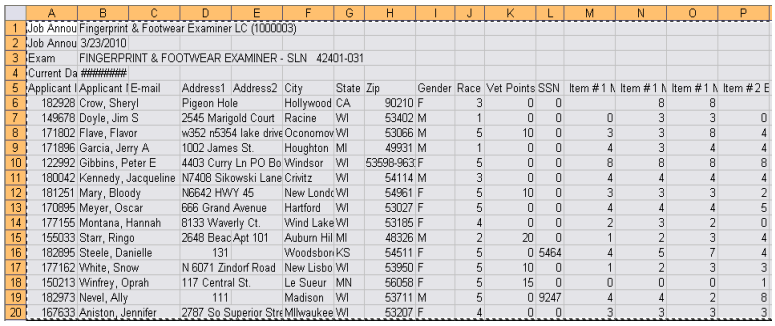
BUREAU OF MERIT RECRUITMENT AND SELECTION

SCORE ANALYSIS for RATED EXAMS

Table of Contents

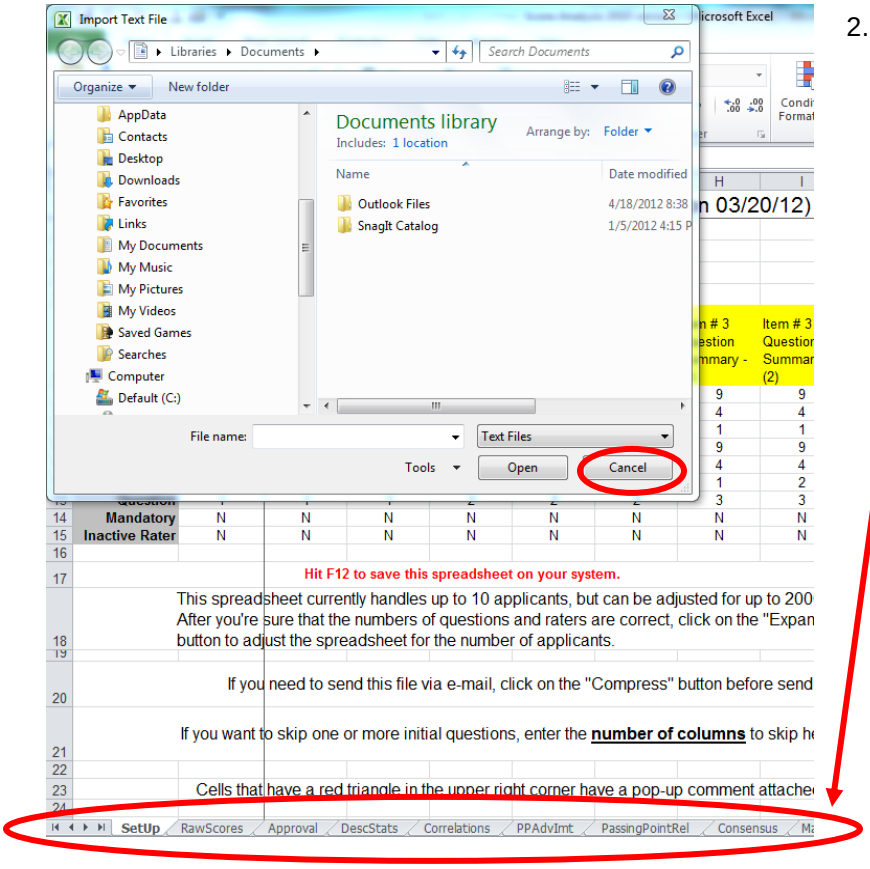
Copy Export of Rater Scores into Score Analysis Spreadsheet:



1. When the export file opens, use these keys in this order:

   Ctrl-A *to select all of the information*

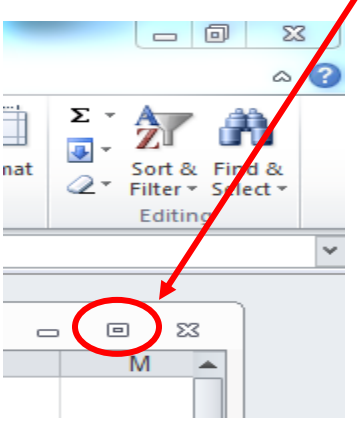
   Ctrl-C *to copy the information*

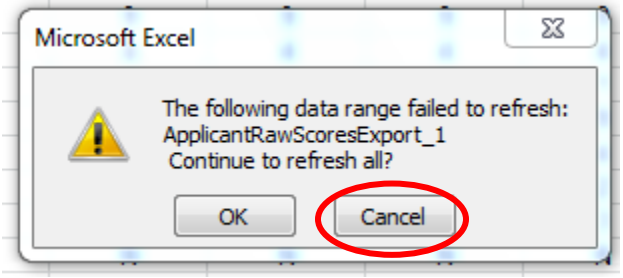


2. Open the **Score Analysis Spreadsheet** saved on the computer.

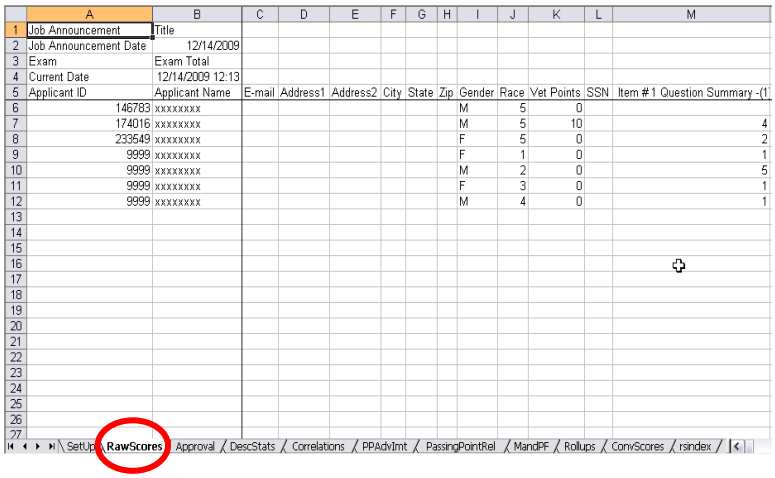
   To "paste" the copied score export data into the spreadsheet select **[Cancel]**

Note: *If you don't see the tabs at the bottom of the spreadsheet, expand the spreadsheet by clicking on box symbol in the upper right corner of the viewing window to fill the screen.*





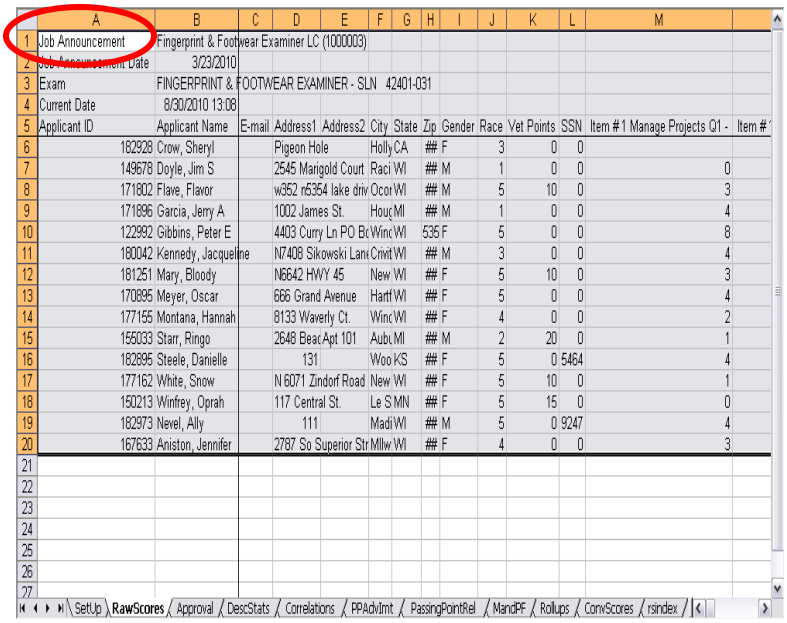
3. At this prompt select **[Cancel]**



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Job Announcement	Title											
2	Job Announcement Date	12/14/2009											
3	Exam	Exam Total											
4	Current Date	12/14/2009 12:13											
5	Applicant ID	Applicant Name	E-mail	Address1	Address2	City	State	Zip	Gender	Race	Vet Points	SSN	Item # 1 Question Summary -(1
6	146783	xxxxxxxx							M	5	0		
7	174016	xxxxxxxx							M	5	10		4
8	233549	xxxxxxxx							F	5	0		2
9	9999	xxxxxxxx							F	1	0		1
10	9999	xxxxxxxx							M	2	0		5
11	9999	xxxxxxxx							F	3	0		1
12	9999	xxxxxxxx							M	4	0		1

