THE EFFECT OF NON-NORMALITY ON THE CUTSCORE OPERATING FUNCTION: ESTIMATION CORRECTNESS IN NON-NORMAL MONTE CARLO SIMULATIONS

Jesse R. Pace
09/06/2019

TODAY'S AGENDA

Literature

- \- Focused on the problem statement and why it matters

Methods

Results

Discussion

- Focused on answers to research questions AND
- Focused on what all this work means for the real world

*Topic font indicates relative time allotment

LITERATURE

Classification

- Any time we give a test to put examinees in categories, we are classifying them

Any time we make a classification, there is a chance that we make an error

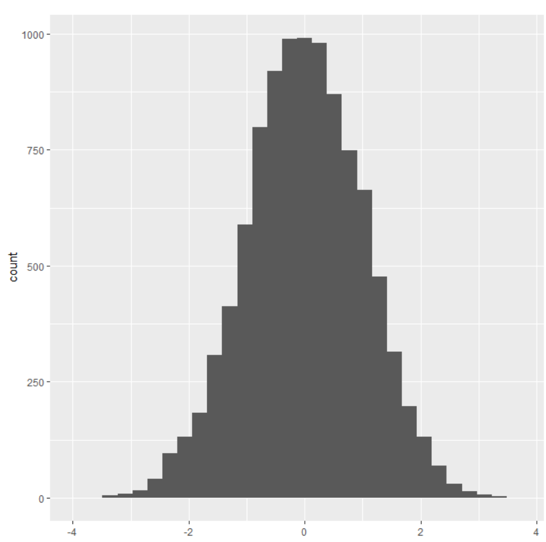
- Errors come in two forms, False Positive (FP) and False Negative (FN)

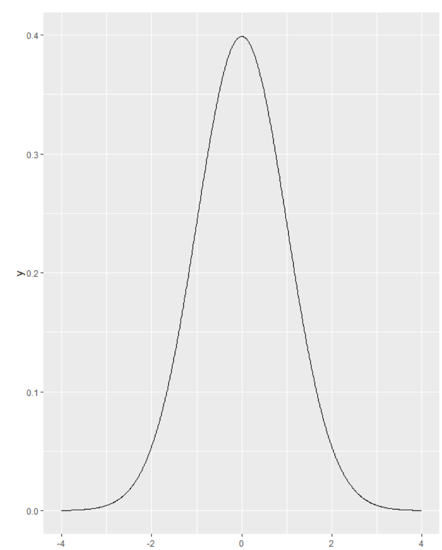
CUTSCORES

Divide a continuous latent trait into categories, such as students who will be labeled ‘competent’ and ‘non-competent’

The location of the cutscore is often informed by standard setting

Standard setting panels stand to benefit from having additional information to make cutscore decisions





GRABOVSKY AND WAINER'S METHOD

Grabovsky and Wainer, 2017 (we'll call this GW-CSOF from hereon)

Used to predict error

Intent was to provide standard setting committees additional info

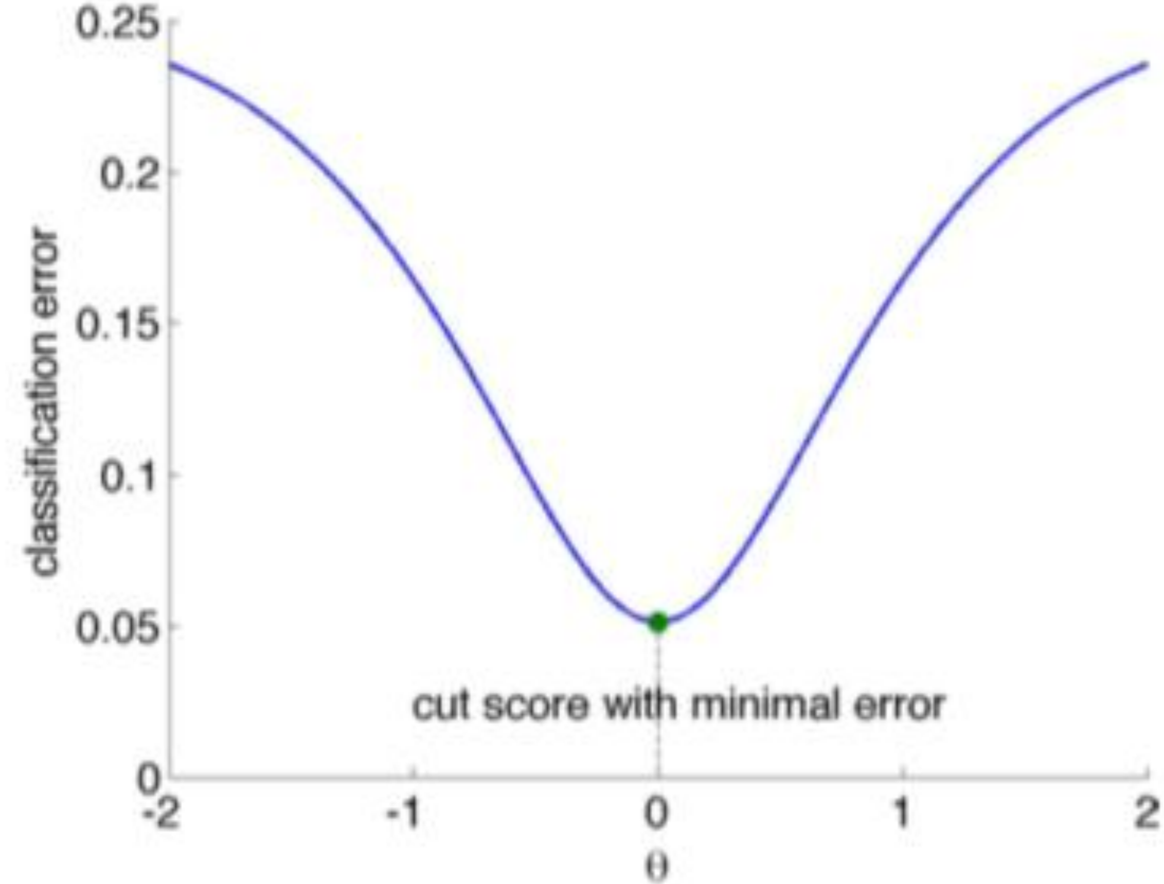
Provides estimated classification error at all possible cutscores

Optimal cutscore: the point where false positive and false negative errors are simultaneously minimized

- I.e., the cutscore which, if chosen, would lead to the smallest number of wrongly classified examinees

** Requires the notion of a 'true cutscore' and an 'observed cutscore'

- X=T+E and $\rho = \sigma_t^2 / \sigma_x^2$ where σ_x^2 is the variation among examinees' observed scores
  - So, true cutscores are on the T continuum, and observed cutscores are in the observed (x) continuum among examinees



From Grabovsky & Wainer, 2017

GW-CSOF COULD PROVIDE USEFUL INFORMATION, BUT... WHEN DOES IT NOT?

Normal examinee ability distribution is assumed

- This is potentially **problematic**

Research has shown that this is problematic assumption

- Micceri (1989) and skewness, bimodality, kurtosis

Problem statement: Research was necessary to determine the degree to which GW-CSOF estimates match actual values, both when assumptions are met, and especially, when this assumption is violated

RESEARCH QUESTIONS

1) Do GW-CSOF estimates of error at the true cutscore location match actual error rates, and does the match change as non-normality increases?

- Hyp.I) The GW-CSOF method would produce error estimates close to actual error values when the normality assumptions of the true score distribution were met,
- Hyp.II) Increased non-normality in the true score distribution would increase incorrectness in error estimates.

RESEARCH QUESTIONS (CONTINUED)

2) Do GW-CSOF estimates of optimal cutscores match the actual location of the optimal cutscore, and does the match change as non-normality increases?

