BCIRP Survey Analytics Workshop

Presented by:
Stephanie McKeown & Scott Emerson
UBC's Okanagan Campus (OPAIR)

Thursday, June 12

Overview

I. **What are composite scores?**

II. **Why would we want to use them?**

III. **Step-by-step demonstration: creating a composite**

IV. **Interpreting composite scores**

V. **Multilevel Validity: testing levels of aggregation**

What are composite scores?

- A composite is a set of related survey questions(items)
- Items are carefully selected, and together represent an underlying trait/ ability/ construct
- Multiple items, not a single item
- One score (an index) that represents the composite

Why use composite scores?

- A common approach to analyze survey data is to report item results individually
  - frequency distribution
  - % positive/ % negative
  - an average rating per group
- However, this can be a problematic approach for several reasons…

Defining Constructs

- Many of the concepts surveys measure are complex social phenomena, such as attitudes, perceptions, belonging (Furr & Bacharach, 2013)

- Such constructs are considered "latent" variables as compared to "manifest" variables and, hence, cannot be measured directly
  - Use multiple "indicators" or items to tap into the construct

- Using a single item can be problematic as it limits the opportunities an individual has to respond to the question and provide valid and reliable information regarding the latent variable

Measuring Constructs

- Using a single item to measure a construct can be problematic as it limits the opportunities an individual has to respond to the question and provide valid and reliable information regarding the latent variable
  - Can result in under-representation of the construct
  - One item may not aptly reflect diversity (variability) of respondents and may not sufficiently differentiate individuals

Benefits of composites

- Using a multi-item composite yields several benefits:
  - Greater validity as it more accurately reflects the complexity and multi-dimensional nature of a construct than one item would
  - Better captures variability in responses/ respondents

  → Multiple items provide a broader catchment of information

  - Allows weighting of items in accordance to their relative contributions to the composite – the composite score is thus better reflection of the collection of items

Psychometrics: Reliability

- Important psychometric aspect of surveys

- Can be understood as consistency

  → "whether an instrument can be interpreted consistently across different situations" (Field, 2009)

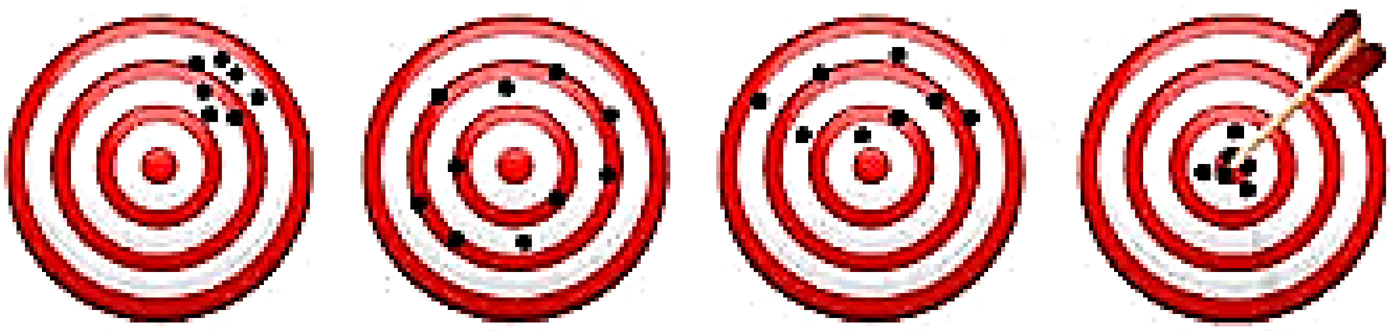
- If an individual responds to different items regarding the same construct, will the same information be yielded? Or will the responses vary? …e.g., Cronbach's Alpha

Psychometrics: Validity

- Reliability is referred to as consistency and validity is referred to as accuracy.
- Extent to which a set of items measure what they are purported to measure and we can interpret the results in a meaningful way.
- Both reliability and validity are not static attributes of an item/ set of items but need to be explored for each administration

  → Better understood as the interpretations or conclusions drawn from the survey scores

Reliability and validity



Not valid
Reliable

Low validity
Low reliability

Not valid
Not reliable

Valid and
Reliable

Image from: https://explorable.com/validity-and-reliability

Validation Process

- Modern validity theory is focused on the interpretation or conclusions drawn from the survey results rather than the actual survey responses (Borden & Young, 2008).
- The validation process involves first specifying the proposed inferences and assumptions that will be made from the findings and then provide empirical evidence to support them (Kane, 2006; Zumbo, 2007, 2009).
- The amount of evidence required depends on the importance of the interpretation or survey use.

Validity – dimensionality

- Dimensions are themes, clusters of associated items
- Sets of items may be uni-dimensional or have several dimensions, multi-dimensional (Furr & Bacharach, 2013)
- If the results suggest more than one theme, it may be more appropriate to create multiple composites

Creating composites

- A composite is a set/ cluster of related survey items that together aim at capturing a certain theme (e.g., sense of belonging on campus)

- It's essential that one is careful when selecting the items that make up a composite. Three important considerations:

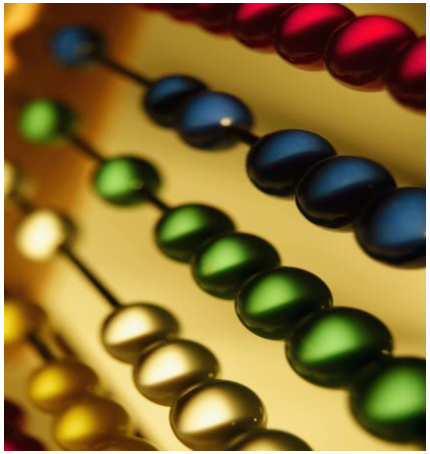
  I) Theoretical/ Practical
  II) Psychometric evaluation
  III) Reasoning/ Appropriateness

Data preparation

- Before creating a composite, it's important to prepare the survey data for further analysis:
  - Check the data to account for missing values (different approaches to managing missing data)
  - Determine if sample size (for items of interest) is sufficient
  - Ensure survey items are on the same scale (5-point, 4-point etc.), or transform to a common scale
  - Check for reverse scored items, ensure all are scored in same direction (e.g., higher numbers = higher level of construct)

Creating Composite Scores

- **Non-refined Methods:**
  - Sum scores by factor*
  - Sum scores – above a cut-off value
  - Sum scores – standardized variables
  - Weighted sum scores*