SetUp | RawScores | Approval | DescStats | Correlations | PPAdvImt | PassingPointRel | MandPF | Rollups | ConvScores | rsindex

4. Select the **[Raw Scores]** tab on the analysis spreadsheet.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Job Announcement	Fingerprint & Footwear Examiner LC (1000003)											
2	Job Announcement Date	3/23/2010											
3	Exam	FINGERPRINT & FOOTWEAR EXAMINER - SLN 42401-031											
4	Current Date	8/30/2010 13:08											
5	Applicant ID	Applicant Name	E-mail	Address1	Address2	City	State	Zip	Gender	Race	Vet Points	SSN	Item # 1 Manage Projects Q1 -	Item #
6	182928	Crow, Sheryl		Pigeon Hole		Holly	CA	##	F	3	0	0		
7	149678	Doyle, Jim S		2545 Marigold Court		Raci	WI	##	M	1	0	0	0	
8	171802	Flave, Flavor		w352 n5354 lake driv		Ocor	WI	##	M	5	10	0	3	
9	171896	Garcia, Jerry A		1002 James St.		Hou	MI	##	M	1	0	0	4	
10	122992	Gibbins, Peter E		4403 Curry Ln PO Bo		Winc	WI	535	F	5	0	0	8	
11	180042	Kennedy, Jacqueline		N7408 Sikowski Lan		Crivit	WI	##	M	3	0	0	4	
12	181251	Mary, Bloody		N6642 HWY 45		New	WI	##	F	5	10	0	3	
13	170895	Meyer, Oscar		666 Grand Avenue		Hartf	WI	##	F	5	0	0	4	
14	177155	Montana, Hannah		8133 Waverly Ct.		Winc	WI	##	F	4	0	0	2	
15	155033	Starr, Ringo		2648 Beac	Apt 101	Aubu	MI	##	M	2	20	0	1	
16	182895	Steele, Danielle		131		Woo	KS	##	F	5	0	5464	4	
17	177162	White, Snow		N 6071 Zindorf Road		New	WI	##	F	5	10	0	1	
18	150213	Winfrey, Oprah		117 Central St.		Le S	MN	##	F	5	15	0	0	
19	182973	Nevel, Ally		111		Madi	WI	##	M	5	0	9247	4	
20	167633	Aniston, Jennifer		2787 So Superior Str		Mllw	WI	##	F	4	0	0	3	

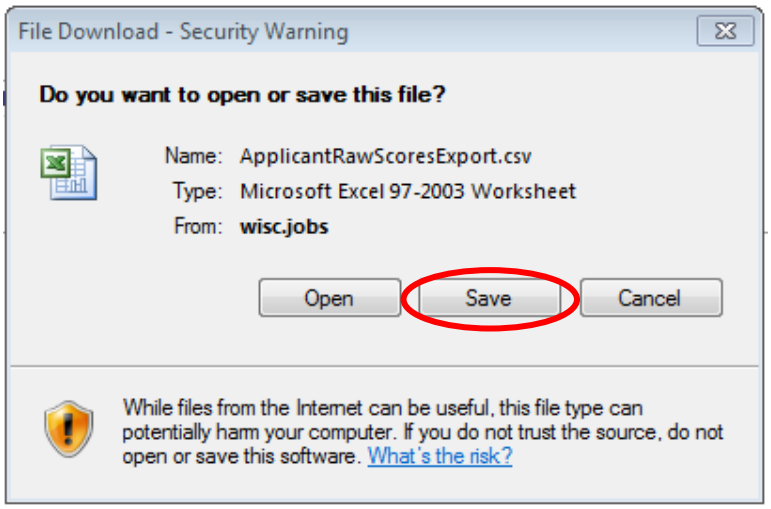
SetUp | RawScores | Approval | DescStats | Correlations | PPAdvImt | PassingPointRel | MandPF | Rollups | ConvScores | rsindex

5. Highlight **cell A1** (if not already), paste the copied information from the Rater Export file into the **[RawScores]** tab by using **Ctrl-V**.

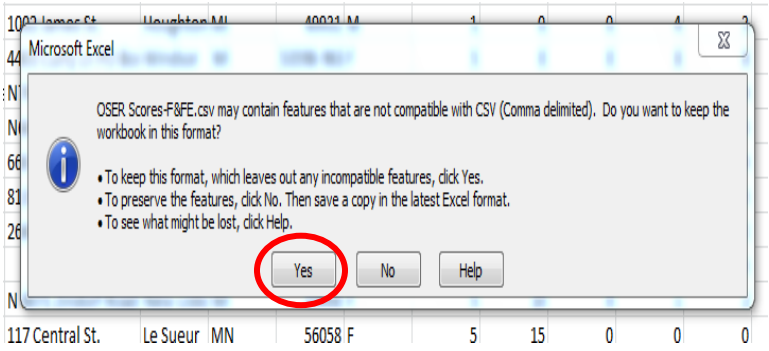
Go to section [Verifying and Adjusting the Set Up:]

Saving the Export File to the Desktop:

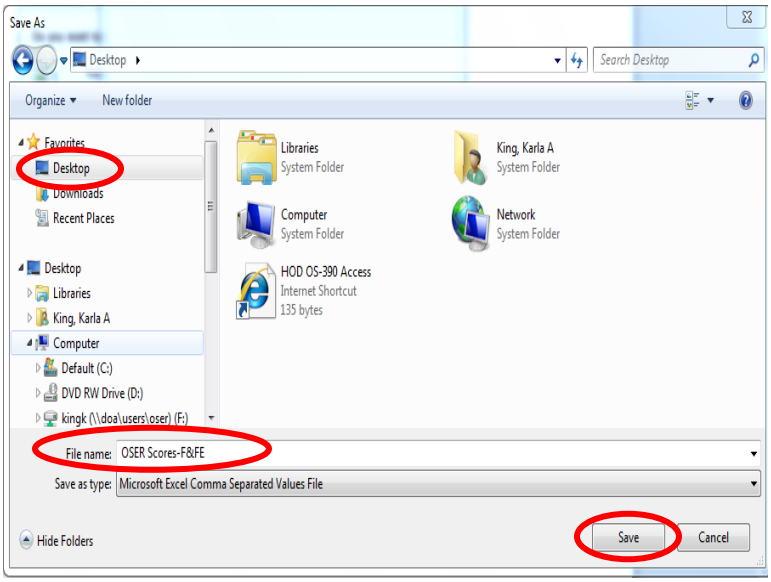
Note: Open the Score Analysis Spreadsheet only after the export file has been saved to the computer.



1. At the **File Download** prompt, click on **[Save]**. Save the export file to the desktop or a folder on the computer where it can be accessed to import into the Score Analysis Spreadsheet.

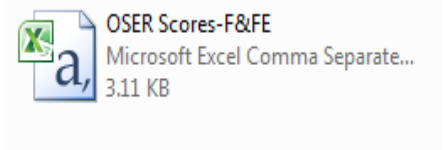


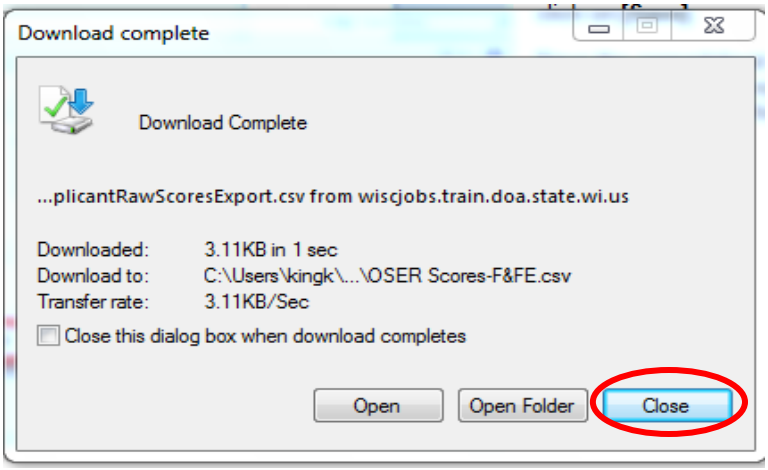
2. At this prompt, click on **[Yes]** to save in the current format (.csv).