- Hyp. III) the GW-CSOF method would estimate a location for the optimal cutscore near the location of the actual optimal cutscore when normality assumptions of the true score distribution were met,
- Hyp.IV) Increased non-normality in the true score distribution would cause increased incorrectness in GW-CSOF estimates of the optimal cutscore.

METHODS

Simulations

-Why?

- Specific non-normality manipulations

Skewness, bimodality, and kurtosis

50 increasingly non-normal manipulations for each condition

4 true cutscores were looked at:

45, 47.5, 52.5, and *55*

Corresponding to -1, -.5, +.5, and +1 standard deviations form the mean

SIMULATIONS

Skewness

- 1) Exponentially modified normal
  - Manipulating skew between 0 and 2

Bimodality

- 2) Mixture of two normals with different means
  - Mainpulation was conducted between 0 and 10

Kurtosis

- 3) Mixture of two normals with different variances
  - Manipulating from 3 to 6

SPEARMAN'S RHO

Used to correlate the degree of non-normality (e.g., the skewness, difference of modes, or kurtosis) with the difference between GW-CSOF and actual estimates

Also correlated the difference in location of the optimal cutscore with the degree of non-normality

RESULTS

Simulation quality: simulations performed as hoped

Figure E 1 : Iteration # 1 Distribution Skewness of 0)
True Score Frequencies with Skewness = -0.03

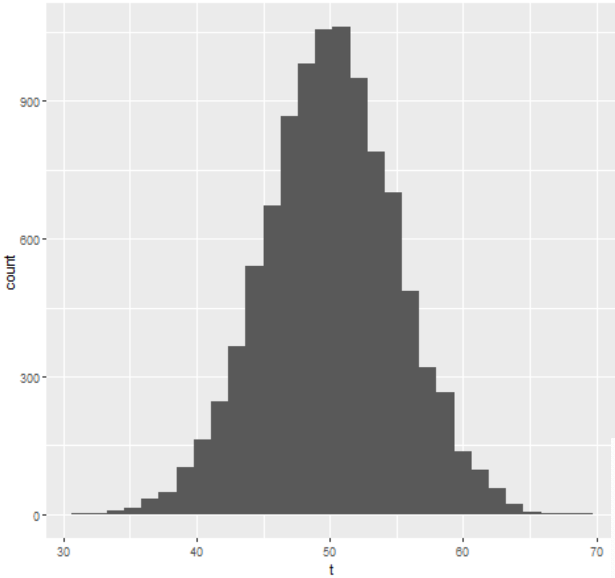


Figure F 1 : Iteration # 26 (distance of in mixture means= 5)
True Score Frequencies with D undefined

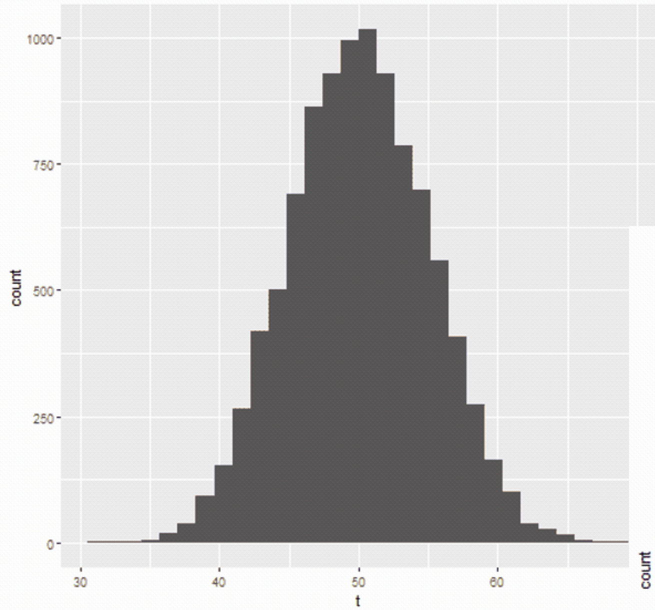
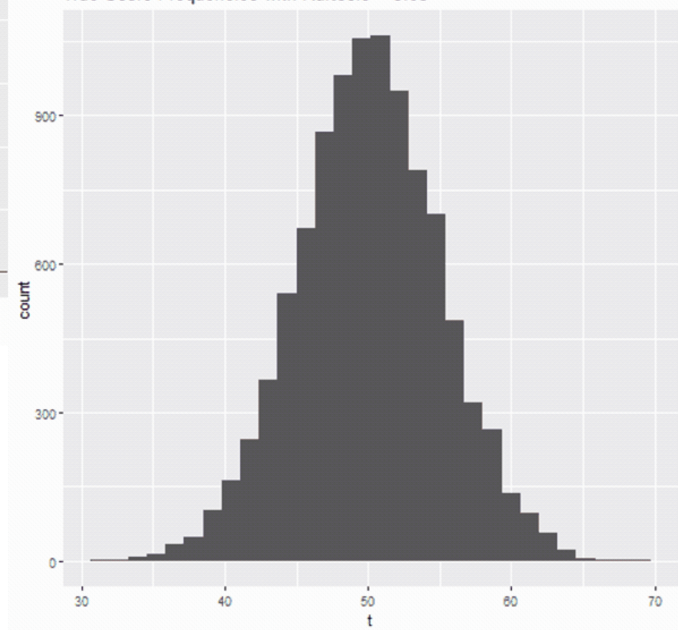


Figure G 1 : Iteration # 1 (Distribution Kurtosis of 3)
True Score Frequencies with Kurtosis = 3.03



RESEARCH QUESTIONS

Do GW-CSOF estimates of error at the true cutscore location match actual error rates, and does the match change as non-normality increases?

- Hyp.I) The GW-CSOF method would produce error estimates close to actual error values when the normality assumptions of the true score distribution were met,
- Hyp.II) Increased non-normality in the true score distribution would increase incorrectness in error estimates. **Partially Supported**

Do GW-CSOF estimates of optimal cutscores match the actual location of the optimal cutscore, and does the match change as non-normality increases?

- Hyp. III) the GW-CSOF method would estimate a location for the optimal cutscore near the location of the actual optimal cutscore when normality assumptions of the true score distribution were met, **Supported**
- Hyp.IV) Increased non-normality in the true score distribution would cause increased incorrectness in GW-CSOF estimates of the optimal cutscore. **Partially Supported**

HYPOTHESIS 1: EVIDENCE TO SUPPORT

	Actual Error	GW-CSOF Estimated Error
45	0.088	0.095
47.5	0.134	0.131
52.5	0.131	0.134
55	0.098	0.099

Thus, fewer than 1 in every 100 examinees would be classified differently between the GW-CSOF estimates and the actual values.

HYPOTHESIS 2: EVIDENCE TO SUPPORT

Skewness		
True Cut Location	Error Rate at True Cutscore	
	Spearman's Rho	p
45	0.97	<.001
47.5	0.98	<.001
52.5	0.93	<.001
55	0.97	<.001

Bimodality		
True Cut Location	Error Rate at True Cutscore	
	Spearman's Rho	p
45	0.88	<.001
47.5	0.83	<.001
52.5	0.87	<.001
55	0.9	<.001

Kurtosis		
True Cut Location	Error Rate at True Cutscore	
	Spearman's Rho	p
45	0.96	<.001
47.5	0.49	<.001
52.5	0.55	<.001
55	0.96	<.001

HYPOTHESIS 3: EVIDENCE TO SUPPORT

	Actual Optimal Cutscore Location	GW-CSOF Estimated Optimal Cutscore Locataion
45	43.7	43.6
47.5	46.8	46.8
52.5	53	53.1
55	56.8	56.2

HYPOTHESIS 4: EVIDENCE TO SUPPORT

There's a lot to unpack here, so we will proceed with one condition at a time

Skewness Results

True Cut Location	Optimum Cutscore Location	
	Spearman's Rho	p
45	0.87	<.001
47.5	0.58	<.001
52.5	0.58	<.001
55	-0.25	0.078