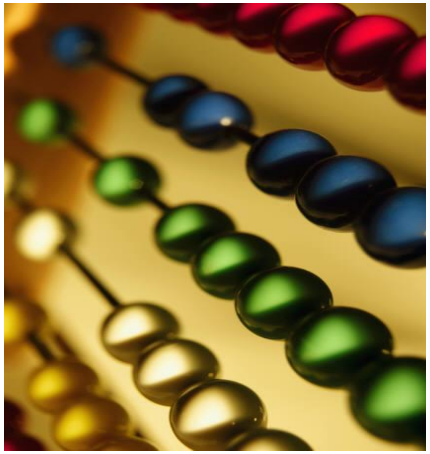


- **Refined Methods:**
  - Regression Scores*
  - Bartlett Scores
  - Anderson-Rubin Scores

(DiStefano, Zhu, & Mîndrilă, 2009)

Method 1: Sum Scores by Factor

- Summing all items together, items assumed to have equal contributions



- Summing up all items together as one composite and deriving an overall score
  - Example: I select 5 items on a 4-point scale about satisfaction with teaching
- Need to be on a common scale or converted to one
- Can either report a totalled score or averaged score
  - For above example, could average to be /4 or total out of /20

Method 2: Weighted Sum Scores

- Factor analysis has many uses, one of which is to better explore and understand the data (e.g., reduce many items into smaller, themes/ clusters)
- Factor analysis output will yield factor (component) scores
- Factor scores should be weighted according to their contribution to the composite
  - Note: It is possible that all items in a set may contribute very similarly to the construct, and therefore weighting may not be needed

Method 3: Regression Scores

- Regression factor scores predict the location of each individual on the factor or component
- Method 3 uses an underlying model to predict an “optimal” factor score
- IVs in the regression equation are the standardized values of the items in the estimated factors, The factor scores are the DVs
- A multivariate procedure – takes into account the correlation between the factors, between factors and observed variables (via factor loadings), the correlation among observed variables, and also the correlation among oblique factors (for oblique EFA solutions).

LET’S CREATE COMPOSITE SCORES!

Programme for International Student Assessment (PISA)

- Worldwide study by the Organisation for Economic Co-operation and Development (OECD)
- First implemented in 2000 and every three years afterwards; latest was in 2012
- 15-year-old students across 65 nations and territories
- More than 510,000 students participated
- Randomly selected schools
- Translated into many different languages
- Free-access data to analyze
- Purpose of results are to improve education policies and outcomes

What does PISA measure?

- Mixture of open-ended and multiple-choice questions that are organized in groups based on a passage setting out of a real-life situation
  - 2 hour examination
  - A total of about 390 minutes of test items are covered
  - Students take different combination of different tests
- Students’ scholastic performance on mathematics, science and reading, with a focus on one subject in each year of assessment
- Students and school principals answer questionnaires to provide information about the students’ backgrounds, schools and learning experiences and about the broader school system and learning environment

Sample to use here…

- 130 schools in Canada
- 1075 students
- Minimum number of students is 5 per school
- Wrote the test in English, but speak French at home

PISA 2006 dataset

- **School-level variables**
  - School size
  - School mean SES (socio-economic status)
  - School principals' perceptions of shortages acting as hindrances

- **Student-level variables**
  - Gender
  - PISA scores – 5 plausible values each for science and reading
  - SES – PISA index of economic, social, and cultural status

Step-by-step demonstration

- Composite will be created in SPSS, using a sample data set…

Step 1: Prepare data (missing data, size, response scales)

Step 2: Select items for composite

Step 3: Evaluate psychometric properties (reliability, validity)

Step 4: Decide on final items for inclusion in composite

Step 5: Calculate relative weightings for each item

Step 6: Apply weights to each item in the composite

Step 7: Derive score for the composite

Step 1: Prepare data

- Items of interest may have different scales
(e.g., 1 - 4, 1 - 5, 1 - 6…)

- All items included in the composite need to be scored according to the same scale, or else scores can be exaggerated!
(e.g., 3 on a 4-point scale vs. 3 on a 6-point scale)

- Using a formula, one can convert the scores in terms of a different response category: Convert 6-point scale to 5-point

6 point	1	2	3	4	5	6
5 point	1	1.8	2.6	3.4	4.2	5

Conversion	Formulae (A=original value, B=converted value)
1 - 6 → 1 - 5	B = (A*4/5) + 1/5
1 - 5 → 1 - 6	B = (A-3)*(5/4) + 3.5
1 - 4 → 1 - 6	B = (A*5/3) – 2/3

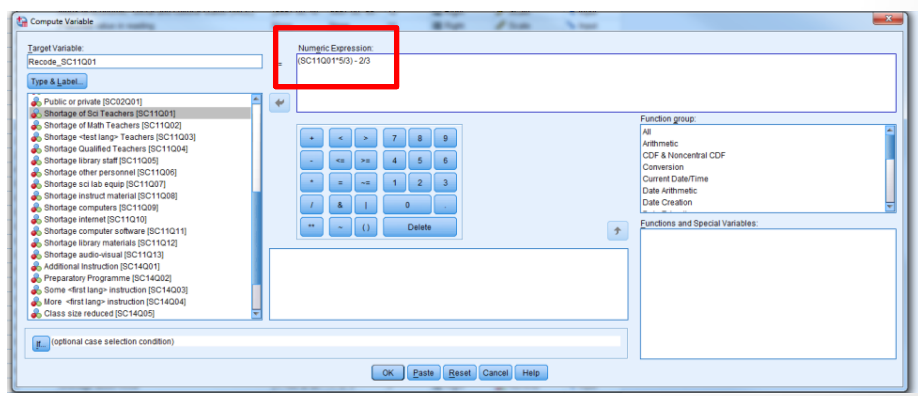
Conversion	Original Score (A)					
	1	**2**	**3**	**4**	**5**	**6**
6 → 5	1.00	1.80	2.60	3.40	4.20	5.00
5 → 6	1.00	2.25	3.50	4.75	6.00	
4 → 6	1.00	2.67	4.33	6.00		