3. Click on **[Desktop]** and give the file a specific name then click to **[Save]**.

   This file will be saved on the desktop:

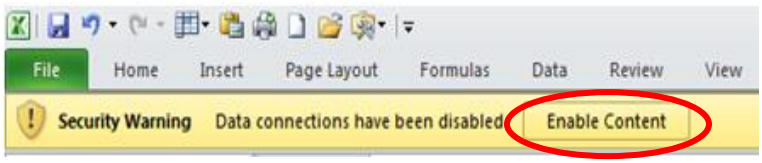




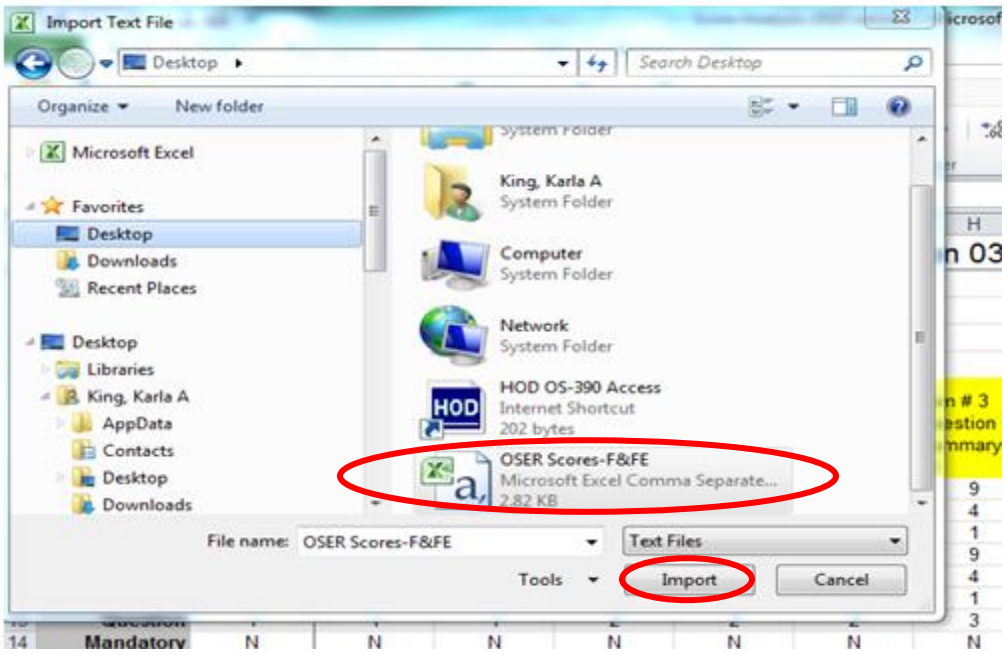
4. Once Download is complete, click on **[Close].**

Import the Score Export File into the Score Analysis Spreadsheet:

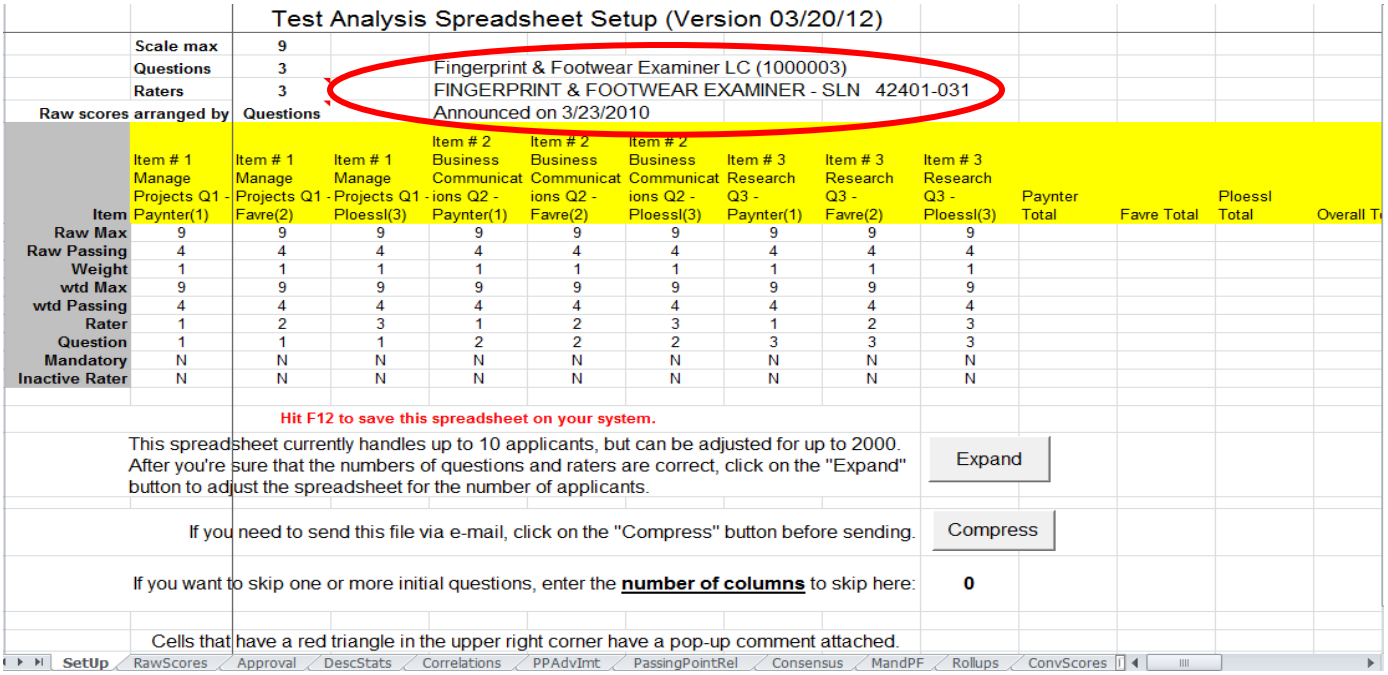
1. **Open Score Analysis Spreadsheet.** If this appears at the top of the screen, click on **[Enable Content]** in order to continue.



2. **Click on the export file saved on the desktop, then click on [Import].** This will automatically load the exam score information into the spreadsheet.



3. **Exam specific information will appear at the top of the Set Up tab.**



Test Analysis Spreadsheet Setup (Version 03/20/12)

	Scale max	9
	Questions	3
	Raters	3
Raw scores	arranged by	Questions

Fingerprint & Footwear Examiner LC (1000003)
FINGERPRINT & FOOTWEAR EXAMINER - SLN 42401-031
Announced on 3/23/2010

Item	Item # 1 Manage Projects Q1 - Paynter(1)	Item # 1 Manage Projects Q1 - Favre(2)	Item # 1 Manage Projects Q1 - Ploessl(3)	Item # 2 Business Communications Q2 - Paynter(1)	Item # 2 Business Communications Q2 - Favre(2)	Item # 2 Business Communications Q2 - Ploessl(3)	Item # 3 Research Q3 - Paynter(1)	Item # 3 Research Q3 - Favre(2)	Item # 3 Research Q3 - Ploessl(3)	Paynter Total	Favre Total	Ploessl Total	Overall T
Raw Max	9	9	9	9	9	9	9	9	9				
Raw Passing	4	4	4	4	4	4	4	4	4				
Weight	1	1	1	1	1	1	1	1	1				
wtd Max	9	9	9	9	9	9	9	9	9				
wtd Passing	4	4	4	4	4	4	4	4	4				
Rater	1	2	3	1	2	3	1	2	3				
Question	1	1	1	2	2	2	3	3	3				
Mandatory	N	N	N	N	N	N	N	N	N				
Inactive Rater	N	N	N	N	N	N	N	N	N				

Hit F12 to save this spreadsheet on your system.