As we an see, as the skew increases to the right, there is less and less density to the left of the truecutscore(45). Thus, positioning the observed cutscore to the left allows for: Greatly decreased FN for a small increase in FP

At an x near 39, the number of examinees with t>45 and x<39 is tiny

The number of examinees with t<45 and x>39 is larger than it would have been for, say x>45, but not so large as to overpower the FN decrease

45 Condition

Figure E 1 : Iteration # 1 Distribution Skewness of 0)
True Score Frequencies with Skewness = -0.03

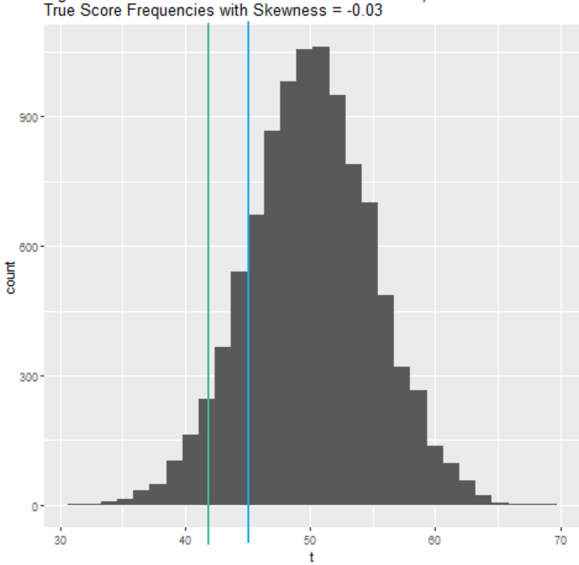


Figure E 7 : Iteration # 30 Distribution Skewness of 1.16)
True Score Frequencies with Skewness = 1.26

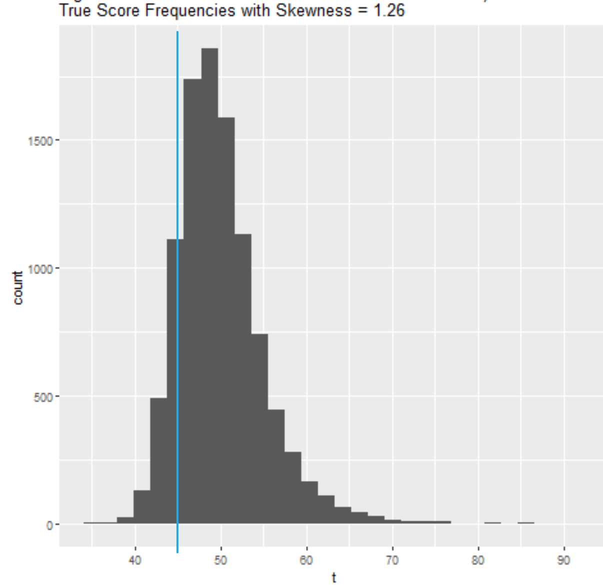
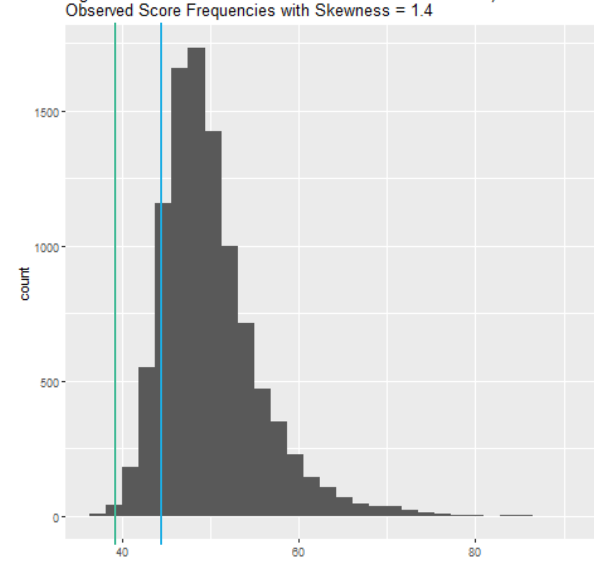


Figure E 12 : Iteration # 50 Distribution Skewness of 1.96)
Observed Score Frequencies with Skewness = 1.4



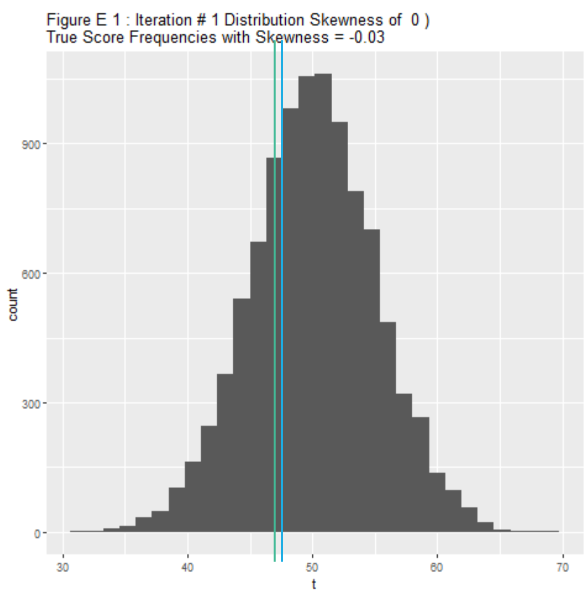
HYPOTHESIS 4: EVIDENCE TO SUPPORT

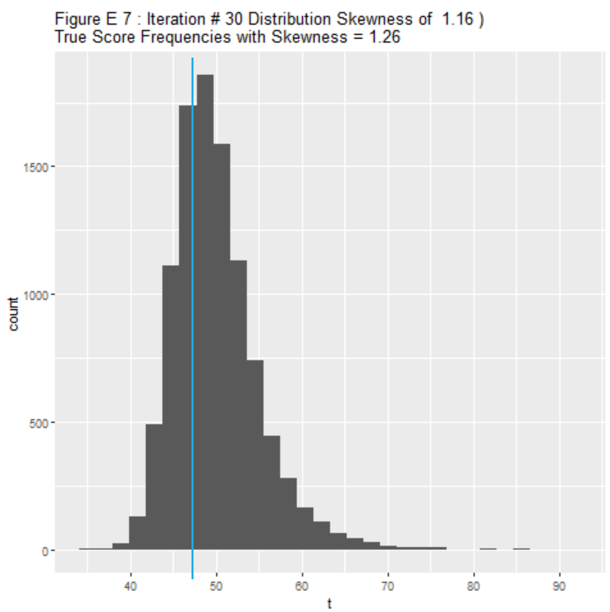
Skewness Results

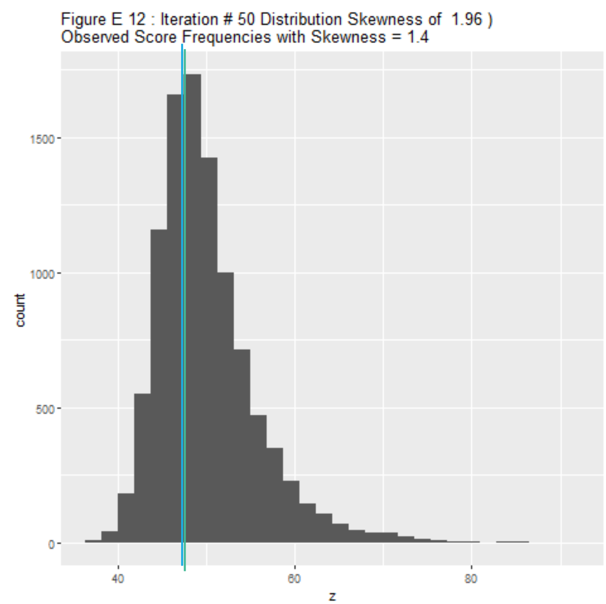
True Cut Location	Optimum Cutscore Location	
	Spearman's Rho	p
45	0.87	<.001
47.5	0.58	<.001
52.5	0.58	<.001
55	-0.25	0.078