Converting response scales

- Transform → Compute Variable

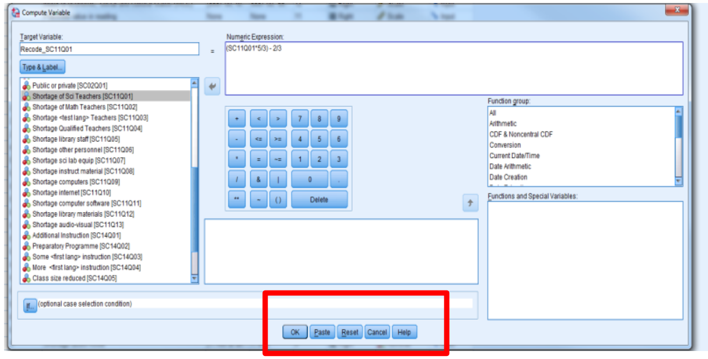
- Name Target Variable (e.g., “Q1_6scale) and select variable to be converted by double-clicking (Q1_rescale4point)

- Enter conversion formula for variable: Q1_rescale4point*5/3 – 2/3



Syntax

- Select **paste** at bottom
- **Syntax Editor** window opens, with the command syntax written out
- Can be saved (File → Save)
  - Saves time
  - Allows for review
- Click the green play button to run the calculation



- Once the variable has been converted, it will appear in the data set at the bottom (may wish to label variable too)

- Analyze → Descriptive statistics
  → frequencies (select ok)

- Response categories have been converted from 1 – 4 to 1 – 6

- If using variables with converted response scales, remember to use those in the analysis … **not the original variables**!

Check to see conversion

Shortage of Sci Teachers (4 point)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	824	76.7	76.7	76.7
	2	97	9.0	9.0	85.7
	3	129	12.0	12.0	97.7
	4	25	2.3	2.3	100.0
	Total	1075	100.0	100.0	

Shortage of Sci Teachers (6 point)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	824	76.7	76.7	76.7
	2.67	97	9.0	9.0	85.7
	4.33	129	12.0	12.0	97.7
	6.00	25	2.3	2.3	100.0
	Total	1075	100.0	100.0	

Re-score direction of responses

- Reverse items may be problematic:
  - Intuitively, we tend to consider higher numbers as higher level of a construct. For instance:
  - ‘4’ on a 5-point satisfaction scale as greater satisfaction than ‘2’
  - May be the opposite! Lower scores = greater

- Some scales are ‘reverse scored’
- Scores need to be all in one direction,
  - May need to convert scores so higher scores = higher level of construct (positively scored)

- **Transform → Recode into Different Variables**

- Select Q1_reverseCoding (double-click)
  - Name and label variable as Q1_revised

- Select **Old and New Values…**

- **Old value**: Value (enter current value)
- **New Value**: Value (enter desired value)

- **Continue → Change → OK/Paste**

	Not at all	Very little	To some extent	A lot
Old (rev)	4	3	2	1
New	1	2	3	4

Missing data

- Important to check for missing data
- Analyze → Descriptive Statistics → Frequencies

Statistics

		Shortage of Sci Teachers	Shortage of Math Teachers	Shortage <test lang> Teachers	Shortage Qualified Teachers	Shortage library staff	Shortage other personnel	Shortage sci lab equip
N	Valid	1075	1075	1075	1075	1075	1075	1075
	Missing	0	0	0	0	0	0	0

Statistics

		Shortage instruct material	Shortage computers	Shortage internet	Shortage computer software	Shortage library materials	Shortage audio-visual
N	Valid	1075	1075	1075	1075	1075	1075
	Missing	0	0	0	0	0	0

- No missing data for these variables, which is good
- If considerable amount of data missing, there are several procedures to take (see

Step 2: Select items for composite

- Which items should we choose for the composite from those available?
- A theme?
- Consider theory and purpose behind creating composite

Step 3: Evaluate psychometric properties

Validity

- Factor analysis will provide factor loadings, contribution of each item to the composite

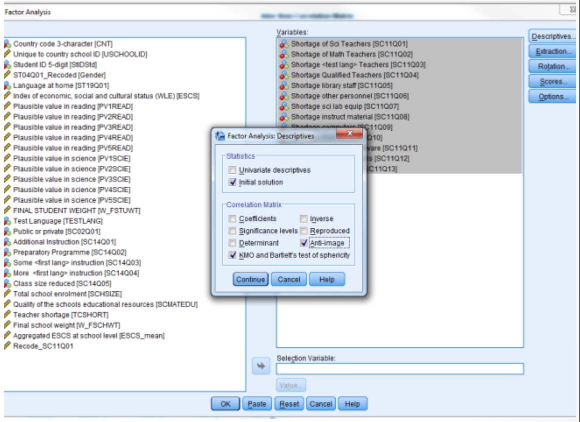
Reliability

- One method is Cronbach's alpha, which estimates:
  - Reliability for individual items
  - Reliability for overall composite (inter-item reliability)
  - Reliability if a given item was removed
    - Can be useful as it can highlight a possibly weak (unreliable) item

Factor analysis

- Analyze → Dimension Reduction → Factor

- Variables = SC11Q01 – SC11Q13
- **Descriptives**
  - Univariate descriptives
  - Anti-image
  - KMO and Bartlett's test of sphericity



- **Extraction**
  - Method: Principal components
  - Unrotated factor solution
  - Based on Eigenvalue greater than 1

- **Options**
  - Suppress small coefficients, absolute value below .32

- Click paste and save syntax (then run)

Interpreting factor analysis

- Factor output will provide estimates of how much variability in the data is captured by the items

- Will suggest which items are contributing in a meaningful way to the composite concept (loadings)

- Will highlight the dimensionality of the data, suggest themes/ clusters within the items

    - Comrey and Lee (1992) suggest loadings > 0.71 are excellent, 0.63 very good, 0.55 good, 0.45 fair, and 0.32 poor

    - Useful to re-consider any items with loadings around < 0.40 as they may not belong with the other items in the set

FA Output

- **Communalities**
- **Total Variance Explained**

Communalities

	Initial	Extraction
Shortage of Sci Teachers	1.000	.757
Shortage of Math Teachers	1.000	.748
Shortage <test lang> Teachers	1.000	.588
Shortage Qualified Teachers	1.000	.474
Shortage library staff	1.000	.362
Shortage other personnel	1.000	.526
Shortage sci lab equip	1.000	.485
Shortage instruct material	1.000	.505
Shortage computers	1.000	.665
Shortage internet	1.000	.608
Shortage computer software	1.000	.749
Shortage library materials	1.000	.670
Shortage audio-visual	1.000	.763

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.537	42.596	42.596	5.537	42.596	42.596	4.529	34.840	34.840
2	2.364	18.188	60.784	2.364	18.188	60.784	3.373	25.944	60.784
3	.930	7.157	67.940						
4	.799	6.143	74.083						
5	.619	4.762	78.846						
6	.558	4.291	83.137						
7	.476	3.659	86.796						
8	.425	3.267	90.062						
9	.372	2.862	92.924						
10	.354	2.726	95.650						
11	.236	1.812	97.462						
12	.171	1.317	98.779						
13	.159	1.221	100.000						

Extraction Method: Principal Component Analysis.