This spreadsheet currently handles up to 10 applicants, but can be adjusted for up to 2000. After you're sure that the numbers of questions and raters are correct, click on the "Expand" button to adjust the spreadsheet for the number of applicants. [Expand]

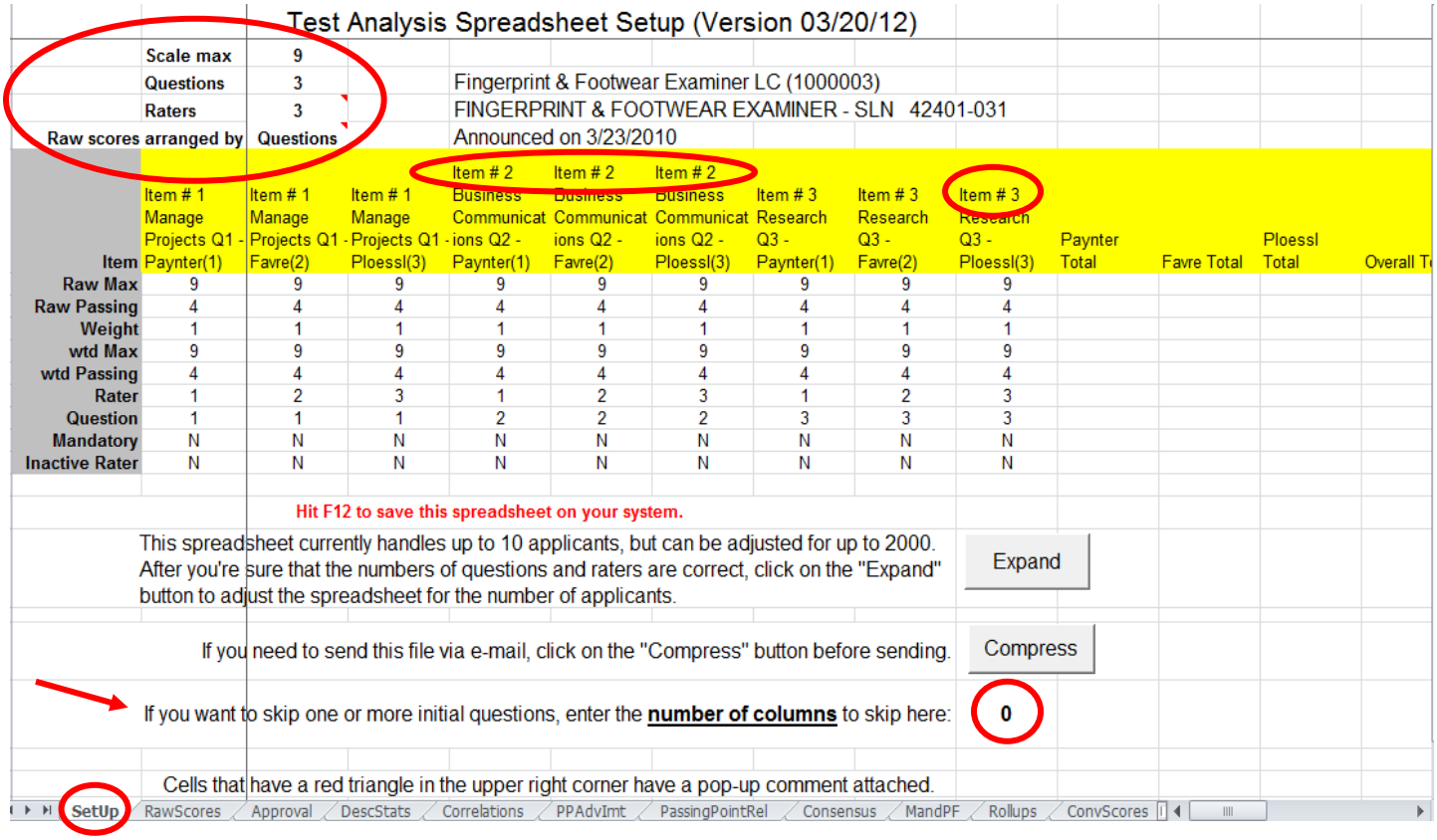
If you need to send this file via e-mail, click on the "Compress" button before sending. [Compress]

If you want to skip one or more initial questions, enter the **number of columns** to skip here: **0**

Cells that have a red triangle in the upper right corner have a pop-up comment attached.

SetUp | RawScores | Approval | DescStats | Correlations | PPAdvImt | PassingPointRel | Consensus | MandPF | Rollups | ConvScores

1. **Click on the [Set Up] Tab.**



Test Analysis Spreadsheet Setup (Version 03/20/12)

Scale max	9	
Questions	3	Fingerprint & Footwear Examiner LC (1000003)
Raters	3	FINGERPRINT & FOOTWEAR EXAMINER - SLN 42401-031
Raw scores arranged by	Questions	Announced on 3/23/2010

Item	Item # 1 Manage Projects Q1 - Paynter(1)	Item # 1 Manage Projects Q1 - Favre(2)	Item # 1 Manage Projects Q1 - Ploessl(3)	Item # 2 Business Communications Q2 - Paynter(1)	Item # 2 Business Communications Q2 - Favre(2)	Item # 2 Business Communications Q2 - Ploessl(3)	Item # 3 Research Q3 - Paynter(1)	Item # 3 Research Q3 - Favre(2)	Item # 3 Research Q3 - Ploessl(3)	Paynter Total	Favre Total	Ploessl Total	Overall T
Raw Max	9	9	9	9	9	9	9	9	9				
Raw Passing	4	4	4	4	4	4	4	4	4				
Weight	1	1	1	1	1	1	1	1	1				
wtd Max	9	9	9	9	9	9	9	9	9				
wtd Passing	4	4	4	4	4	4	4	4	4				
Rater	1	2	3	1	2	3	1	2	3				
Question	1	1	1	2	2	2	3	3	3				
Mandatory	N	N	N	N	N	N	N	N	N				
Inactive Rater	N	N	N	N	N	N	N	N	N				

Hit F12 to save this spreadsheet on your system.

This spreadsheet currently handles up to 10 applicants, but can be adjusted for up to 2000. After you're sure that the numbers of questions and raters are correct, click on the "Expand" button to adjust the spreadsheet for the number of applicants. [Expand]

If you need to send this file via e-mail, click on the "Compress" button before sending. [Compress]

If you want to skip one or more initial questions, enter the **number of columns** to skip here: 0

Cells that have a red triangle in the upper right corner have a pop-up comment attached.

SetUp | RawScores | Approval | DescStats | Correlations | PPAdvImt | PassingPointRel | Consensus | MandPF | Rollups | ConvScores

2. **Review the following fields to make any necessary adjustments:**

Scale Max: the highest possible score per question = 9
- This number should match the max # found in the **Exam Score Criteria** per question.

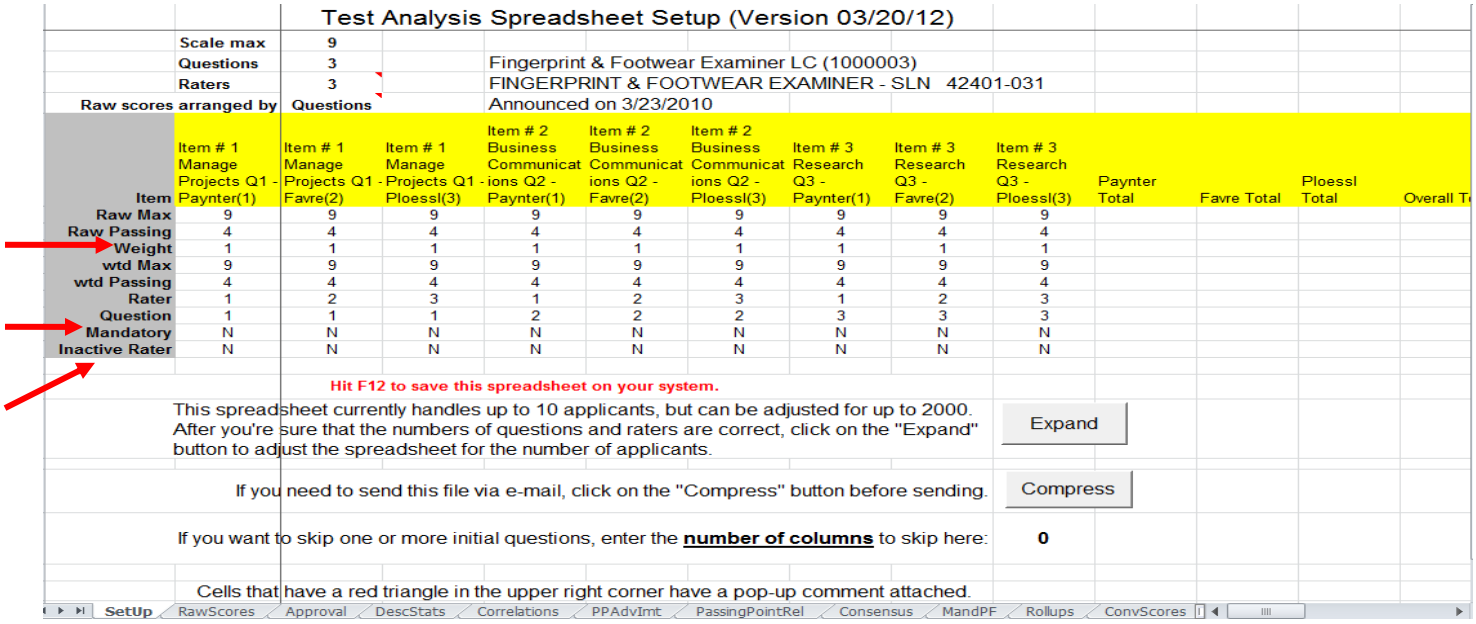
Questions: the total # of items/questions in the exam (see top yellow header on right e.g. Item #3)
- This number should match the # of questions on the **[RawScores]** tab.
- If different, the data may need to be pasted again into cell A1 of the **[RawScores]** tab OR if the data looks fine, just enter the correct # of questions.

