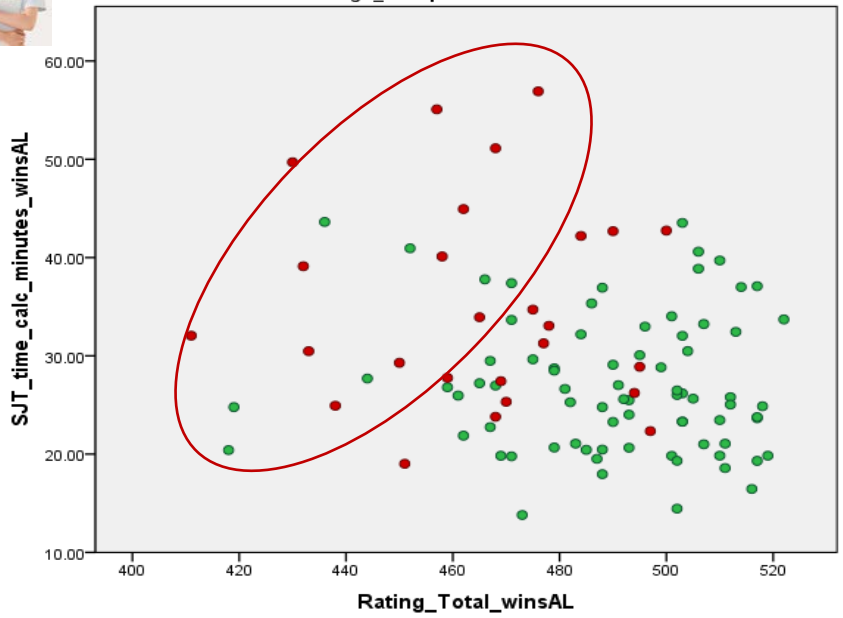
Language_Group: Non-English



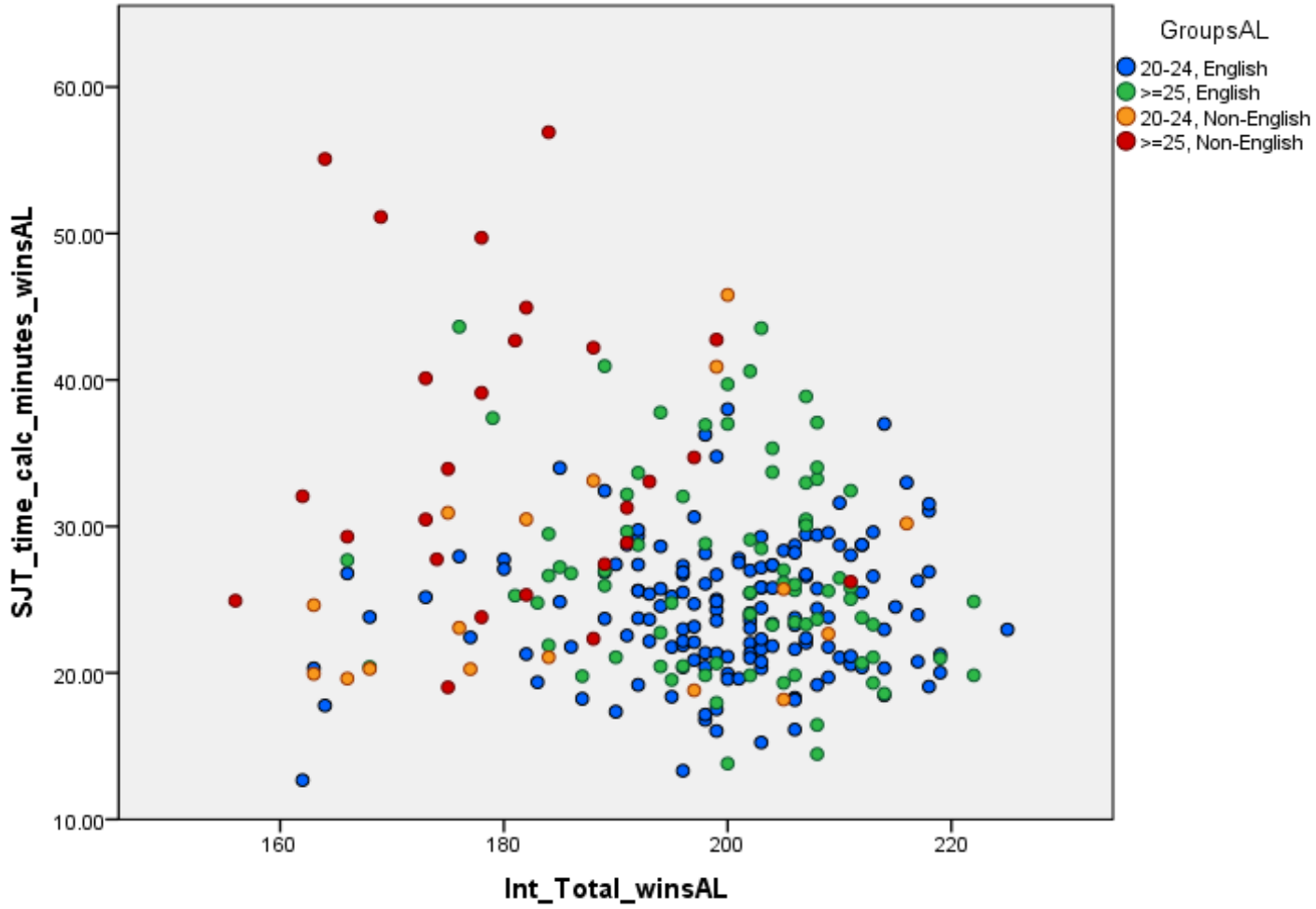
Age_Group: 20-24

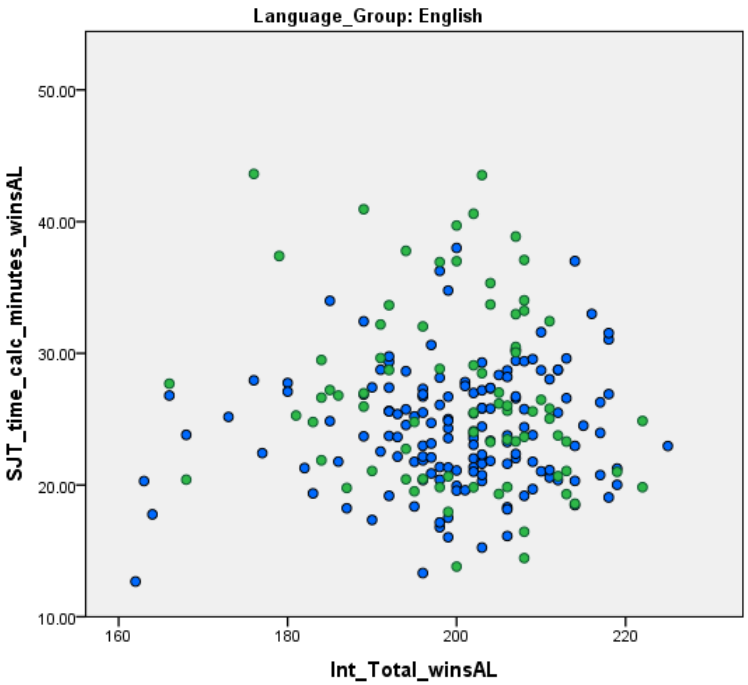