- Univariate descriptives

Overview of items means and SDs (spread) lower Means

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
Shortage of Sci Teachers	1.40	.787	1075
Shortage of Math Teachers	1.44	.852	1075
Shortage <test lang> Teachers	1.17	.489	1075
Shortage Qualified Teachers	1.63	.797	1075
Shortage library staff	1.39	.744	1075
Shortage other personnel	1.81	.913	1075
Shortage sci lab equip	1.58	.768	1075
Shortage instruct material	1.92	.972	1075
Shortage computers	1.89	.938	1075
Shortage internet	1.60	.846	1075
Shortage computer software	1.77	.834	1075
Shortage library materials	1.66	.746	1075
Shortage audio-visual	1.78	.849	1075

- KMO
- Sphericity

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.846
Bartlett's Test of Sphericity	Approx. Chi-Square	7978.888
	df	78
	Sig.	.000

Kaiser-Meyer-Olkin Measure of Sampling Adequacy (Kaiser, 1970)

- KMO measure of sampling adequacy Represents the ratio of the squared correlation between variables to the squared partial correlation between variables. varies between 0 - 1

- Value of 0 indicates diffusion in the pattern of correlations (hence, factor analysis is likely to be inappropriate)

- Value closer to 1 suggests patterns of correlations are relatively compact and so factor analysis should yield distinct and reliable factors

- Mediocre: 0.5 - 0.7, Good: 0.7 - 0.8
- Great: 0.8 - 0.9, Superb: ≥ 0.9 (Hutcheson & Sofroniou, 1999)

Bartlett's test of sphericity

- Tests null hypothesis that original correlation matrix is an identity matrix
- For factor analysis to work, need some relationships between variables
- Want this to be significant (< .05), which indicates there are some relationships between the variables we hope to include in the analysis
- Since Bartlett's test is sig (< .001), factor analysis is appropriate

(Field, 2009)

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.846
Bartlett's Test of Sphericity	Approx. Chi-Square	7978.888
	df	78
	Sig.	.000

- Anti-image matrices
- Can be useful for identifying possible items for deletion
- The anti-image matrix of covariances and correlations. These matrices contain measures of sampling adequacy for each variable along the diagonal and the negatives of the partial correlation/covariances on the off-diagonals (Field, 2009)
- The diagonal elements should be > .05…if the sample is adequate for a given set of variables (Field, 2009, p.651)

Look at correlations, rather than covariances…more informative (Field, 2009)

Anti-image Matrices

		Shortage of Sci Teachers	Shortage of Math Teachers	Shortage <test lang> Teachers	Shortage Qualified Teachers	Shortage library staff	Shortage other personnel	Shortage sci lab equip	Shortage instruct material	Shortage computers	Shortage internet	Shortage computer software	Shortage library materials	Shortage audio-visual
Anti-image Covariance	Shortage of Sci Teachers	.328	-.173	-.100	-.018	-.173	.073	.053	-.106	-.059	-.020	.061	-.027	.018
	Shortage of Math Teachers	-.173	.378	-.123	-.065	-.006	.005	-.052	-.014	.035	.022	-.001	.050	-.054
	Shortage <test lang> Teachers	-.100	-.123	.528	-.047	.045	.002	-.004	-.044	.085	.049	-.085	-.006	-.001
	Shortage Qualified Teachers	-.018	-.065	-.047	.580	.042	-.147	-.066	-.085	.017	.036	-.021	-.051	.017
	Shortage library staff	-.173	-.006	.045	.042	.532	-.193	-.052	.113	.026	.014	-.023	-.068	.019
	Shortage other personnel	.073	.005	.002	-.147	-.193	.473	-.029	-.042	-.066	-.055	.033	.007	-.069
	Shortage sci lab equip	.053	-.052	-.004	-.066	-.052	-.029	.564	-.144	-.055	-.073	.022	-.064	.007
	Shortage instruct material	-.106	-.014	-.044	-.085	.113	-.042	-.144	.530	-.011	-.031	.003	-.028	-.013
	Shortage computers	-.059	.035	.085	.017	.026	-.066	-.055	-.011	.431	-.044	-.116	-.031	-.023
	Shortage internet	-.020	.022	.049	.036	.014	-.055	-.073	-.031	-.044	.428	-.158	.085	-.051
	Shortage computer software	.061	-.001	-.085	-.021	-.023	.033	.022	.003	-.116	-.158	.270	-.048	-.064
	Shortage library materials	-.027	.050	-.006	-.051	-.068	.007	-.064	-.028	-.031	.085	-.048	.300	-.147
	Shortage audio-visual	.018	-.054	-.001	.017	.019	-.069	.007	-.013	-.023	-.051	-.064	-.147	.248
Anti-image Correlation	Shortage of Sci Teachers	.707[a]	-.490	-.240	-.041	-.415	.185	.122	-.255	-.156	-.053	.204	-.085	.064
	Shortage of Math Teachers	-.490	.802[a]	-.276	-.139	-.014	.013	-.113	-.031	.088	.055	-.004	.148	-.176
	Shortage <test lang> Teachers	-.240	-.276	.840[a]	-.085	.085	.005	-.008	-.083	.178	.104	-.226	-.016	-.002
	Shortage Qualified Teachers	-.041	-.139	-.085	.909[a]	.075	-.281	-.116	-.154	.035	.072	-.053	-.122	.045
	Shortage library staff	-.415	-.014	.085	.075	.762[a]	-.384	-.095	.213	.055	.029	-.060	-.172	.051
	Shortage other personnel	.185	.013	.005	-.281	-.384	.854[a]	-.056	-.083	-.147	-.122	.093	.020	-.200
	Shortage sci lab equip	.122	-.113	-.008	-.116	-.095	-.056	.915[a]	-.263	-.111	-.148	.055	-.155	.019
	Shortage instruct material	-.255	-.031	-.083	-.154	.213	-.083	-.263	.892[a]	-.023	-.066	.008	-.071	-.036
	Shortage computers	-.156	.088	.178	.035	.055	-.147	-.111	-.023	.902[a]	-.103	-.338	-.085	-.069
	Shortage internet	-.053	.055	.104	.072	.029	-.122	-.148	-.066	-.103	.830[a]	-.463	.238	-.157
	Shortage computer software	.204	-.004	-.226	-.053	-.060	.093	.055	.008	-.338	-.463	.833[a]	-.169	-.246
	Shortage library materials	-.085	.148	-.016	-.122	-.172	.020	-.155	-.071	-.085	.238	-.169	.852[a]	-.539
	Shortage audio-visual	.064	-.176	-.002	.045	.051	-.200	.019	-.036	-.069	-.157	-.246	-.539	.870[a]

a. Measures of Sampling Adequacy(MSA)