Essentially what is happening is that the 47.5 has equal density just to the left and right, resulting in no worthwhile trade of FP vs FN as we move in either direction. Thus, the optimal cutscore effectively settles on the true cutscore

47.5 Condition







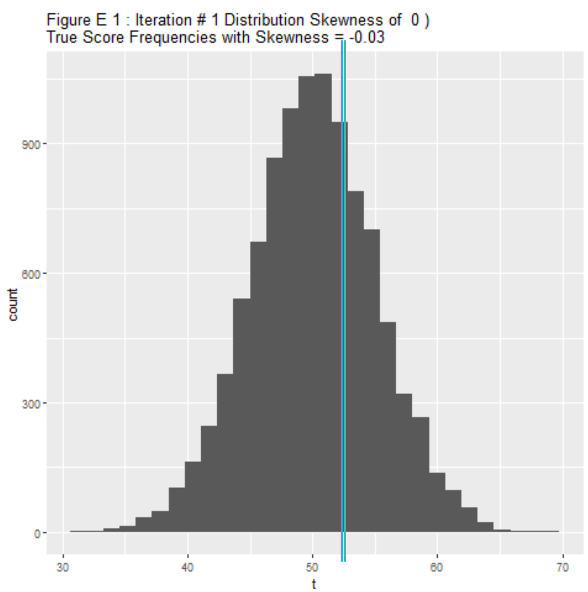
HYPOTHESIS 4: EVIDENCE TO SUPPORT

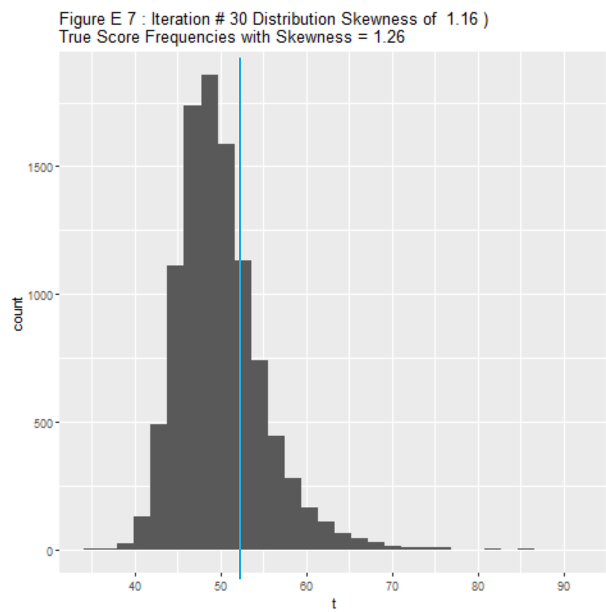
Skewness Results

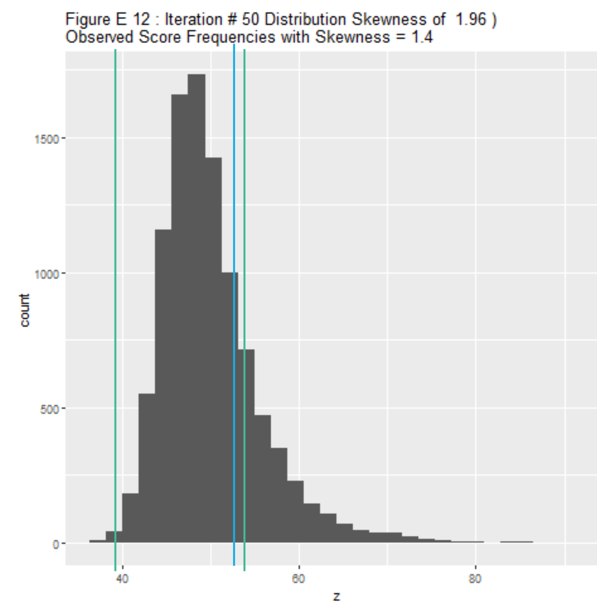
True Cut Location	Optimum Cutscore Location	
	Spearman's Rho	p
45	0.87	<.001
47.5	0.58	<.001
52.5	0.58	<.001
55	-0.25	0.078

Essentially what is happening here is that the shape of the distribution just to the left and to the right of the true cutscore remains more or less constant throughout. So, we do get a slight movement to the right as skew increases, but not a whole lot

52.5 Condition







HYPOTHESIS 4: EVIDENCE TO SUPPORT

Skewness Results

True Cut Location	Optimum Cutscore Location	
	Spearman's Rho	p
45	0.87	<.001
47.5	0.58	<.001
52.5	0.58	<.001
55	-0.25	0.078

Essentially what is happening here is that the shape of the distribution just to the left and to the right of the true cutscore remains more or less constant throughout. We see no real movement at all.

55 Condition

Figure E 1 : Iteration # 1 Distribution Skewness of 0)
True Score Frequencies with Skewness = -0.03

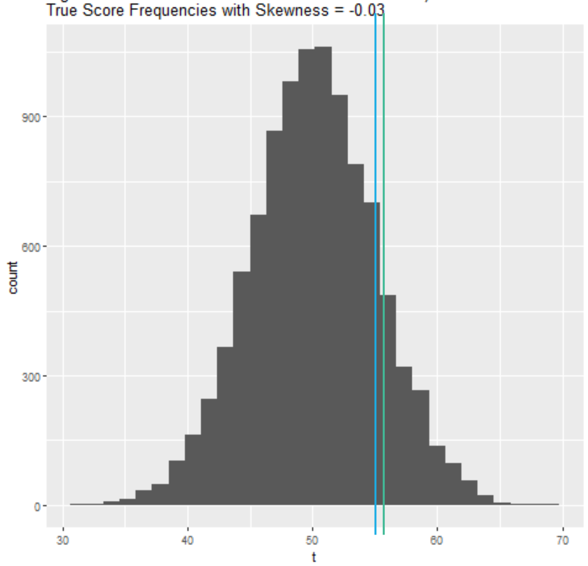


Figure E 7 : Iteration # 30 Distribution Skewness of 1.16)
True Score Frequencies with Skewness = 1.26

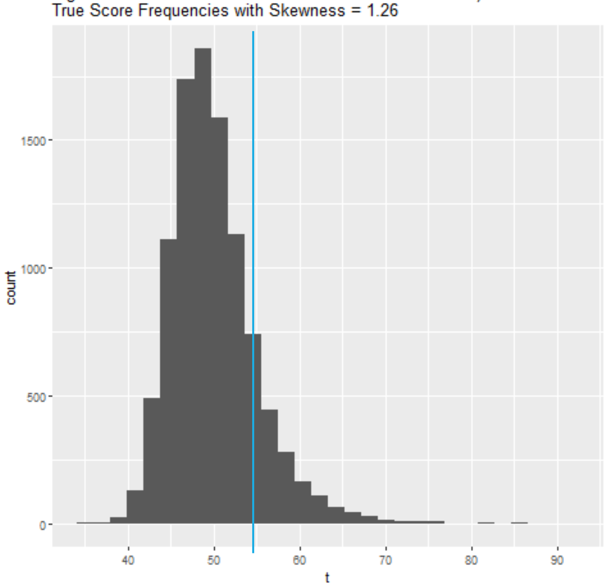
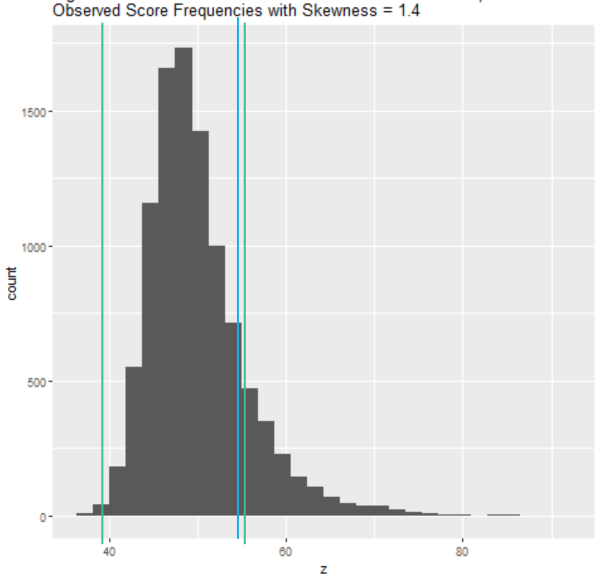