Raters: the total # of raters used to score the exam; most times is = 3 (see top header, there are 3 columns for Item #2) A minimum of 2 raters is required by statue, however, there can more than 3 raters used in certain situations)
- This number should match the # of raters on the **[RawScores]** tab
- If different, enter the correct # of raters
- If an OIQ, # of raters = 1

Raw scores arranged by:
- The default is set as “Questions” if using Rater Export file to populate this spreadsheet. The other choice is by “Raters”

Optional: number of columns to skip - if the exam combines Non-Rated with Rated questions (i.e., Minimum Requirements with narrative responses) determine the # of non-rated questions (at the beginning of the exam); this is the # of columns to skip when performing the score analysis

3. **Review the setup and/or the Exam Plan Checklist for exact details about the following variables.**



Test Analysis Spreadsheet Setup (Version 03/20/12)

Scale max	9	
Questions	3	Fingerprint & Footwear Examiner LC (1000003)
Raters	3	FINGERPRINT & FOOTWEAR EXAMINER - SLN 42401-031
Raw scores arranged by	Questions	Announced on 3/23/2010

Item	Item # 1 Manage Projects Q1 - Paynter(1)	Item # 1 Manage Projects Q1 - Favre(2)	Item # 1 Manage Projects Q1 - Ploessl(3)	Item # 2 Business Communications Q2 - Paynter(1)	Item # 2 Business Communications Q2 - Favre(2)	Item # 2 Business Communications Q2 - Ploessl(3)	Item # 3 Research Q3 - Paynter(1)	Item # 3 Research Q3 - Favre(2)	Item # 3 Research Q3 - Ploessl(3)	Paynter Total	Favre Total	Ploessl Total	Overall T
Raw Max	9	9	9	9	9	9	9	9	9				
Raw Passing	4	4	4	4	4	4	4	4	4				
Weight	1	1	1	1	1	1	1	1	1				
wtd Max	9	9	9	9	9	9	9	9	9				
wtd Passing	4	4	4	4	4	4	4	4	4				
Rater	1	2	3	1	2	3	1	2	3				
Question	1	1	1	2	2	2	3	3	3				
Mandatory	N	N	N	N	N	N	N	N	N				
Inactive Rater	N	N	N	N	N	N	N	N	N				

Hit F12 to save this spreadsheet on your system.

This spreadsheet currently handles up to 10 applicants, but can be adjusted for up to 2000. After you're sure that the numbers of questions and raters are correct, click on the "Expand" button to adjust the spreadsheet for the number of applicants. Expand

If you need to send this file via e-mail, click on the "Compress" button before sending. Compress

If you want to skip one or more initial questions, enter the **number of columns** to skip here: 0

Cells that have a red triangle in the upper right corner have a pop-up comment attached.

SetUp | RawScores | Approval | DescStats | Correlations | PPAdvImt | PassingPointRel | Consensus | MandPF | Rollups | ConvScores

Weight: If all questions are weighted equally, this = 1

If not, adjust based on the weight score criteria (must be adjusted for each rater per question).

Mandatory: auto populates to N for "No"

Enter "Y" for Yes if a passing score is required on a question in order to pass the exam.

Inactive Rater: auto populates to N or "No"

Enter "Y" for Yes, if it has been determined that a rater's scores should not count.

The following fields are populated by the loaded or adjusted data and usually do not require any further adjustment:

Raw Max: the raw score maximum for each question. The top header (in yellow) has 9 columns with data. This is because the spreadsheet is expecting a raw max score (as well as other data) for each rater on each question. Since there are 3 questions and 3 raters, there are 9 columns (3 X 3 = 9). This data will auto populate for the 0-9 point scale with a passing point of 4, but can be typed over if it needs to be adjusted.

Raw Passing: Raw passing point per question, auto populates as 4.

wtd Max: This is the weighted maximum for each question. It will auto calculate based on what is entered in the Raw Max and Weight cells.

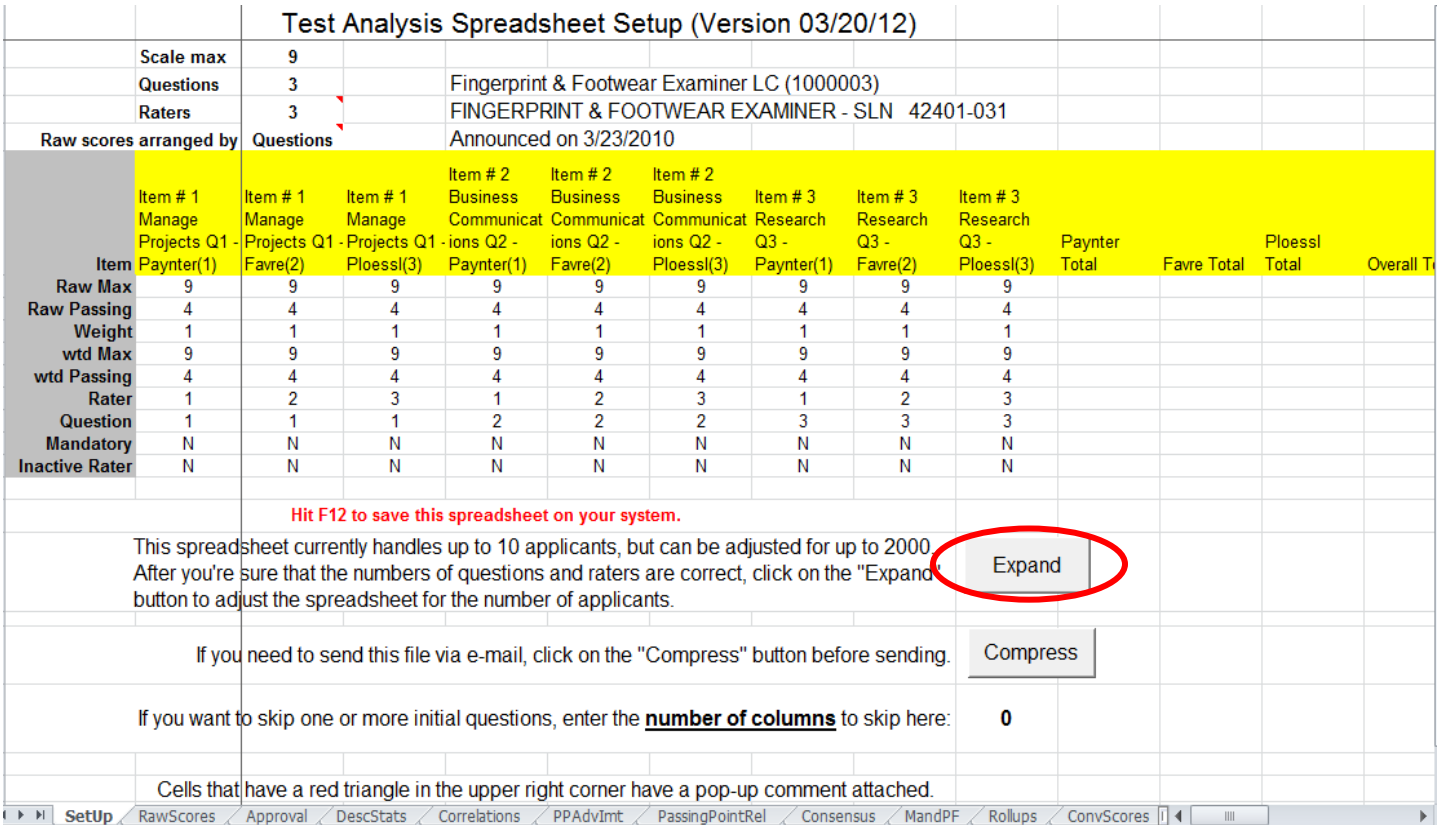
wtd Passing: This is the weighted passing point for each question. It will auto calculate based on what is entered in for the Raw Passing and Weight cells.

Rater: the rater #, it auto populates based on the number of raters entered in the top section.

Question: the question #, it auto populates based on the number of questions entered in the top section.

Initiate the Score Analysis:

Click [Expand], once this is done the screen will flash because the macros are running to perform the calculations.