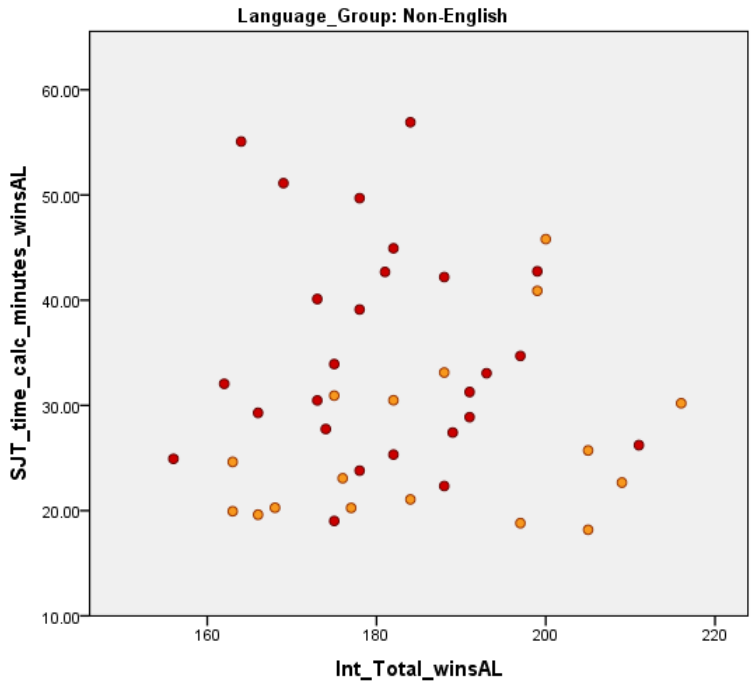
Age_Group: >=25

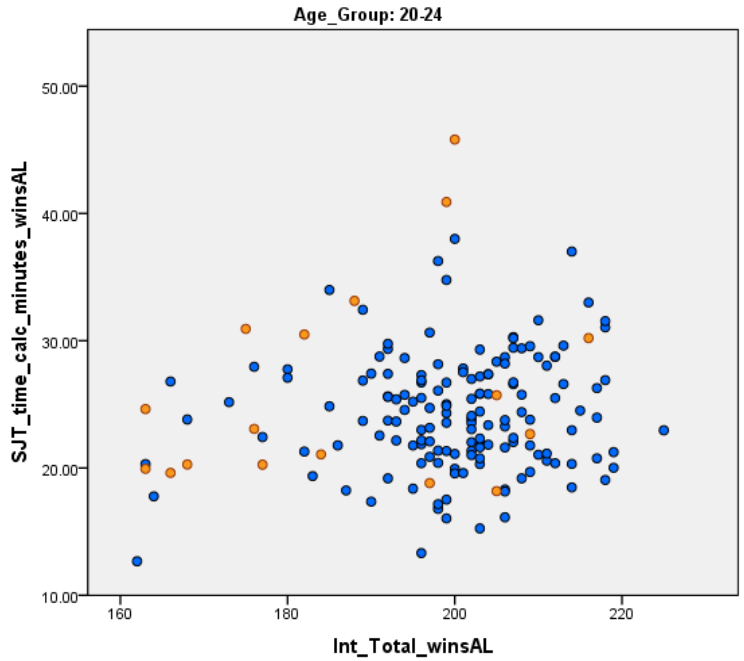


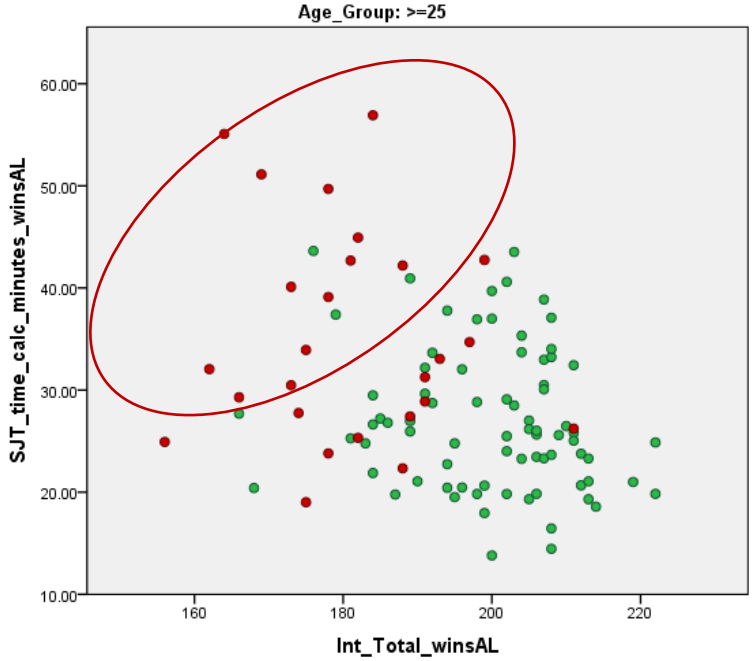
ESL and age effects on SJT scores, SJT Total, Integrity