Figure E 12 : Iteration # 50 Distribution Skewness of 1.96)
Observed Score Frequencies with Skewness = 1.4



HYPOTHESIS 4 (CONTINUED)

Bimodal Results

	Optimum Cutscore Location	
True Cut Location	Spearman's Rho	p
45	0.4	0.004
47.5	0.86	<.001
52.5	0.87	<.001
55	0.5	<.001

45 Condition

Figure F 21 : Iteration # 36 (distance of mixture means= 7)
True Score Frequencies with D = 1.5

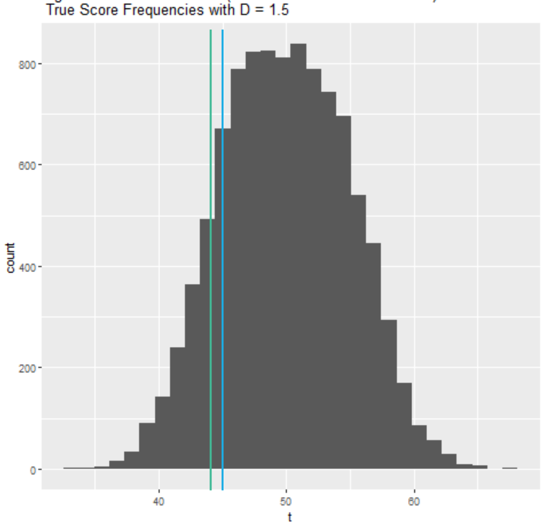


Figure F 29 : Iteration # 40 (distance of mixture means= 7.8)
True Score Frequencies with D = 7.4

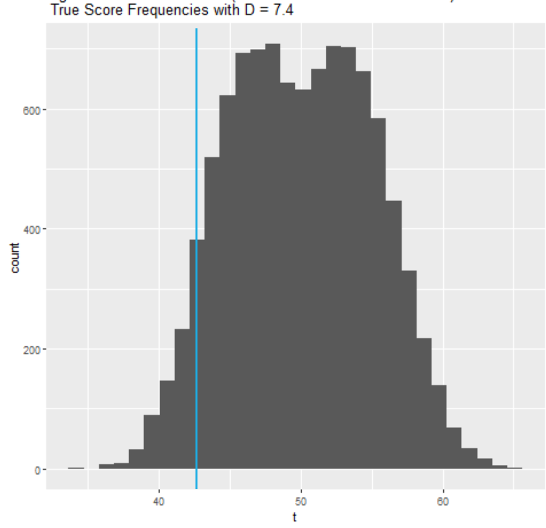
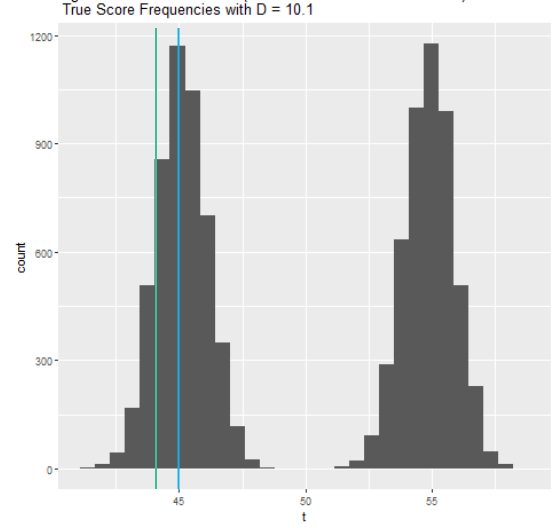


Figure F 49 : Iteration # 50 (distance of mixture means= 9.8)
True Score Frequencies with D = 10.1

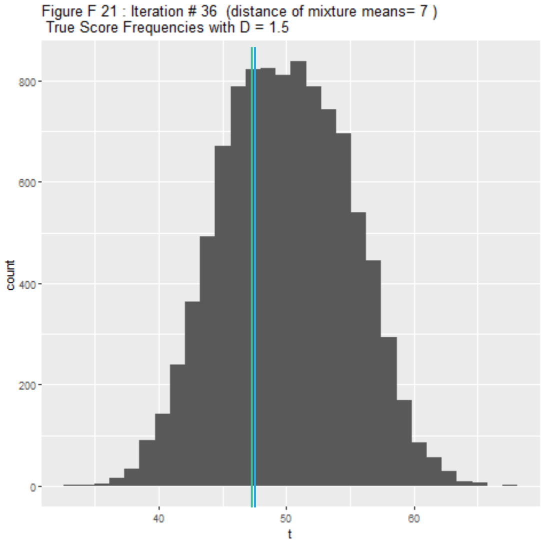


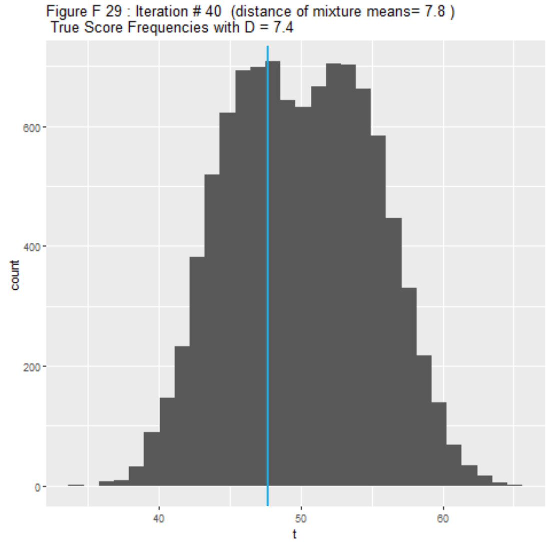
HYPOTHESIS 4 (CONTINUED)

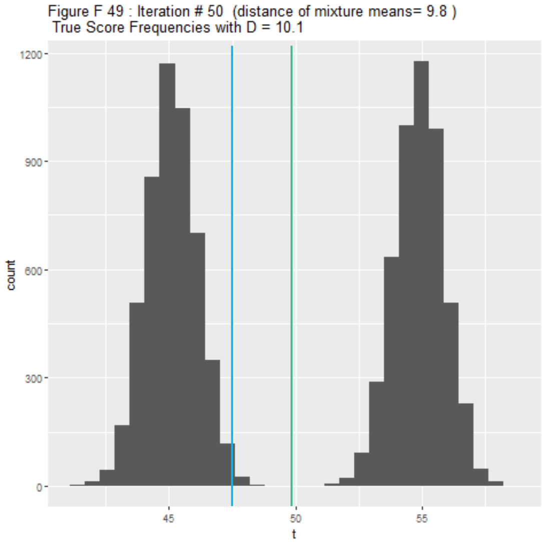
Bimodal Results

	Optimum Cutscore Location	
True Cut Location	Spearman's Rho	p
45	0.4	0.004
47.5	0.86	<.001
52.5	0.87	<.001
55	0.5	<.001