Test Analysis Spreadsheet Setup (Version 03/20/12)

Scale max	9	
Questions	3	Fingerprint & Footwear Examiner LC (1000003)
Raters	3	FINGERPRINT & FOOTWEAR EXAMINER - SLN 42401-031
Raw scores arranged by	Questions	Announced on 3/23/2010

Item	Item # 1 Manage Projects Q1 - Paynter(1)	Item # 1 Manage Projects Q1 - Favre(2)	Item # 1 Manage Projects Q1 - Ploessl(3)	Item # 2 Business Communications Q2 - Paynter(1)	Item # 2 Business Communications Q2 - Favre(2)	Item # 2 Business Communications Q2 - Ploessl(3)	Item # 3 Research Q3 - Paynter(1)	Item # 3 Research Q3 - Favre(2)	Item # 3 Research Q3 - Ploessl(3)	Paynter Total	Favre Total	Ploessl Total	Overall T
Raw Max	9	9	9	9	9	9	9	9	9				
Raw Passing	4	4	4	4	4	4	4	4	4				
Weight	1	1	1	1	1	1	1	1	1				
wtd Max	9	9	9	9	9	9	9	9	9				
wtd Passing	4	4	4	4	4	4	4	4	4				
Rater	1	2	3	1	2	3	1	2	3				
Question	1	1	1	2	2	2	3	3	3				
Mandatory	N	N	N	N	N	N	N	N	N				
Inactive Rater	N	N	N	N	N	N	N	N	N				

Hit F12 to save this spreadsheet on your system.

This spreadsheet currently handles up to 10 applicants, but can be adjusted for up to 2000. After you're sure that the numbers of questions and raters are correct, click on the "Expand" button to adjust the spreadsheet for the number of applicants. [Expand]

If you need to send this file via e-mail, click on the "Compress" button before sending. [Compress]

If you want to skip one or more initial questions, enter the **number of columns** to skip here: **0**

Cells that have a red triangle in the upper right corner have a pop-up comment attached.

SetUp | RawScores | Approval | DescStats | Correlations | PPAdvImt | PassingPointRel | Consensus | MandPF | Rollups | ConvScores

Analyzing the Results:

Before reviewing the analysis check to make sure the data loaded correctly into the Score Analysis Spreadsheet.

1. Randomly select an applicant on the Rater Export file (sorting the data by CSS to make it easier to compare with The Score Analysis Spreadsheet)
   - Look at their **CSS #** and compare it to the **CS #** on the Score Analysis Spreadsheet.
   - If the #s do not match or if the highest # in the **CS** column is greater than 100, the **[SetUp]** tab information is incorrect or the data was pasted incorrectly into the spreadsheet.
   - Re-verify the # of Raters and Questions for the exam. ***Go to section*** ***Verifying and Adjusting . . .***

2. If the Rater Export file is blank, the exam needs to be sent through the queue to be scored to begin the process. ***Go to section*** ***Score the Exam . . .***

3. If more than one scoring criteria, be sure to choose the correct one from the dropdown box for the classification/position being scored.

Passing Point Reliability [PassingPointRel] tab - *(required documentation for recruitment file)*

Passing Point Reliability Calculations

Values of x indicate missing raters

Coefficient of Consensus = 0.86

Adjust Passing Point Here -> 16

rank	Name (sex, ethnic, handicap)	Rank	CS	Raw	Fail Mand	Raters 1	2	3	4	5	6	7	8	Questions 1	2	3	4	5
1	xxxxxxxx (M5)	1	81.33	23.556		x	24	29						17	23	24	29	
2	xxxxxxxx (M4)	2	69.00	15.333		4	21	21						10	11	12	13	
3	xxxxxxxx (M2)	3	66.00	13.333		14	11	15						10	11	9	10	
4	xxxxxxxx (M5)	4	63.50	11.667		9	13	13						12	11	7	5	
5	xxxxxxxx (F5)	5	60.00	9.333		6	11	11						8	6	6	8	
6	xxxxxxxx (F3)	6	58.00	8.000		4	8	12						6	6	6	6	
7	xxxxxxxx (F1)	6	58.00	8.000		4	8	12						6	6	6	6	
8																		

Coefficient of Consensus (aka the reliability of the passing point)

- This is the rate of agreement on pass/fail decisions across all raters and applicants. In this example, there are 3 raters and 7 applicants.

Looking at the Raters section above:

- Only two of three raters rated applicant #1, because Rater 1 has an X in that column.
- Both rater scores for applicant #1 are at or above the passing point of 16, therefore they "agreed" on this question.
- Rater 1 failed applicant #2, the 4 is highlighted because it is below the passing point of 16.
- Raters 2 & 3 passed applicant #2, so the raters "disagreed" on this candidate.
- All raters "agreed" that applicants 3 thru 7 failed, so there is agreement on 6 out of the 7 applicants. There was only disagreement on one - applicant # 2. The Coefficient of Consensus equals 6/7, or .86.

Evaluating the Coefficient of Consensus #:

- .90+ = very good
- .75 - .90 = OK
- .50 - .75 = barely acceptable

Below .50 indicates severe issues with the rating process; rater orientation needs improvement to gain better clarity on benchmark application or benchmarks need to be more clear.

Determining whether the Passing Point needs to be adjusted or lowered:

1. In the Raters section look for areas where applicants were passed by the majority of raters usually 2 out of 3 raters (white = passing score / teal = failing score).
2. If this happens, the Raw score for an applicant can potentially fall below the passing point noted in the yellow highlighted area (see applicant #2).
3. To adjust the Passing Point, change the yellow highlighted # to the Raw score # of the applicant that was passed by the majority of raters. Ex. Applicant #2 has a Raw score # of 15.333 so the new adjusted passing point would be 15.3 (only use 1 decimal place when adjusting the passing point).
4. With the new adjusted passing point entered the applicant's **CS** should now be 70 or above and the teal areas will have changed in the **CS** column reflecting the new passing point.
5. If the Passing Point is adjusted in the Score Analysis Spreadsheet then it also needs to be adjusted in the system as well.

Passing Point Reliability Calculations

Values of x indicate missing raters

Coefficient of Consensus = 0.71

Adjust Passing Point Here -> 15.3

rank	Name (sex, ethnic, handicap)	Rank	CS	Raw	Fail Mand	Raters 1	2	3	4	5	6	7	8	Questions 1	2	3	4	5
1	xxxxxxxx (M5)	1	80.19	22.333		14	24	29						13	17	18	19	
2	xxxxxxxx (M4)	2	70.05	15.333		4	21	21						10	11	12	13	
3	xxxxxxxx (M2)	3	67.15	13.333		14	11	15						10	11	9	10	
4	xxxxxxxx (M5)	4	64.73	11.667		9	13	13						12	11	7	5	
5	xxxxxxxx (F5)	5	61.35	9.333		6	11	11						8	6	6	8	
6	xxxxxxxx (F3)	6	59.42	8.000		4	8	12						6	6	6	6	
7	xxxxxxxx (F1)	6	59.42	8.000		4	8	12						6	6	6	6	
8																		

Test Reliability for Fingerprint & Footwear Examiner LC (1000003)

Test Title = FINGERPRINT & FOOTWEAR EXAMINER - SLN 42401-031

Corrected r = 0.878 Average r = 0.706 Average z = 0.879 Assigned raters = 3 Active raters = 3

Inter Rater Total Correlations

	rater 1	rater 2	rater 3	rater 4	rater 5	rater 6	rater 7	z transformation
rater 2	0.690							0.849
rater 3	0.737	0.687						0.944 0.843
rater 4								
rater 5								
rater 6								
rater 7								
rater 8								

Interrater-Interquestion Correlations

	Item # 1 Manage Projects Q1 - Paynter(1)	Item # 1 Manage Projects Q1 - Favre(2)	Item # 1 Manage Projects Q1 - Ploessl(3)	Item # 2 Business Communications Q2 -	Item # 2 Business Communications Q2 -	Item # 2 Business Communications Q2 -	Item # 3 Research Q3 - Paynter(1)	Item # 3 Research Q3 - Favre(2)	Item # 3 Research Q3 - Ploessl(3)
Item # 1 Manage Projects Q1 - Favre(2)	0.612								
Item # 1 Manage Projects Q1 - Ploessl(3)	0.725	0.579							
Item # 2 Business Communications Q2 - Paynter(1)	0.941	0.526	0.651						
Item # 2 Business Communications Q2 - Favre(2)	0.348	0.626	0.438	0.425					
Item # 2 Business Communications Q2 - Ploessl(3)	0.455	0.446	0.642	0.450	0.417				
Item # 3 Research Q3 - Paynter(1)	0.822	0.623	0.613	0.826	0.393	0.325			
Item # 3 Research Q3 - Favre(2)	0.696	0.729	0.594	0.691	0.599	0.346	0.769		
Item # 3 Research Q3 - Ploessl(3)	0.726	0.601	0.717	0.661	0.375	0.424	0.724	0.753	

Approval DescStats **Correlations** PPAdvImt PassingPointRel Consensus MandPF Rollups

Corrected r - This number corrects the **Average r** statistic to account for the total number of measurements. This is done because the average r is based only on one rater, since we have 3 raters (3 measurements) this needs to be taken in to account. To do this, the Spearman Brown prophecy formula is used.