Anti-Image Matrices (correlations)

Anti-image Matrices

		Shortage of Sci Teachers	Shortage of Math Teachers	Shortage <test lang> Teachers	Shortage Qualified Teachers	Shortage library staff	Shortage other personnel	Shortage sci lab equip	Shortage instruct material	Shortage computers	Shortage internet	Shortage computer software	Shortage library materials	Shortage audio-visual
Anti-image Covariance	Shortage of Sci Teachers	.328	-.173	-.100	-.018	-.173	.073	.053	-.106	-.059	-.020	.061	-.027	.018
	Shortage of Math Teachers	-.173	.378	-.123	-.065	-.006	.005	-.052	-.014	.035	.022	-.001	.050	-.054
	Shortage <test lang> Teachers	-.100	-.123	.528	-.047	.045	.002	-.004	-.044	.085	.049	-.085	-.006	-.001
	Shortage Qualified Teachers	-.018	-.065	-.047	.580	.042	-.147	-.066	-.085	.017	.036	-.021	-.051	.017
	Shortage library staff	-.173	-.006	.045	.042	.532	-.193	-.052	.113	.026	.014	-.023	-.068	.019
	Shortage other personnel	.073	.005	.002	-.147	-.193	.473	-.029	-.042	-.066	-.055	.033	.007	-.069
	Shortage sci lab equip	.053	-.052	-.004	-.066	-.052	-.029	.564	-.144	-.055	-.073	.022	-.064	.007
	Shortage instruct material	-.106	-.014	-.044	-.085	.113	-.042	-.144	.530	-.011	-.031	.003	-.028	-.013
	Shortage computers	-.059	.035	.085	.017	.026	-.066	-.055	-.011	.431	-.044	-.116	-.031	-.023
	Shortage internet	-.020	.022	.049	.036	.014	-.055	-.073	-.031	-.044	.428	-.158	.085	-.051
	Shortage computer software	.061	-.001	-.085	-.021	-.023	.033	.022	.003	-.116	-.158	.270	-.048	-.064
	Shortage library materials	-.027	.050	-.006	-.051	-.068	.007	-.064	-.028	-.031	.085	-.048	.300	-.147
	Shortage audio-visual	.018	-.054	-.001	.017	.019	-.069	.007	-.013	-.023	-.051	-.064	-.147	.248
Anti-image Correlation	Shortage of Sci Teachers	.707[a]	-.490	-.240	-.041	-.415	.185	.122	-.255	-.156	-.053	.204	-.085	.064
	Shortage of Math Teachers	-.490	.802[a]	-.276	-.139	-.014	.013	-.113	-.031	.088	.055	-.004	.148	-.176
	Shortage <test lang> Teachers	-.240	-.276	.840[a]	-.085	.085	.005	-.008	-.083	.178	.104	-.226	-.016	-.002
	Shortage Qualified Teachers	-.041	-.139	-.085	.909[a]	.075	-.281	-.116	-.154	.035	.072	-.053	-.122	.045
	Shortage library staff	-.415	-.014	.085	.075	.762[a]	-.384	-.095	.213	.055	.029	-.060	-.172	.051
	Shortage other personnel	.185	.013	.005	-.281	-.384	.854[a]	-.056	-.083	-.147	-.122	.093	.020	-.200
	Shortage sci lab equip	.122	-.113	-.008	-.116	-.095	-.056	.915[a]	-.263	-.111	-.148	.055	-.155	.019
	Shortage instruct material	-.255	-.031	-.083	-.154	.213	-.083	-.263	.892[a]	-.023	-.066	.008	-.071	-.036
	Shortage computers	-.156	.088	.178	.035	.055	-.147	-.111	-.023	.902[a]	-.103	-.338	-.085	-.069
	Shortage internet	-.053	.055	.104	.072	.029	-.122	-.148	-.066	-.103	.830[a]	-.463	.238	-.157
	Shortage computer software	.204	-.004	-.226	-.053	-.060	.093	.055	.008	-.338	-.463	.833[a]	-.169	-.246
	Shortage library materials	-.085	.148	-.016	-.122	-.172	.020	-.155	-.071	-.085	.238	-.169	.852[a]	-.539
	Shortage audio-visual	.064	-.176	-.002	.045	.051	-.200	.019	-.036	-.069	-.157	-.246	-.539	.870[a]

a. Measures of Sampling Adequacy(MSA)

Note each value along the diagonal is well above 0.5 (most ~ 0.8 / 0.9), indicating they can remain in the analysis (Field, 2009)

Also important to check those off the diagonal: Want to be small as they represent partial correlations between variables (see Field, 2009, p. 659)

Component Matrix: how items are clustered, 'fall together'

- One main factor (1),
  One smaller factor (2)

- Look at each item,
  identify highest loading

Component Matrix[a]

	Component	
	1	2
Shortage of Sci Teachers	.470	.732
Shortage of Math Teachers	.504	.703
Shortage <test lang> Teachers	.463	.612
Shortage Qualified Teachers	.641	
Shortage library staff	.556	
Shortage other personnel	.699	
Shortage sci lab equip	.695	
Shortage instruct material	.661	
Shortage computers	.686	-.441
Shortage internet	.602	-.496
Shortage computer software	.741	-.446
Shortage library materials	.805	
Shortage audio-visual	.831	

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Another option - Rotation

- Same factor analysis procedure as before, but in **Rotation** menu (below Descriptive & Extraction buttons), select **Varimax**
- This examines how they 'cluster' after rotation
- Axes of the factors can be rotated within the multidimensional variable space
- Program determines the best fit between the variables and factors
- See Field (2009) p.664 for information about rotation

- **Component 2 items** (Q01 - Q03)

- Results suggest one dimension
  - Conceptually similar (teachers!)