47.5 Condition





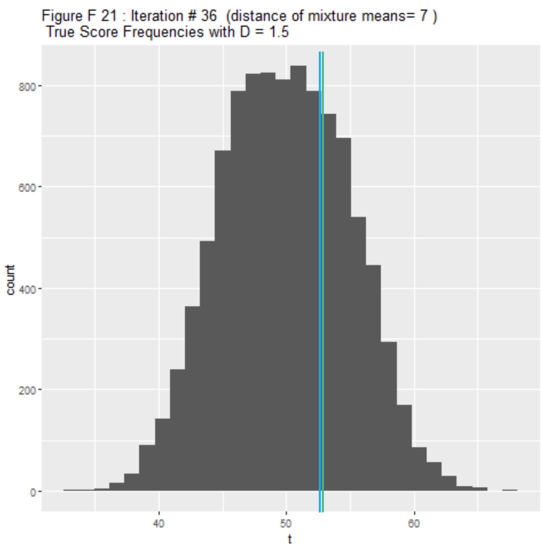


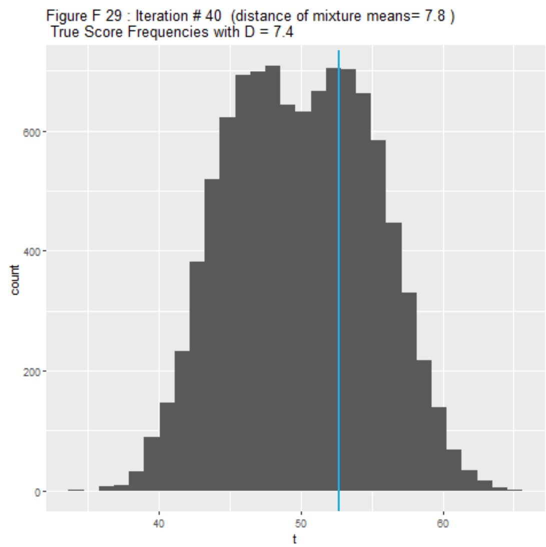
HYPOTHESIS 4 (CONTINUED)

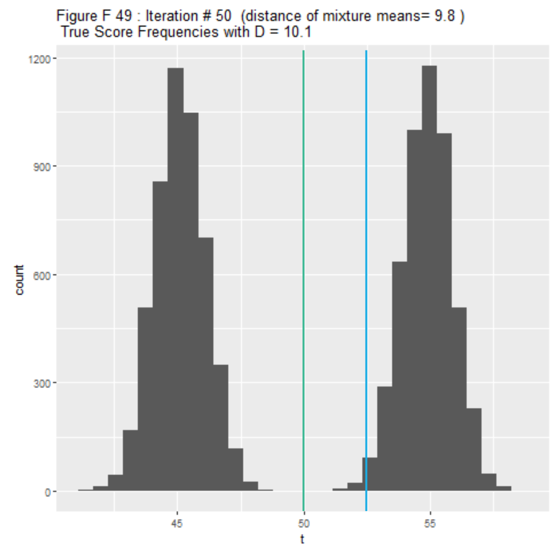
Bimodal Results

	Optimum Cutscore Location	
True Cut Location	Spearman's Rho	p
45	0.4	0.004
47.5	0.86	<.001
52.5	0.87	<.001
55	0.5	<.001

52.5 Condition







HYPOTHESIS 4 (CONTINUED)

Bimodal Results

	Optimum Cutscore Location	
True Cut Location	Spearman's Rho	p
45	0.4	0.004
47.5	0.86	<.001
52.5	0.87	<.001
55	0.5	<.001

55 Condition

Figure F 21 : Iteration # 36 (distance of mixture means= 7)
True Score Frequencies with D = 1.5

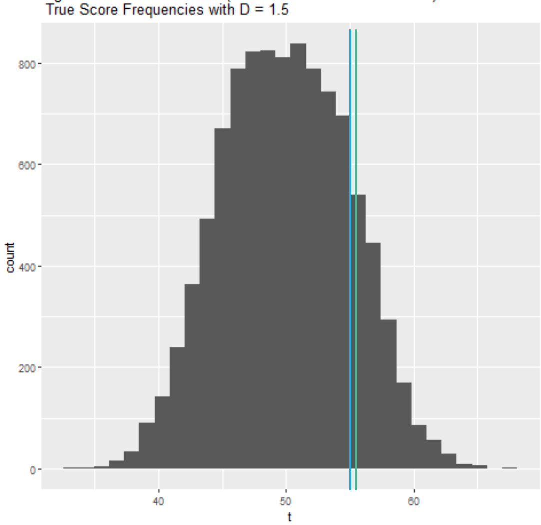


Figure F 29 : Iteration # 40 (distance of mixture means= 7.8)
True Score Frequencies with D = 7.4

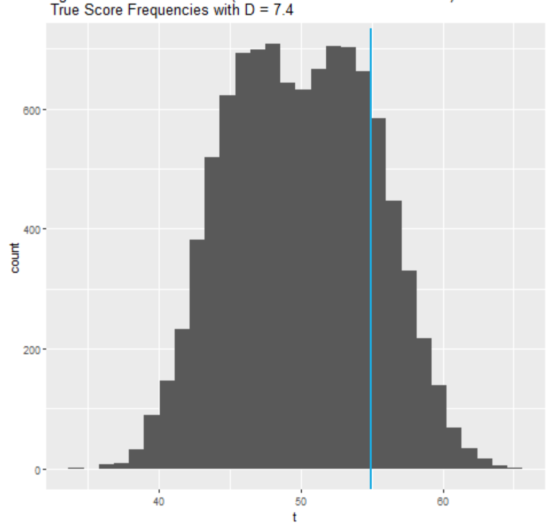
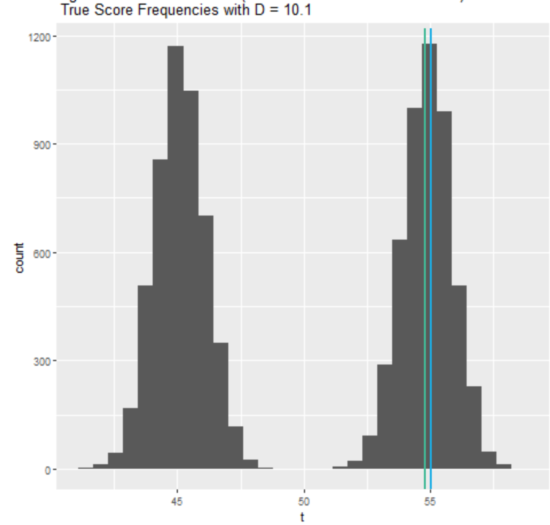


Figure F 49 : Iteration # 50 (distance of mixture means= 9.8)
True Score Frequencies with D = 10.1



HYPOTHESIS 4 (CONTINUED)

Kurtosis Results

	Optimum Cutscore Location	
True Cut Location	Spearman's Rho	p
45	0.35	0.011
47.5	0.89	<.001
52.5	0.9	<.001
55	0.22	0.118

45 Condition

Figure G 1 : Iteration # 1 (Distribution Kurtosis of 3)
True Score Frequencies with Kurtosis = 3.03

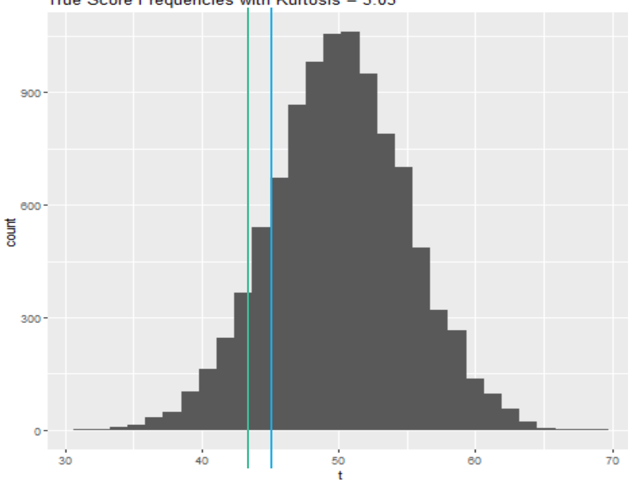


Figure G 7 : Iteration # 30 (Distribution Kurtosis of 4.74)
True Score Frequencies with Kurtosis = 5.02