- **MUST BE ABOVE .70, - if below, OSER will NOT approve, indicates major problems**

Inter-rater/Inter-question Correlations - This displays the Pearson r correlation between one rater's total scores on a given question against another rater's total scores on another question. ***This is very useful information in determining where error levels may be attributed to in the scoring process (benchmarks, raters, etc.).***

In the example above:

- Ideally the **r** within a question among raters *(#'s highlighted yellow)* is higher on average than the rest of the colors. The practical meaning of this is that items that are meant to be related (scores assigned on the same question) have a higher correlation than items that are not meant to be closely related (scores assigned on different questions).
- The Intra-rater/Inter-question correlations for Rater 1 *(#'s highlighted red)* are very high (.941, .822, .826). This could potentially be attributed to the halo or horns effect.

This spreadsheet utilizes correlations to calculate reliability, rather than R-Alpha. A major point about correlations is that they depend upon independent ratings. If ratings are not independent, this info is meaningless. Another difference is that R-Alpha is a lower bound estimate (meaning the R-Alpha tells us the reliability is no lower than that estimate). However, when using R-Alpha, one cannot attribute low levels of reliability to a specific question, but Pearson r correlations helps to do this.

Average r - this is the average correlation among raters. To get to this number, a z score for each rater is calculated by transforming each of the Pearson r correlations (above these are .690, .737, .687) to a normally distributed z value. These z values are then averaged to get the average z (.879), and the Fisher inverse is taken to get the Average Pearson r correlation (.706).

Inter-Rater Total Correlations - This is the Pearson r correlation when comparing the total rater scores assigned for each applicant by one rater against the total rater scores assigned for each applicant by another rater. In this example, the correlation between Rater 1 and Rater 2 is .690, the correlation between Rater 1 and Rater 3 is .737, and the correlation between Rater 2 and Rater 3 is .687. Therefore, the strongest correlation in this example is between Rater 1 and Rater 3, and the weakest correlation is between Rater 2 and Rater 3.

Title

Test Title: Exam Total

Announced on 12/14/2009

Weighted Rater-Question Statistics						Weighted Rater Statistics						Weighted Question Statistics					
Item	N	Mean	SD	Min	Max	Rater	N	Mean	SD	Min	Max	Quest	N	Mean	SD	Min	Max
Item # 1 Question Summary -(1)	6	2.33	1.75	1	5	1	6	6.83	4.02	4	14	1	7	9.79	3.70	6	17
Item # 1 Question Summary -(2)	7	2.71	0.95	2	4	2	7	13.71	6.32	8	24	2	7	10.50	5.85	6	23
Item # 1 Question Summary - (3)	7	4.29	2.21	3	9	3	7	16.14	6.59	11	29	3	7	10.00	6.56	6	24
Item # 2 Item Summary - (1)	6	1.67	1.03	1	3	4	0					4	7	10.93	8.23	5	29
Item # 2 Item Summary - (2)	7	3.71	1.89	2	7	5	0					5	0				
Item # 2 Item Summary - (3)	7	4.29	1.8	3	8	6	0					6	0				
Item # 3 Question Summary - (1)	6	1.33	0.82	1	3	7	0					7	0				
Item # 3 Question Summary - (2)	7	3.86	2.67	2	9	8	0					8	0				
Item # 3 Question Summary - (3)	7	3.86	1.68	2	7							9	0				
Item # 4 Item Summary - (1)	7	2.43	2.57	1	8							10	0				
Item # 4 Item Summary - (2)	7	3.43	1.81	2	6							11	0				
Item # 4 Item Summary - (3)	7	3.71	1.38	2	6							12	0				

This tab will point out if one rater is always rating high or low.

Under the **Weighted Rater-Question Statistics**

- Look at **Min/Max** columns, there should be variation.
- If there were all 6's for the Max, when the Max could be 9, look at the benchmarks to see if "ANDs" were used a lot. Example: The applicant needs ____ experience AND _____ experience.

Under the **Weighted Rater Statistics**

- Look at the **Mean** to make sure these numbers are fairly close.

N = Number of applicants that received a score for a particular question or rater

Mean = Average Score

SD = Standard Deviation for the particular measurement

NOTE: *All sections below are weighted based on information entered into the* ***SetUp*** *tab*

Weighted Rater-Question Statistics - This section lists the data broken out by rater and question. For example, if we look at the first row "Item #1 Question Summary-(1)", we can see that Rater (1) scored 6 candidates on Question 1 (because N=6). The average of the 6 scores was 2.33, and the standard deviation was 1.75. The minimum score that rater (1) assigned on Question 1 was 1 (Min=1) and the maximum score they assigned was 5 (Max=5).

Weighted Rater Statistics - This section sums up the scores for each Rater across all questions. We can see from this example that Rater 1 scored 6 candidates (N=6), the average of their scores was 6.83, and their standard deviation was 4.02. The lowest total score they assigned for a candidate was 4 (Min=4), and the highest total score they assigned was 14 (Max=14).

Weighted Question Statistics - This section sums up the data for each question on the exam. In this example, we can see that overall, 7 applicants were scored on question 1 (N=7), the average score was 9.79, the lowest score on Question 1 when combining all raters scores was 6, and the highest score on Question 1 when combining all raters was 17.

Adverse Impact [PPAdvImt] tab - *(required documentation for recruitment file)*

Adverse Impact Analysis for
Title

Passing Point Impact Analysis for Minorities					Score Stats	
Ethnic Code	Pass	%	Fail	%	Average	SD
1	0	0.00%	1	100.00%	8.000	
2	0	0.00%	1	100.00%	13.333	
3	0	0.00%	1	100.00%	8.000	
4	0	0.00%	1	100.00%	15.333	
5	1	33.33%	2	66.67%	14.852	7.627
9	0		0			
Total	1	14.29%	6	85.71%	12.746	5.5037
All Minorities	0	0.00%	4	100.00%	12.746	5.504
80% Rule	26.67%					
Probability that the difference between Minority & White (5) score averages is due to chance = 0.70						

Passing Point Impact Analysis for Females					Score Stats	
Gender	Pass	%	Fail	%	Average	SD
M	1	25.00%	3	75.00%	15.972	5.273
F	0	0.00%	3	100.00%	8.444	0.770
U	0		0			
Total	1	14.29%	6	85.71%	12.746	5.5037
80% Rule	20.00%					
Probability that the difference between Male & Female score averages is due to chance = 0.06						

Note: The passing point is <u>never</u> adjusted based on the Adverse Impact Analysis due to the generally small sample size of applicants for most recruitments. The analysis can assist with the development of future diversity recruitment goals.

This tab breaks down the number of applicants who pass or fail in each Race/Ethnic category, as well as each Gender category.

Race/Ethnicity Categories:
1 - Black (not Hispanic)
2 - Asian or Pacific Islander
3 - American Indian/Alaska Native
4 - Hispanic
5 - White (not Hispanic)
9 - Unknown

The **All Minorities** row sums up all the different minority categories, and then displays the percentage of applicants that Pass and the percentage that Fail.

The **80% Rule** (4/5 Rule) box is displayed a little differently. This actually displays the percentage of the minority group that would need to pass in order for the passing rate to be at least 80% of the majority group. From the ethnicity example above, the passing rate of White applicants is 1/3 or 33.33%. Therefore, the 80% rule is .3333 x .8, or 26.67%.

The **"Probability that the difference between . . . is due to chance"** info is the significance test. . If the # is below .05 there is statistical significance of adverse impact. *Note: This text only appears when there are at least two scored applicants in the minority group (ethnicity or gender).*

In the **Passing Point Analysis for Females** section, the 3 gender options are M-Male, F-Female, and U-Unknown. In the example, the passing rate for females needs to be 20% in order to be at least 80% of the male passing rate (25%). Since it is less than 20%, "<u>Adverse Impact for Females</u>" will display on our approval tab.