Communalities

	Initial	Extraction
Shortage of Sci Teachers	1.000	.778
Shortage of Math Teachers	1.000	.811
Shortage <test lang> Teachers	1.000	.684

Extraction Method: Principal Component Analysis.

- Note differences in loadings

- Different contributions
  - Hence a need for weightings

Component Matrix[a]

	Component
	1
Shortage of Sci Teachers	.882
Shortage of Math Teachers	.901
Shortage <test lang> Teachers	.827

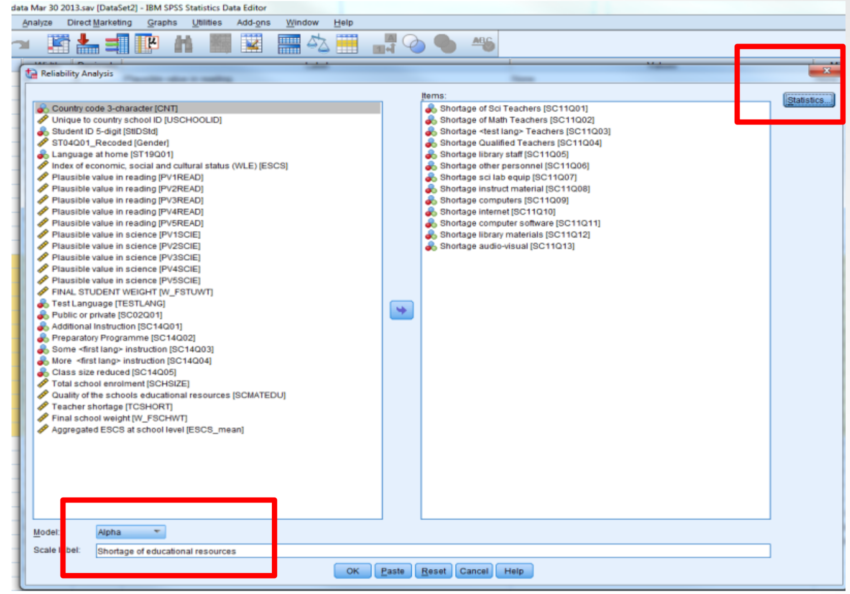
Extraction Method: Principal Component Analysis.

a. 1 components extracted.

- Check reliability estimates:

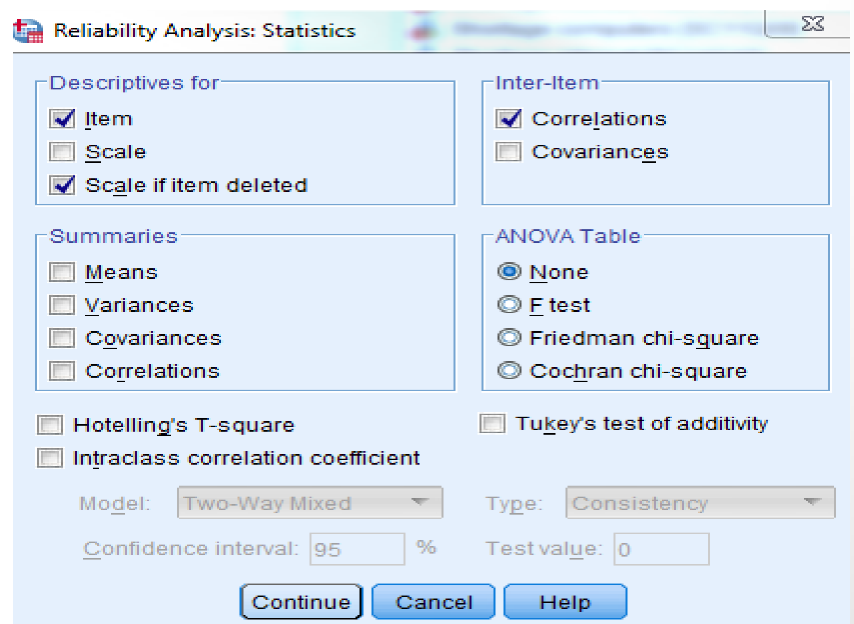
Reliability

- Analyze → Scale → Reliability Analysis
- Items: Select SC11Q01 – SC11Q03
- Label the scale
- Model: Select 'Alpha'
- Select 'Statistics' in the top right



In Reliability Analysis: Statistics pop-up window, check:

- 'Item'
- 'Scale if item deleted'
- Optional: 'Correlations'



Interpreting reliability output

- The Cronbach's alpha coefficient is an overall index of the scale; higher number = higher reliability
- Typically, $\alpha \geq 0.70$ is desirable (Nunnally & Bernstein, 1994).
- For scales containing few items (~10 or less), an $\alpha \geq 0.60$ can be an acceptable indicator of good internal consistency (Loewenthal, 1996).
  - Risky, not ideal

- Useful to paste and save syntax. Run analysis
- Reliability Statistics: Cronbach's Alpha
- Item Statistics: Sense of responses
- Inter-item Correlation Matrix: pattern of links between items
- Item-Total Statistics:

"What if"

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.822	.840	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Shortage of Sci Teachers	2.61	1.476	.738	.551	.692
Shortage of Math Teachers	2.57	1.297	.765	.586	.675
Shortage <test lang> Teachers	2.84	2.319	.636	.407	.840

Next steps…

- Results support our proposed composite (teachers)
  - Qualified teachers item?
- Think conceptually about items and how they fit

Step 4: Decide on final items for inclusion in composite

- May require some ‘trial and error’, re-calculating reliability and factor analysis estimates
- **But** should be based on theory and statistics
- Useful to be guided by both approaches rather than solely one
- Theory-driven (think about items) vs. data-driven (psychometric)

Step 5: Calculate relative weightings for each item

- Using component 2 as example (teachers)
- Take the scores for each item (from component matrix)
- Copy table into excel
- Sum scores to create total
  - Total = 2.610
- Divide each item score by this total
  - Sci Teachers: .882/2.610 = 0.338…

Component Matrix[a]

	Component
	1
Shortage of Sci Teachers	.882
Shortage of Math Teachers	.901
Shortage <test lang> Teachers	.827

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Component Matrix	**Component**	**Weight**
Shortage of Sci Teachers	0.8822	0.338046
Shortage of Math Teachers	0.9008	0.345154
Shortage <test lang> Teachers	0.8268	0.3168
Total	2.6098	1

Step 6: Apply weights to each item in the composite

- Syntax can be very important here as it keeps a record of exact instructions, and can save a lot of time

- Transform → Compute Variable
- Name target variable (Teaching_shortage)
- Numeric Expression:

  (Variable * weight) + (Variable * weight)…

  - (SC11Q01 * 0.3380) + (SC11Q02 * 0.3451) + (SC11Q03 * 0.3168).