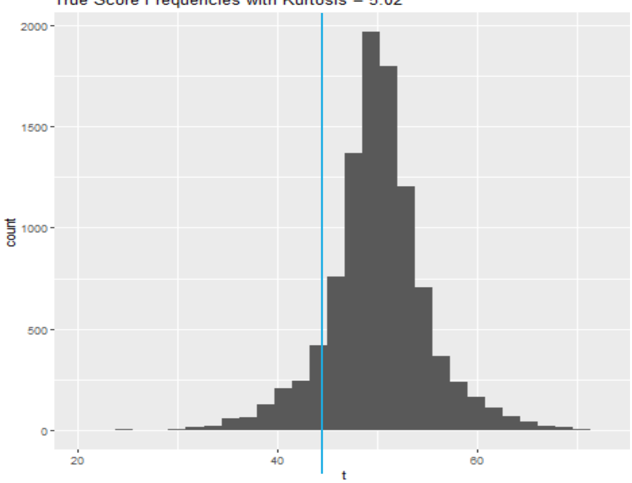
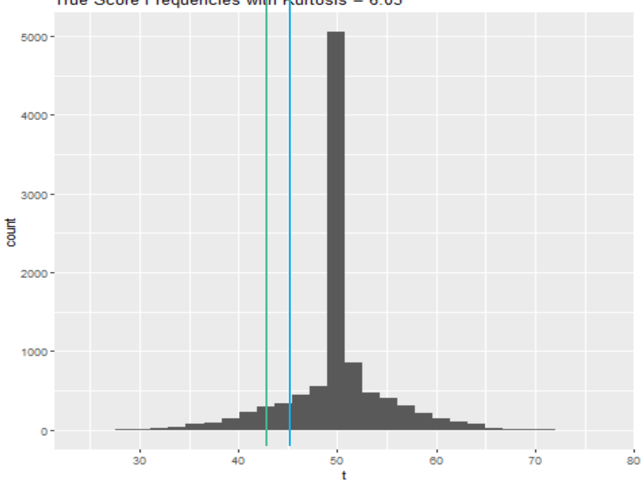


Figure G 11 : Iteration # 50 (Distribution Kurtosis of 5.94)
True Score Frequencies with Kurtosis = 6.03



HYPOTHESIS 4 (CONTINUED)

Kurtosis Results

	Optimum Cutscore Location	
True Cut Location	Spearman's Rho	p
45	0.35	0.011
47.5	0.89	<.001
52.5	0.9	<.001
55	0.22	0.118

47.5 Condition

Figure G 1 : Iteration # 1 (Distribution Kurtosis of 3)
True Score Frequencies with Kurtosis = 3.03

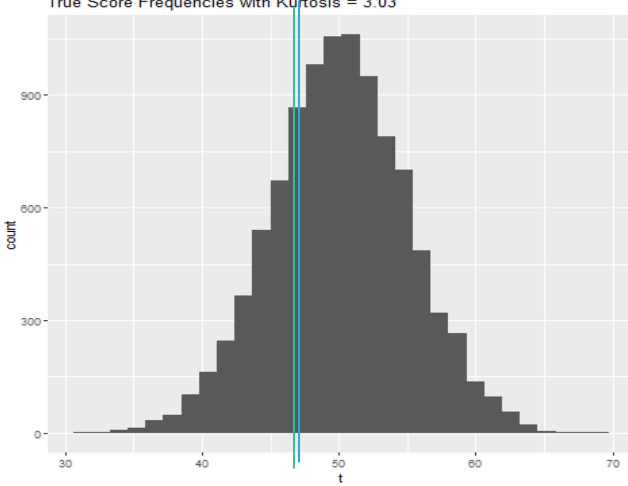


Figure G 7 : Iteration # 30 (Distribution Kurtosis of 4.74)
True Score Frequencies with Kurtosis = 5.02

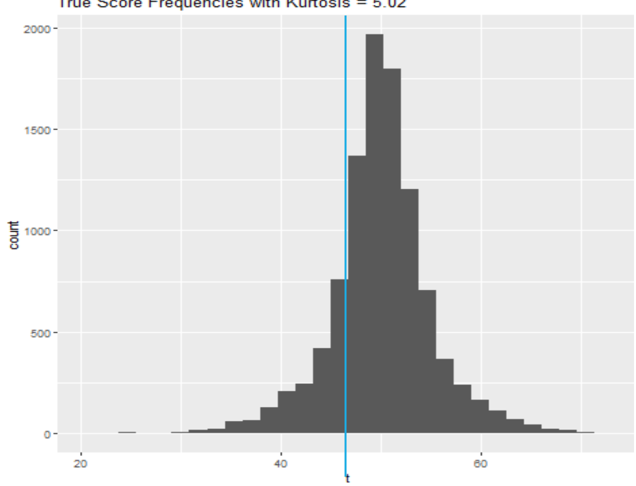
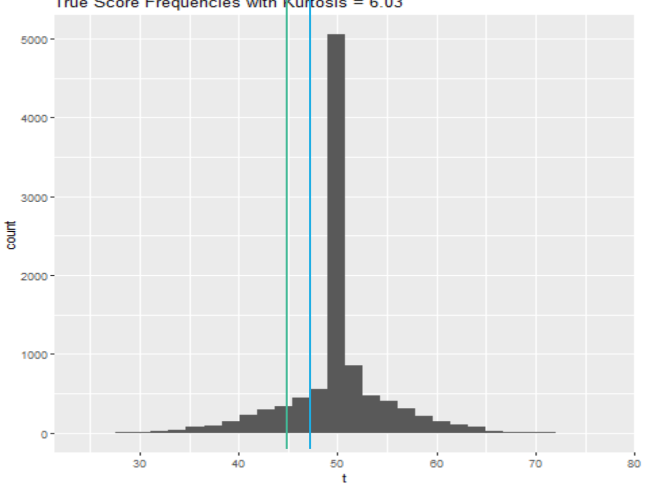


Figure G 11 : Iteration # 50 (Distribution Kurtosis of 5.94)
True Score Frequencies with Kurtosis = 6.03

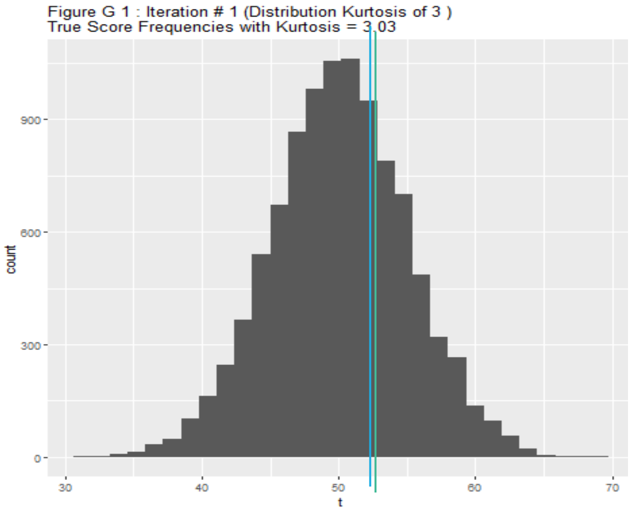


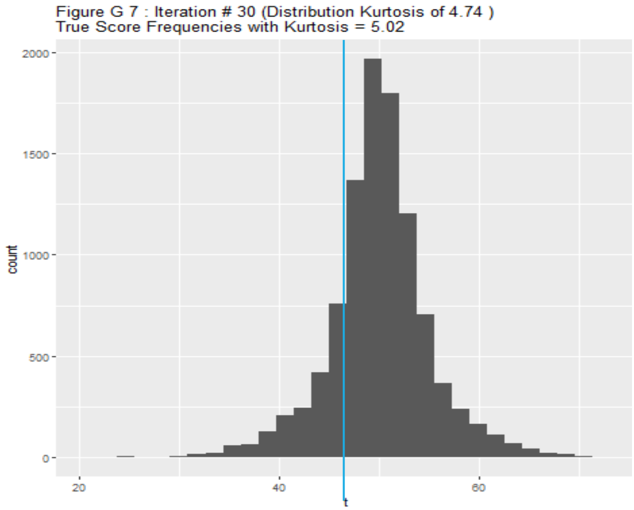
HYPOTHESIS 4 (CONTINUED)