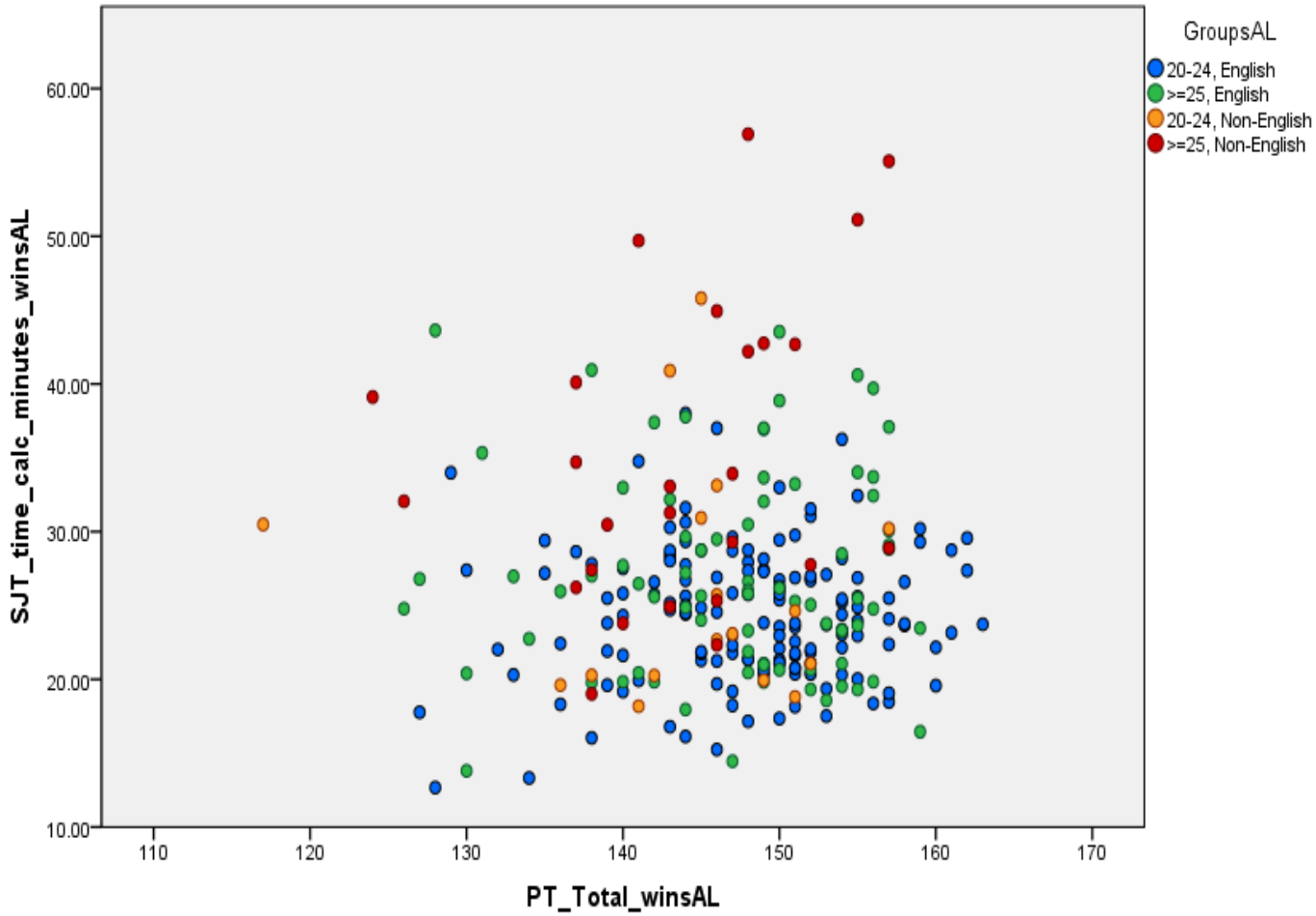

Language_Group: English
SJT_time_calc_minutes_winsAL
Int_Total_winsAL
10.00
20.00
30.00
40.00
50.00
160
180
200
220

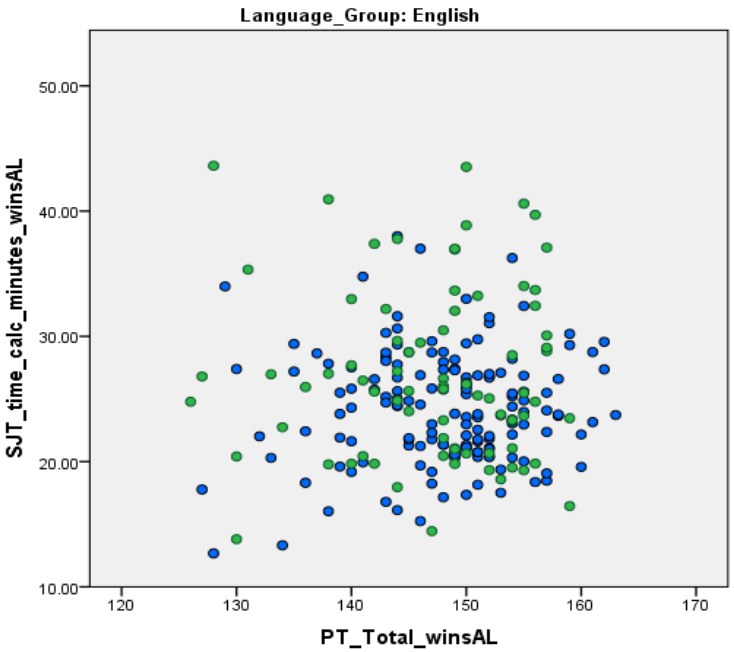

Language_Group: Non-English
SJT_time_calc_minutes_winsAL
Int_Total_winsAL
10.00
20.00
30.00
40.00
50.00
60.00
160
180
200
220

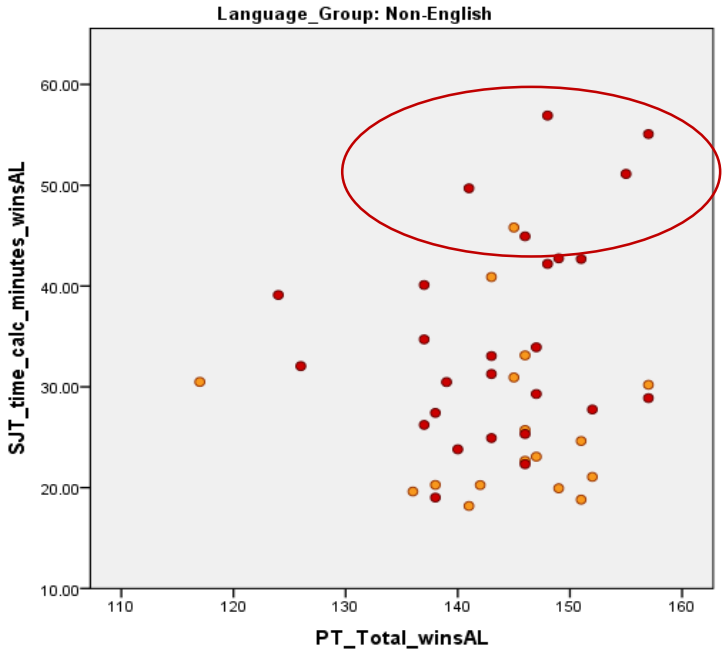

Age_Group: 20-24
SJT_time_calc_minutes_winsAL
Int_Total_winsAL
10.00
20.00
30.00
40.00
50.00
160
180
200
220