Step 7: Derive score for the composite

- After the weights have been applied to the scores variable will (hopefully!) appear at bottom of data set

- To view the score and related details:
  Analyze → descriptive statistics → frequencies

  - Statistics:
    - Mean
    - Median
    - Std. deviation

Interpreting composites

- Remember that it is an estimate of the collection of items, albeit one that has been weighted
- Due to proportional weighting, we obtain an average /4

Statistics

		Teaching_3items	Teaching_3itemsUNWEIGHTED
N	Valid	1075	1075
	Missing	0	0
Mean		1.3422	1.3380
Median		1.0000	1.0000
Std. Deviation		.63060	.62426
Variance		.398	.390
Minimum		1.00	1.00
Maximum		3.68	3.67
Percentiles	25	1.0000	1.0000
	50	1.0000	1.0000
	75	1.3452	1.3333

Importance of weighting

- Example of scores with/ without weighting

Time for a Break

- Composite section is completed
- In about 10 minutes we will proceed to the next section:

Approaches to testing levels of aggregation using ANOVA and the unconditional multilevel approach

Validation Process

- Modern validity theory is focused on the interpretation or conclusions drawn from the survey results rather than the actual survey responses (Borden & Young, 2008).

Multi-level validity

- From this modern perspective, we also need to consider the validity of aggregated survey responses

- When we aggregate individual student ratings to the program level or institutional level, we need to ensure the meaning is the same at both levels – absolute correspondence (Zumbo & Forer, 2011)

- Studies have found that factor analysis at both the individual group level there have been demonstrated distinct or fuzzy latent factors at both levels when making use of perceptual data (D'Haenens, Van Damme, & Onghena, 2008)

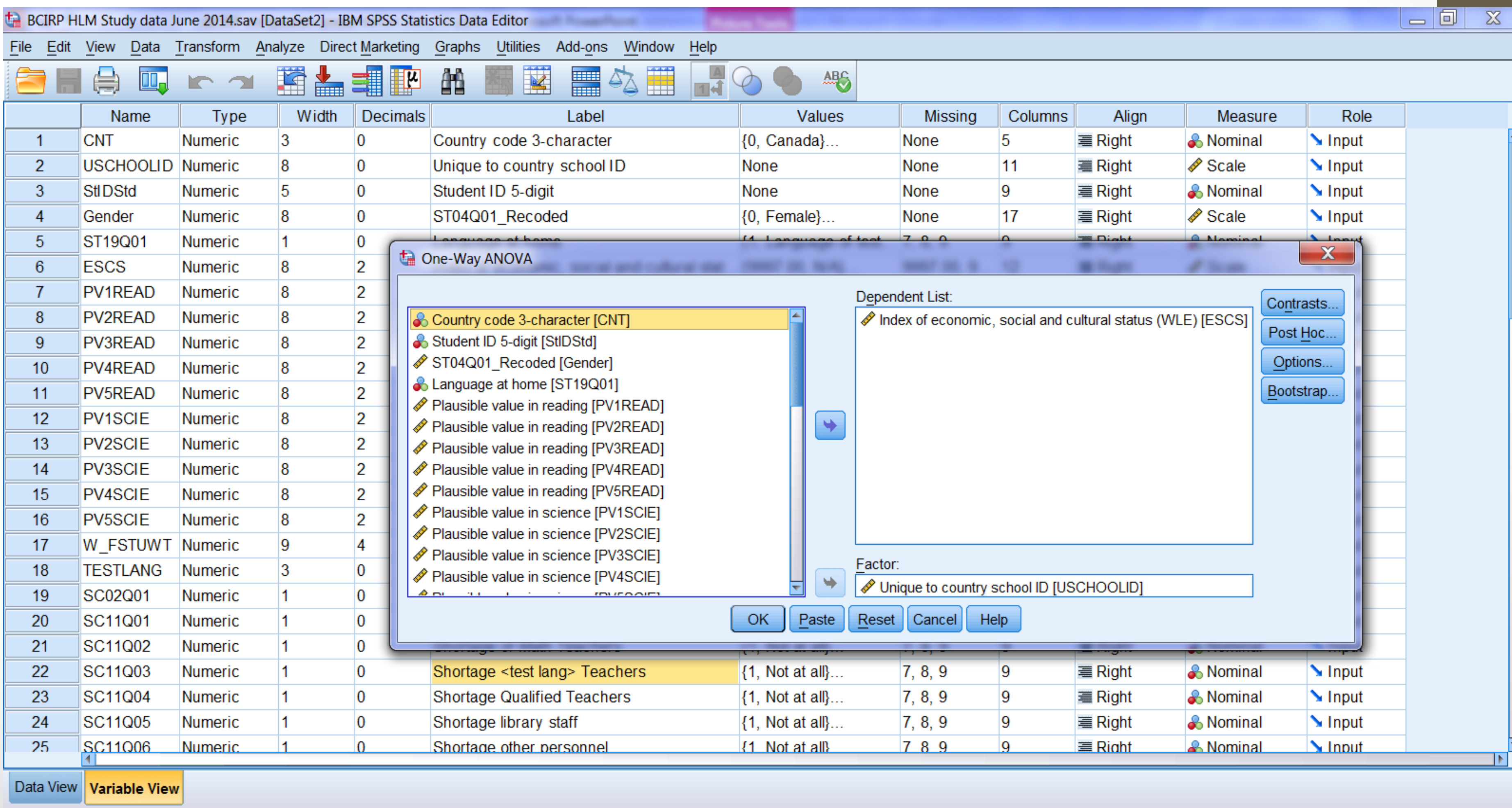
Can we aggregate SES for each school?