Note: This tab is <u>only</u> functional when using MS Office 2007 or newer

Title									
Test Title: Exam Total									
Announced on 12/14/2009									
Rater Category Consensus by Question									
Question	1	2	3	4	5	6	7	8	Average
Passing Point Consensus	0.57	0.43	0.57	0.71					0.57
Overall Category Consensus	0.57	0.43	0.57	0.57					0.54

This tab shows further analysis of the examination results including:

Passing Point Consensus – for each question and the average as opposed to the overall exam passing point consensus which was reviewed previously on the Passing Point Reliability tab (also known as the coefficient of consensus)

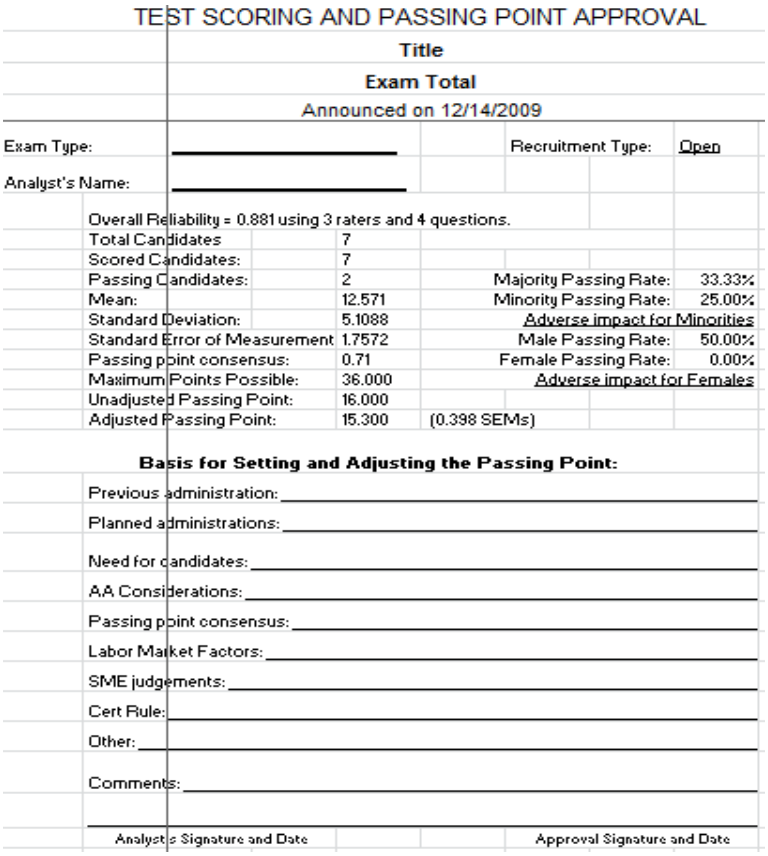
Overall Category Consensus – for each exam benchmark category and the average (i.e. More Than Acceptable, Acceptable, Less Than Acceptable).

In the example above:

- Passing Point Consensus – Q4 has the best rate of .71 (71%) when compared to the other questions. Q2 has a low rate of .43 which means only 43% of the time the raters were in agreement on whether an applicant receives a passing score on this question. This helps to identify which questions caused the greatest rater disagreement.

- Overall Category Consensus - Q1, Q3 and Q4 all have a rate of .57 (57%) which is low. This means only 57% of the time the raters agreed on the benchmark category to score applicant responses. The ideal result would be that this rate (%) were high because this would mean the raters at a minimum agreed on whether an applicant response was considered in the More Than Acceptable, Acceptable or Less Than Acceptable benchmark category. If the benchmarks are very clear and easily applied this rate will be higher.

Approval tab - *(required documentation for recruitment file)*



TEST SCORING AND PASSING POINT APPROVAL

Title

Exam Total

Announced on 12/14/2009

Exam Type: _____ Recruitment Type: Open

Analyst's Name: _____

Overall Reliability = 0.881 using 3 raters and 4 questions.

Total Candidates	7		
Scored Candidates:	7		
Passing Candidates:	2	Majority Passing Rate:	33.33%
Mean:	12.571	Minority Passing Rate:	25.00%
Standard Deviation:	5.1088	Adverse impact for Minorities	
Standard Error of Measurement	1.7572	Male Passing Rate:	50.00%
Passing point consensus:	0.71	Female Passing Rate:	0.00%
Maximum Points Possible:	36.000	Adverse impact for Females	
Unadjusted Passing Point:	16.000		
Adjusted Passing Point:	15.300	(0.398 SEMs)	

Basis for Setting and Adjusting the Passing Point:

Previous administration: _____

Planned administrations: _____

Need for candidates: _____

AA Considerations: _____

Passing point consensus: _____

Labor Market Factors: _____

SME judgements: _____

Cert Rule: _____

Other: _____

Comments: _____

Analyst's Signature and Date

Approval Signature and Date

Total Candidates - The total number of applicants from the rater score export file (applicants that have applied – not necessarily the same # that have completed the exam)

Scored Candidates - Only counts applicants with scores entered (does not count applicants that do not have scores entered because they had blank responses to the exam questions).

Adverse impact for Minorities (or Females) – will only appear when detected

At a minimum fill in these lines:

Exam Type: *T&E (rated exam)*

Analyst Name: *the Agency HR Staffing Analyst*

Passing point consensus: *example = .71*

Cert Rule: *transfer from Exam Plan Checklist*

Use the blanks of the Approval page to document factors relating to adjusting the passing point:

Previous administration: If the same test was given in the past year or so, consider if you should be using the same passing point so that the current group of applicants is evaluated consistently with the previous one. This is particularly important if some of the applicants applied for both recruitments and used the same answers.

Planned administrations: If this test will be given again in the near future and the current group of applicants is too limited to make a fully informed passing point decision, you may want to make an adjustment.

Need for candidates & Labor Market Factors: If labor market conditions are such that the level of qualification within your applicant pool is too small for the number of vacancies and the risk associated with lowering the passing point is minimal, you may want to make a downward adjustment to pass more applicants.

AA Considerations: If the job is underutilized, the test has adverse impact and the risk associated with lowering the passing point is minimal, you may want to make a downward adjustment to pass more protected group applicants.
****** This option should not be the sole basis for lowering the passing point ******

Passing point consensus: Often you will find some disagreement among your raters as to which candidates should pass. Consider adjusting the passing point downwards to the score of the applicant that has a majority of the raters giving a passing score.

SME judgments: If your raters indicate that the benchmarks were unrealistically strict, you may want to adjust the passing point downward. This can be done with your SME's approval either for one or more questions on the Setup tab or globally on the Passing Point Reliability tab.

Comments: Note passing point decisions and any other adjustments made to the recruitment/scoring.

Examples:
- "No need to adjust passing point"
- "Adjusted passing point due to 2 out of 3 rater agreement"
- "Removed mandatory due to inconsistent ratings"

Required Documentation For The Recruitment File:

Non-Delegated Agencies - contact DPM to perform and approve the score analysis then file the documents below

Delegated Agencies – follow the instructions to perform the score analysis then . . .

1. **Print these tabs of the Score Analysis Spreadsheet**
    - **Approval** tab - can use quick print, if delegated follow internal procedure to approve score analysis and use the signature lines at the bottom
    - **PassingPointRel** tab – define the print area including all rows/columns before printing this tab
    - **PPAdvImt** tab – can use quick print
2. **File all scoring related documents in the recruitment folder**
    - Include the completed Exam Security Agreement/Job Expert Certificates

OSER Procedures Performed (applies to Non-Delegated Agencies only):

1. Generate the score analysis
2. Review analysis to determine whether the exam results are acceptable
    - Passing Point Consensus = .5+ (prefer .75+)
    - Pinpoint any issues with raters, questions and/or benchmarks
    - Adjust/lower passing point because of 2 out of 3 rater agreement
4. Print the tabs in step 1 above
    - Sign Approval page on the lower left signature line
3. Have a BMRS Senior/Advanced Analyst review analysis findings
    - Consult with the BMRS Administrator as needed
        - Passing Point Consensus below .5
        - Rater inconsistencies
        - Use of different raters
4. Analysts determine appropriate action
    - Approve
    - Adjust/lower Passing Point
        - Adjust exam passing point in the system
        - Re-score the exam
    - Other
5. Obtain Approval Signature from Senior/Advanced Analyst
6. Scan and email the score analysis documents to the Agency HR Staffing Analyst (retain in recruitment folder)
    - Approval tab
    - PassingPointRel tab
    - PPAdvImt tab
7. Review recruitment folder contents and arrange according to folder checklist
    - Print anything that might be missing (e.g. ERS announcement, final job anno or exam, etc.)
    - Complete the Recruitment Checklist
    - Check off the folder checklist
    - Place the recruitment folder into the basket in the top file cabinet located at the end (closest to the main aisle) of the recruitment folder file cabinets by classification
    - ESC team will label the folder, noting retention dates and will file it back with current recruitment folder records