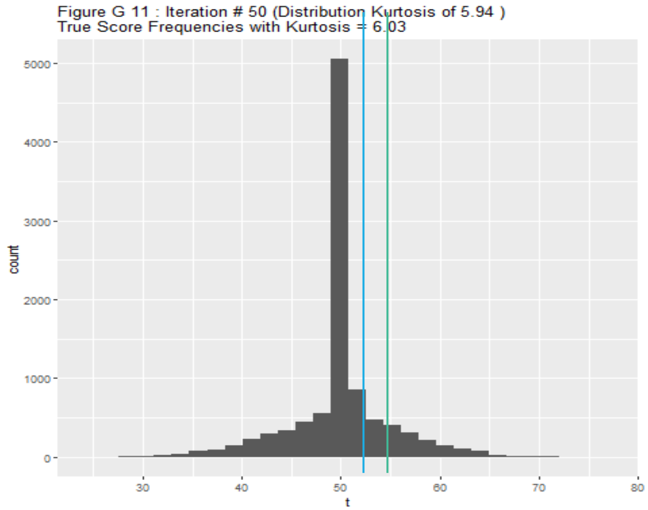
Kurtosis Results

	Optimum Cutscore Location	
True Cut Location	Spearman's Rho	p
45	0.35	0.011
47.5	0.89	<.001
52.5	0.9	<.001
55	0.22	0.118

52.5 Condition





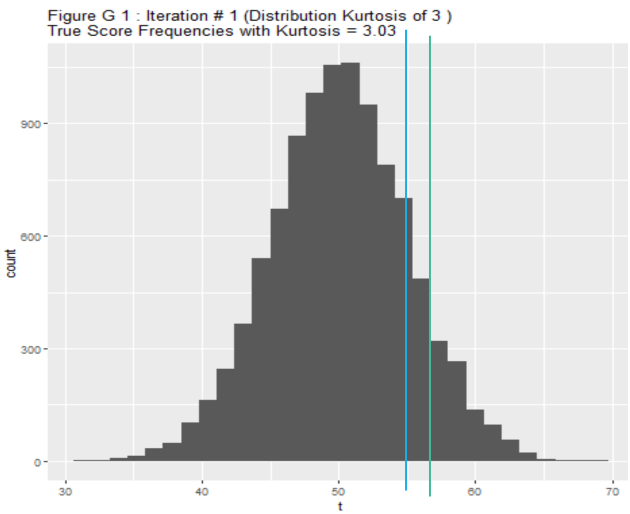


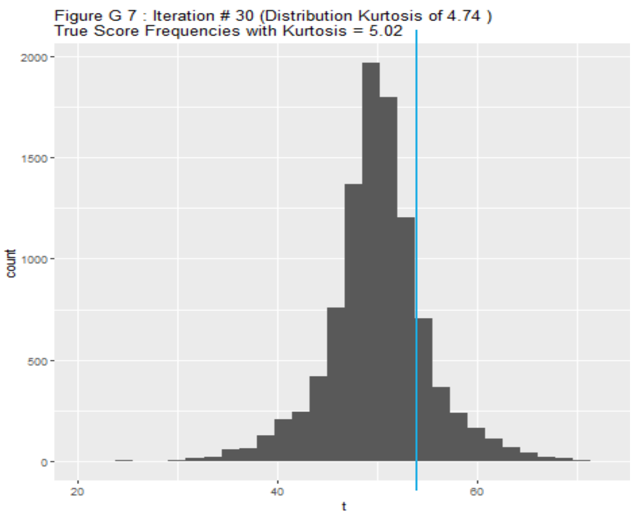
HYPOTHESIS 4 (CONTINUED)

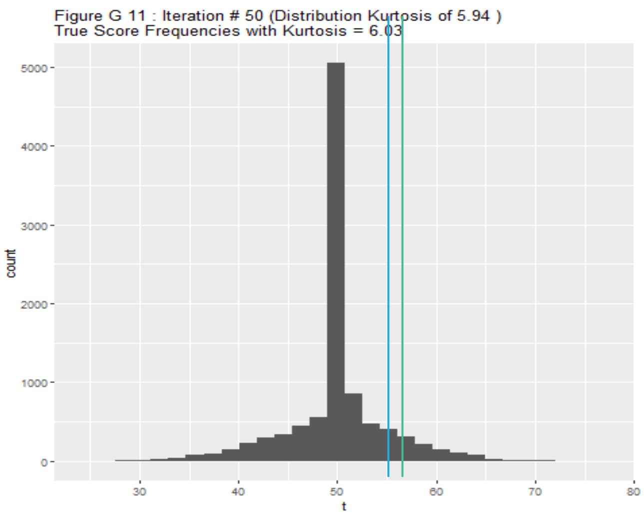
Kurtosis Results

	Optimum Cutscore Location	
True Cut Location	Spearman's Rho	p
45	0.35	0.011
47.5	0.89	<.001
52.5	0.9	<.001
55	0.22	0.118

55 Condition







DISCUSSION

Consequential validity (of the GW-CSOF)

- The single most important factor to consider in weighing the GW-CSOF is by the consequences of using it
- The GW-CSOF has two likely consequences that follow from its use:
  - Information
  - Decisions based on that information
- We consider these two consequences with three different types of examinee distributions
  - Normal to minimally non-normal
  - Moderately non-normal
  - Largely non-normal

MORE IMPORTANTLY…

Are the differences meaningful?!

Results were divided into three mutually exclusive categories per condition

True score skewness of .63, 1.32, and 1.97, denoted ‘minutely non-normal’

True score bimodality D of 6.1, 8.9, and 9.9, denoted ‘moderately non-normal’

True score kurtosis of 3.9, 5.1, and 6.03, denoted ‘largely non-normal’

Difference between actual error at actual optimal cutscore & actual error at estimated optimal cutscore

*Note: Difference is actual optimal error - actual error at the GW-CSOF estimated optimal location. NA's denote non-significant results.

	45	47.5	52.5	55
Minute	ΔTot.	ΔTot.	ΔTot.	ΔTot.
Skew	0.00	0.00	0.00	NA
Bimodal	0.00	-0.01	-0.01	0.00
Kurtosis	NA	0.00	0.00	NA
Moderate				
Skew	0.00	0.00	0.00	NA
Bimodal	0.00	-0.03	-0.02	0.00
Kurtosis	NA	-0.01	-0.01	NA
Large				
Skew	-0.05	-0.01	-0.01	NA
Bimodal	0.00	-0.09	-0.09	-0.01
Kurtosis	NA	-0.03	-0.03	NA

Difference between actual and GW-CSOF estimate of error at true cutscore

*Note: Difference is actual error at true cutscore - GW-CSOF estimate of error at true cutscore.

	45	47.5	52.5	55
Minute	ΔTot.	ΔTot.	ΔTot.	ΔTot.
Skew	0.01	0.01	-0.01	-0.01
Bimodal	0.00	0.00	0.00	0.00
Kurtosis	-0.01	0.02	0.00	-0.01
Moderate				
Skew	0.03	0.04	-0.02	-0.02
Bimodal	0.01	-0.01	-0.01	0.01
Kurtosis	-0.03	0.01	0.01	-0.03
Large				
Skew	0.05	0.07	-0.03	-0.04
Bimodal	0.09	-0.04	-0.04	0.09
Kurtosis	-0.04	0.00	0.00	-0.04

QUESTIONS? COMMENTS?