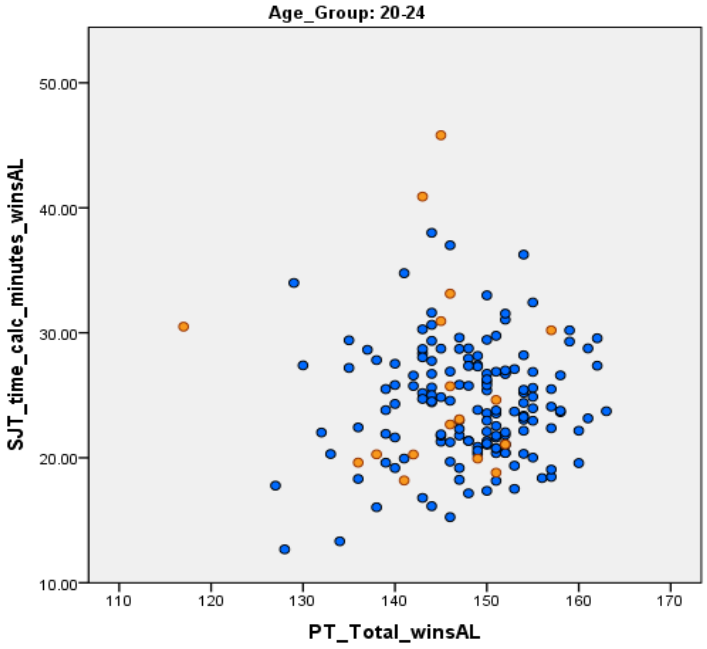

Age_Group: >=25
SJT_time_calc_minutes_winsAL
Int_Total_winsAL
10.00
20.00
30.00
40.00
50.00
60.00
160
180
200
220

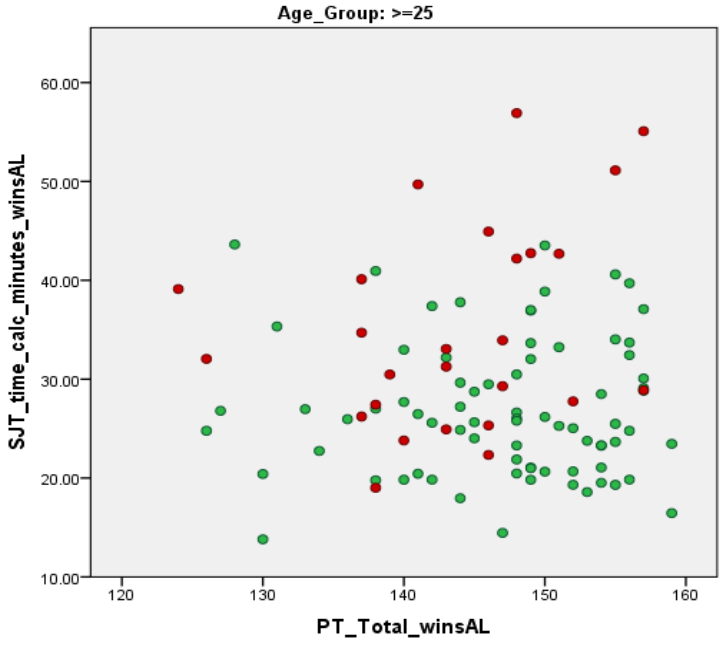
ESL and age effects on SJT scores, SJT Total, Empathy




Language_Group: English
SJT_time_calc_minutes_winsAL
PT_Total_winsAL
50.00
40.00
30.00
20.00
10.00
120
130
140
150
160
170


Language_Group: Non-English
SJT_time_calc_minutes_winsAL
PT_Total_winsAL
60.00
50.00
40.00
30.00
20.00
10.00
110
120
130
140
150
160


Age_Group: 20-24
SJT_time_calc_minutes_winsAL
PT_Total_winsAL
50.00
40.00
30.00
20.00
10.00
110
120
130
140
150
160
170


Age_Group: >=25
SJT_time_calc_minutes_winsAL
PT_Total_winsAL
60.00
50.00
40.00
30.00
20.00
10.00
120
130
140
150
160

Nutrition and Dietetics SJT psychometrics (n=107)

	Mean	SD	Minimum	Maximum
Difficulty [a]	.60	.22	.08	.95
Discriminating Power [b]	.26	.12	-.03	.53
Corrected item total correlations [c]	.34	.20	.04	.79

a: The ideal difficulty for items in this test is .625
b: Items with Discriminating power above .13 are considered reasonable (Cohen & Swerdlik, 2005).
c: An average corrected item-total correlation between .2 and .4 "represents an optimal level of item specificity" (Piedmont & Hyland, 1993).

MONASH University

Investigating Scoring

Sam Henry

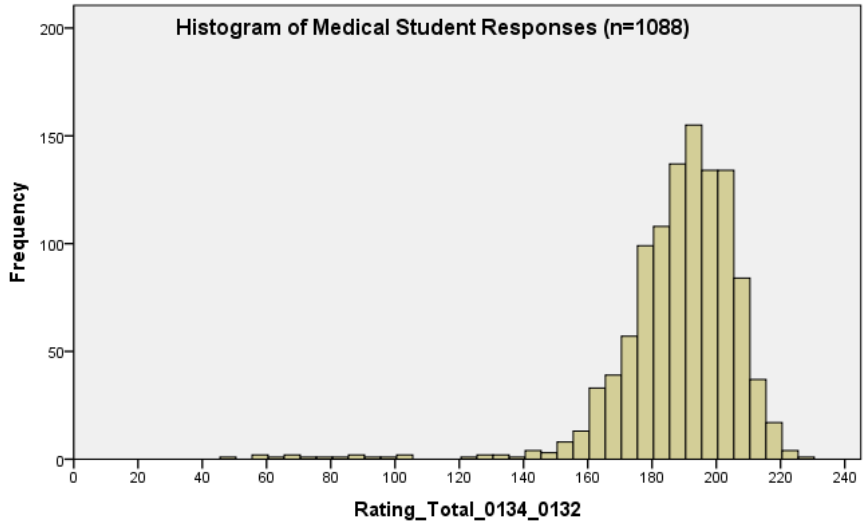
Investigating Scoring

MONASH University

- Our Current Scoring System
- Scoring Problems
  - Item Weightings
  - Perceptions of Distance
- Possible Solutions
  - 2-point Options
  - 3-point Options
  - 4-point Options
- Ideas About Cutoffs

Our Current Scoring System - Rating

	Very Inappropriate (VI)	Inappropriate (I)	Appropriate (A)	Very Appropriate (VA)
VA is Correct	0	1	3	4
A is Correct	0	1	3	2
I is Correct	2	3	1	0
VI is Correct	4	3	1	0

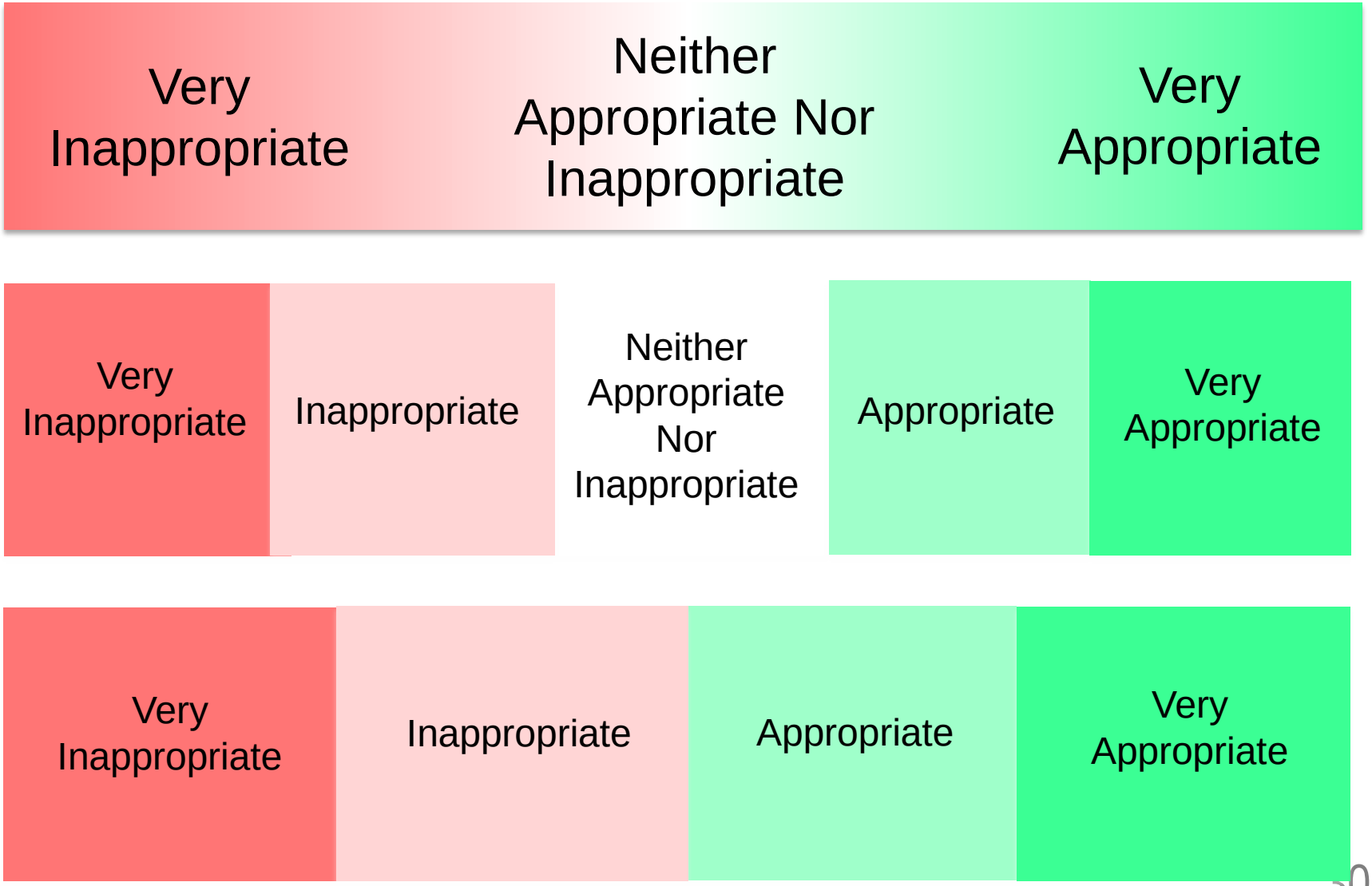


- 2 points between Inappropriate and Appropriate (to represent “Neither Appropriate nor Inappropriate”).
- Some items are out of 3, others are out of 4.

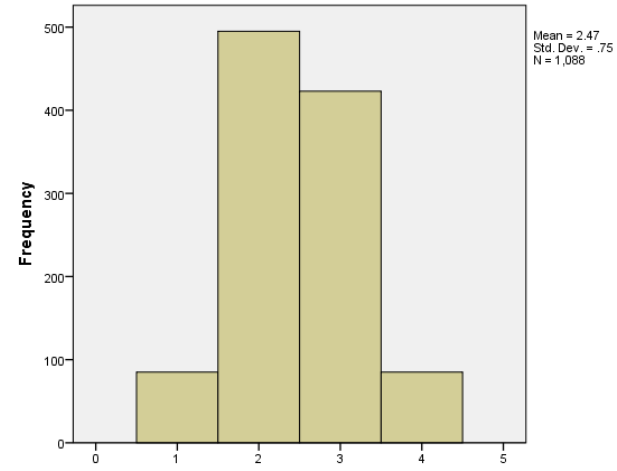
Calculating the total score – Item Weighting Questions

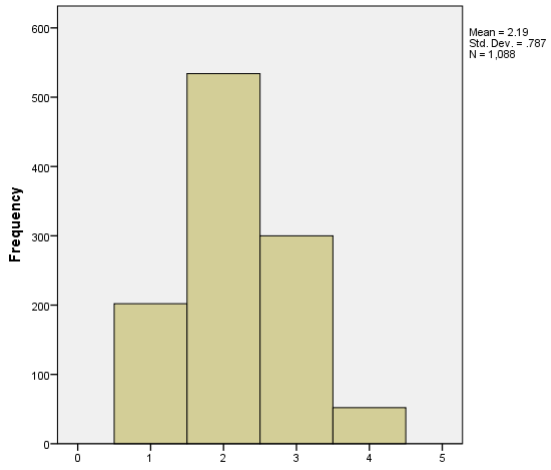
	Very Inappropriate (VI)	Inappropriate (I)	Appropriate (A)	Very Appropriate (VA)	Number of items	Number of items multiplied by maximum
VA is Correct	0	1	3	4	14	56
A is Correct	0	1	3	2	2	6
I is Correct	2	3	1	0	13	39
VI is Correct	4	3	1	0	36	144
					Total	**245**

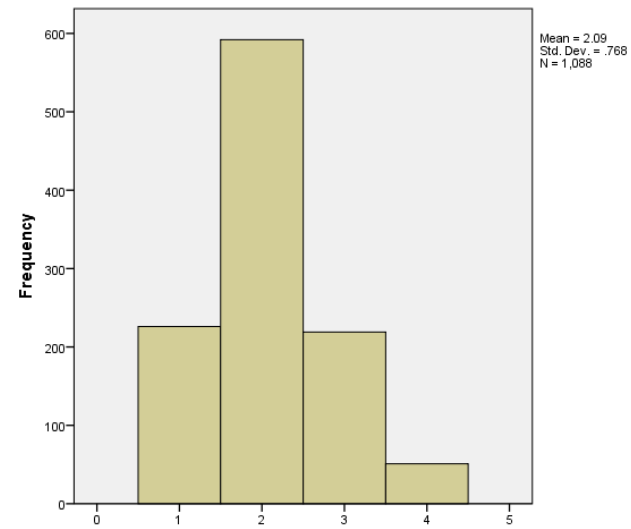
Student's Perceptions of Distance Between Responses



Student's Perceptions of Distance Between Responses







For 11/15 questions where the correct answer is a 2 or a 3, more students answer moderately on the incorrect side, than extreme on the correct side.

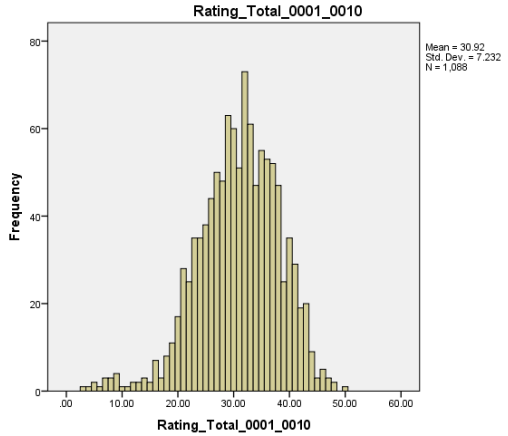
MONASH University

Scoring Options In Practice:
2-Point Options

Binary – Correct Choice

Very Inappropriate	Inappropriate	Appropriate	Very Appropriate
0	0	0	1
0	0	1	0
0	1	0	0
1	0	0	0

- Score is equal to the number of correct answers.



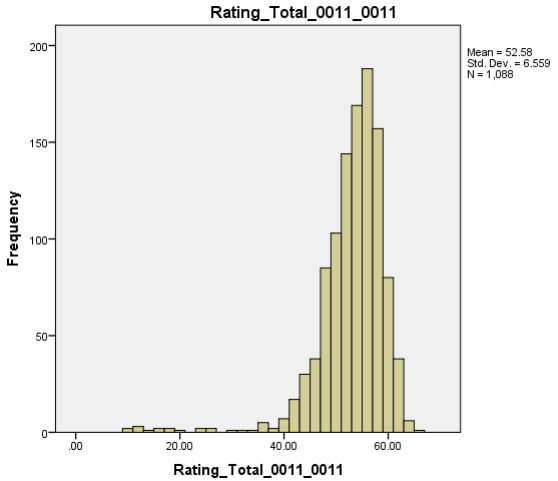
Pearson Correlation Coefficients

	0001 0010	Current System
UMAT Book 2	.065	.054
Interview Score	.095**	.079*
Y1 OSCE	.072	.102*
Y2 OSCE	-.063	.018

* p <.05
** p < .01

Binary – Correct Side

Very Inappropriate	Inappropriate	Appropriate	Very Appropriate
0	0	1	1
0	0	1	1
1	1	0	0
1	1	0	0



- Score is equal to the number of correct appropriate choices.

	0011 0011	Current System
UMAT Book 2	.048	.054
Interview Score	.044	.079*
Y1 OSCE	.092*	.102*
Y2 OSCE	.015	.018

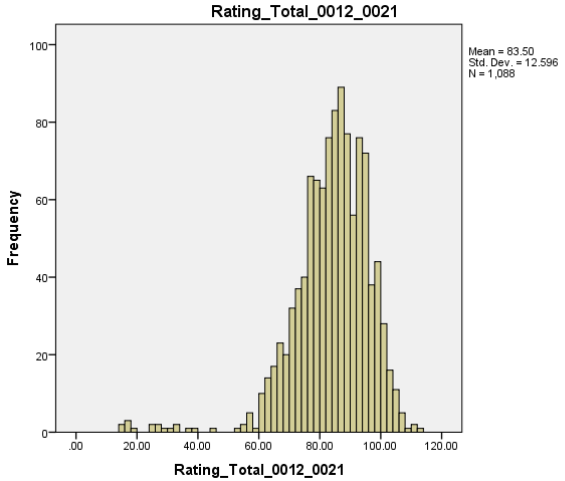
MONASH University

3-Point Options

Correct Side, with bonus for correct choice

Very Inappropriate	Inappropriate	Neither Appropriate Nor Inappropriate	Appropriate	Very Appropriate

Very Inappropriate	Inappropriate	Appropriate	Very Appropriate
0	0	1	2
0	0	2	1
1	2	0	0
2	1	0	0

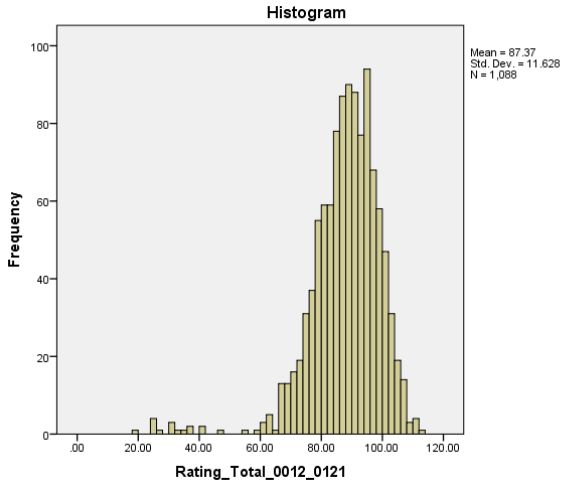


- Only get marks for the correct side, with a bonus mark for the correct choice.

	0012 0021	Current System
UMAT Book 2	.062	.054
Interview Score	.078*	.079*
Y1 OSCE	.090*	.102*
Y2 OSCE	-.033	.018

Proximity Version

Very Inappropriate	Inappropriate	Appropriate	Very Appropriate
0	0	1	2
0	1	2	1
1	2	1	0
2	1	0	0



- Receive two marks for correct answer, and one mark if close.

	0012 0121	Current System
UMAT Book 2	.058	.054
Interview Score	.094*	.079*
Y1 OSCE	.096*	.102*
Y2 OSCE	.002	.018

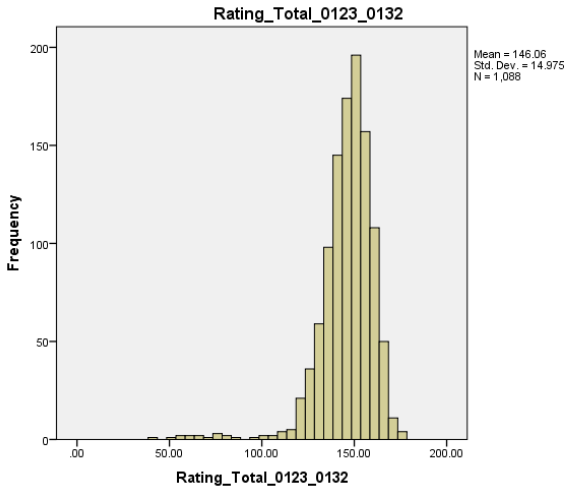
MONASH University

4-Point Options

0123 - 0132

Very Inappropriate	Inappropriate	Neither Appropriate Nor Inappropriate	Appropriate	Very Appropriate

Very Inappropriate	Inappropriate	Appropriate	Very Appropriate
0	1	2	3
0	1	3	2
2	3	1	0
3	2	1	0



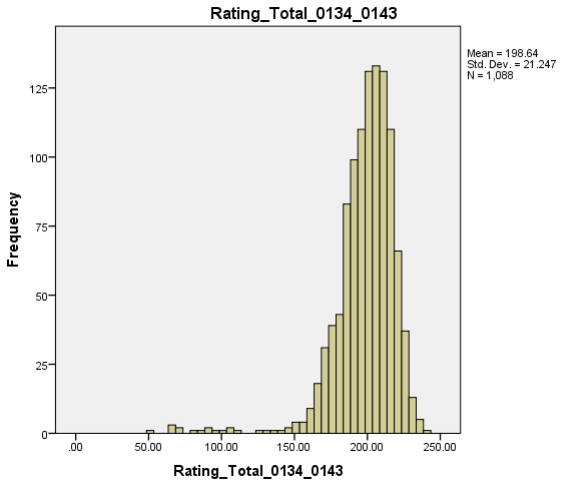
- Score is equal to the number of correct answers multiplied by 3.

	0123 0132	Current System
UMAT Book 2	.058	.054
Interview Score	.077*	.079*
Y1 OSCE	.100*	.102*
Y2 OSCE	-.014	.018

0134 - 0143

Very Inappropriate	Inappropriate	Neither Appropriate Nor Inappropriate	Appropriate	Very Appropriate

Very Inappropriate	Inappropriate	Appropriate	Very Appropriate
0	1	3	4
0	1	4	3
3	4	1	0
4	3	1	0



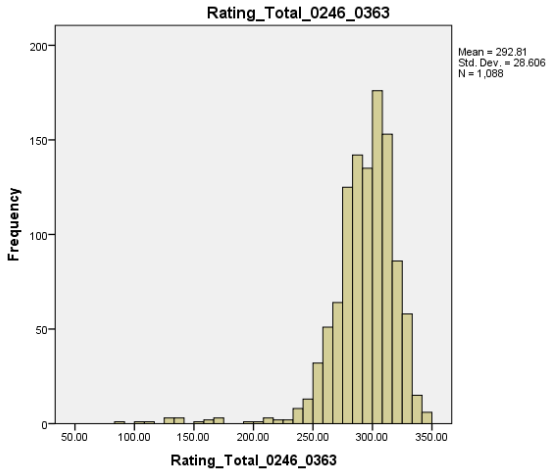
- Score is equal to the number of correct answers multiplied by 4.
- There is a gap in the middle of the scores to indicate “neither appropriate nor inappropriate”.

	0134 0143	Current System
UMAT Book 2	.056	.054
Interview Score	.068*	.079*
Y1 OSCE	.099*	.102*
Y2 OSCE	-.005	.018

0246 - 0363

Very Inappropriate	Inappropriate	Appropriate	Very Appropriate
Very Inappropriate	Inappropriate	Appropriate	Very Appropriate

Very Inappropriate	Inappropriate	Appropriate	Very Appropriate
0	2	4	6
0	3	6	3
3	6	3	0
6	4	2	0



- Score is based on proximity to correct response.
- Maximum score is six, minimum is zero, and all others are equally divided.

	0134 0143	Current System
UMAT Book 2	.055	.054
Interview Score	.089**	.079*
Y1 OSCE	.113*	.102*
Y2 OSCE	.019	.018

MONASH University

Ideas About Cutoffs

Idea for our current system

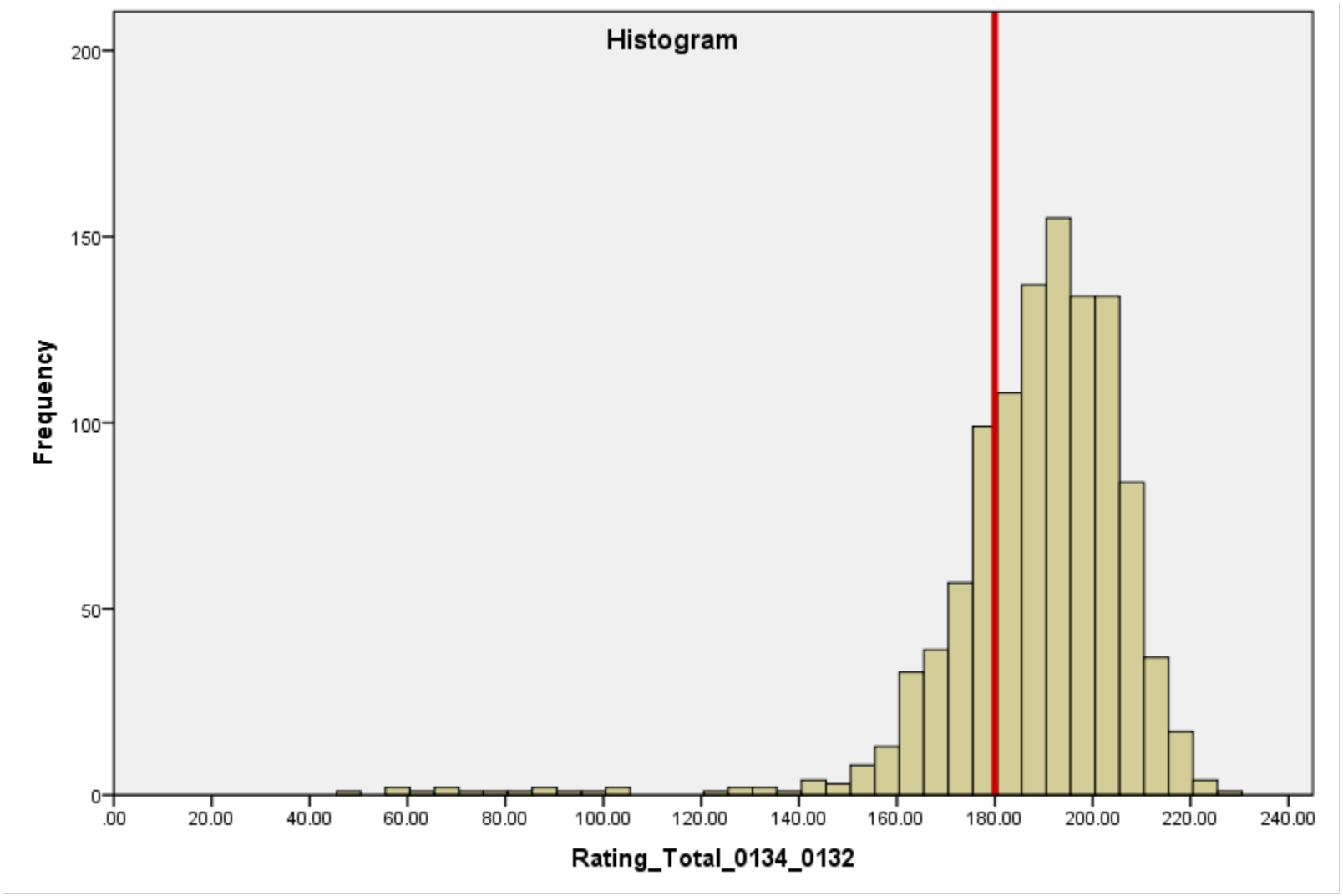
- What if we calculated the minimum score required to choose the correct side?

- For the current system this calculation would be:

$$3 \times max4 + 2 \times max3 = 180$$

	Very Inappropriate (VI)	Inappropriate (I)	Appropriate (A)	Very Appropriate (VA)
VA is Correct	0	1	3	4
A is Correct	0	1	3	2
I is Correct	2	3	1	0
VI is Correct	4	3	1	0

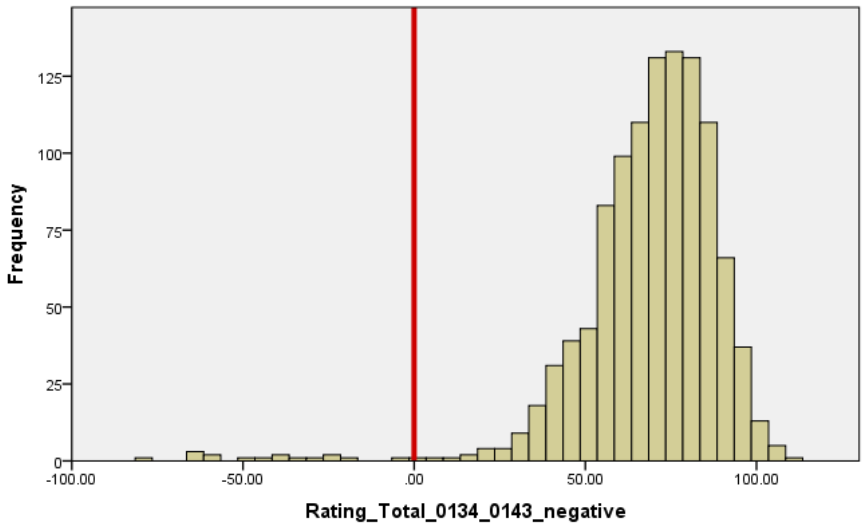
Cut-off for our current system

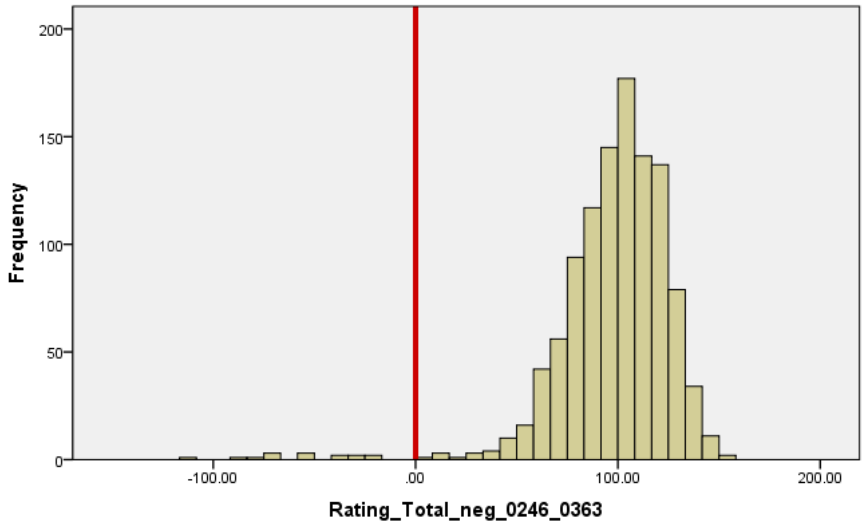


A zero form cut-off?

Very Inappropriate	Inappropriate	Appropriate	Very Appropriate
-2	-1	1	2
-2	-1	2	1
1	2	-1	-2
2	1	-1	-2

Very Inappropriate	Inappropriate	Appropriate	Very Appropriate
-3	-1	1	3
-3	0	3	0
0	3	0	-3
3	1	-1	-3





MONASH University

Monash
Institute for Health
and Clinical Education

Monash
Institute For Health
And Clinical Education

2nd International Selection in the Health Professions Conference

11-13 April 2018