- Analysis of Variance (ANOVA)
  1. Non-independence – Intraclass correlation ICC(1)
  2. Reliability of program means – ICC(2)
  3. Within-group agreement - Rwg

- Unconditional Multilevel model
  1. Proportion of variance explained
  2. Extent to which member responses on the dependent variables indicate group responses
  3. Extent variation in the dependent variable can distinguish group membership

Assess non-independence

Analyze → compare means → one-way ANOVA



Assess non-independence

ANOVA

Index of economic, social and cultural status (WLE)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	336.346	129	2.607	4.313	.000
Within Groups	571.317	945	.605		
Total	907.663	1074			

$$\text{ICC}(1) = \frac{MS_B - MS_W}{MS_B + (k-1)MS_W} \qquad 0.29 = \frac{2.607 - 0.605}{2.607 + (8-1) * 0.605}$$

MSb = mean square between groups

MSw = mean square within groups

K = average number of individuals per group*

Assess reliability of means

$$\text{ICC}(2) = \frac{k\ x\ ICC(1)}{1 + (k-1) x\ ICC(1)}$$

$$0.29 = \frac{2.607 - 0.605}{2.607 + (8-1) * 0.605}$$

$$0.77 = \frac{8\ x\ 0.29}{1 + (8-1) x\ 0.29}$$

Assess within-group agreement

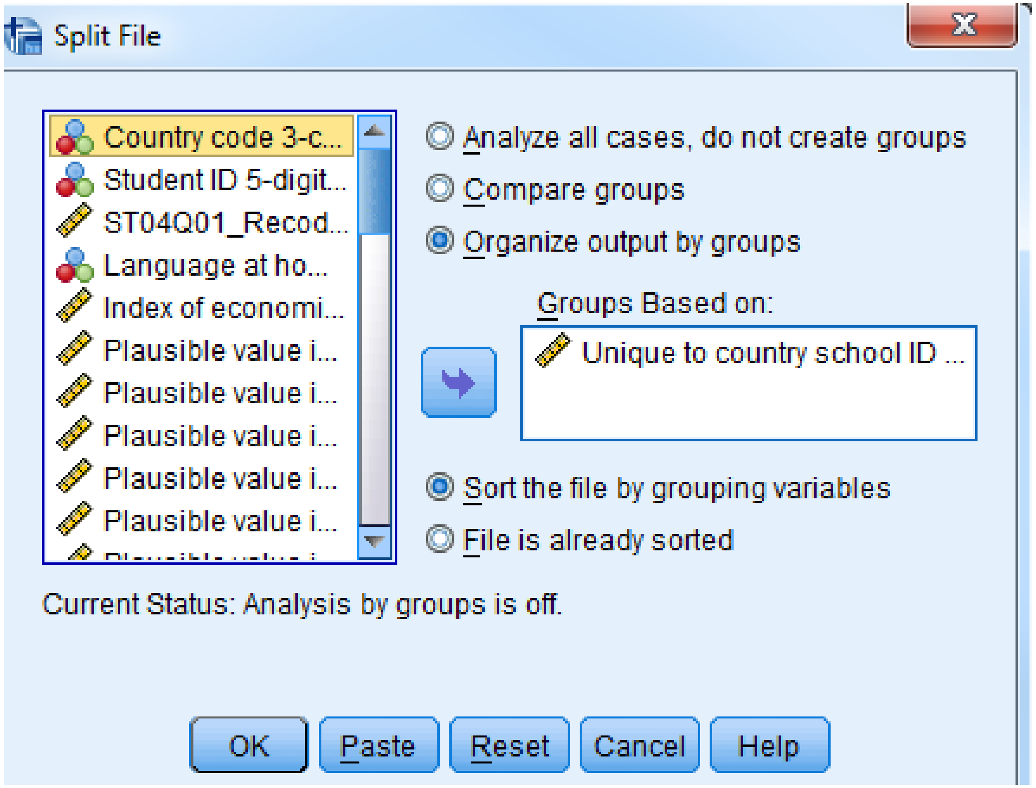
$$r_{WG} = 1 - \frac{s_x^2}{s_{EU}^2}$$

$$s_{EU}^2 = \frac{A^2 - 1}{12}$$

- ICC(1) and ICC(2) were omnibus tests, but the within-group agreement is calculated separately for each school and then averaged across.
- Run variances for each school…

- Data→ Split File… → Organize output by groups
  - Compare groups: Unique to country school ID
  - Groups Based on: Unique to country school ID

- Analyze → descriptives → select options…check variance



Assess within-group agreement

- determine the null distribution to which compare your sample to, which is determined on the type of data you have, either multi-items or single items and the number of response options available to raters

- In this example, we had one item, ESCS, that was standardized and ranged from -3.0 to +3.0

- Using this information, the average rwg statistic was calculated at 0.78

ANOVA Results…

ANOVA STEPS	Calculated on this Sample	Criteria
1. non-independence	0.29	> 0.12
2. reliability/ correspondence	0.77	> 0.83
3. within-group agreement	0.78	> 0.70

- ESCS is associated with group membership (e.g., the school), and we can say that about 29% of variance in ESCS was related to school membership
- The reliability of the means is high, but not quite as high as we might want; could be that some schools have greater within-school variability among ESCS
- There appears to be high within-group agreement, on average, meaning that ESCS seems to be homogenous within schools.

Another Aggregation Approach

- Unconditional multilevel model
- Considers the hierarchical structure of the data
  - students are nested in schools
- No variables are included in the model other than the dependent variable

Level 1: $Y_{ij} = \beta_{0j} + r_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

Unconditional multilevel model

- SPSS SYNTAX TO RUN AN UNCONDITIONAL MULTILEVEL MODEL:

```
MIXED ESCS
/PRINT = SOLUTION TESTCOV
/METHOD = REML
/FIXED = INTERCEPT
/RANDOM = INTERCEPT | SUBJECT (USCHOOLID).
```

Unconditional Multilevel Model

Fixed Effects

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	131.120	54.099	.000

a. Dependent Variable: Index of economic, social and cultural status (WLE).

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	.363639	.049439	131.120	7.355	.000	.265837	.461441

a. Dependent Variable: Index of economic, social and cultural status (WLE).

Covariance Parameters

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		.603518	.027707	21.782	.000	.551585	.660340
Intercept [subject = USCHOOLID]	Variance	.236948	.039096	6.061	.000	.171478	.327415

a. Dependent Variable: Index of economic, social and cultural status (WLE).

Unconditional MLM results

Covariance Parameters

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Residual		.603518	.027707	21.782	.000	.551585	.660340
Intercept [subject = USCHOOLID]	Variance	.236948	.039096	6.061	.000	.171478	.327415

a. Dependent Variable: Index of economic, social and cultural status (WLE).

- Proportion of Variance Explained = $\hat{\tau}_{00}/(\hat{\tau}_{00} + \hat{\sigma}^2)$

= 0.236948 / (0.23695 + 0.60352)

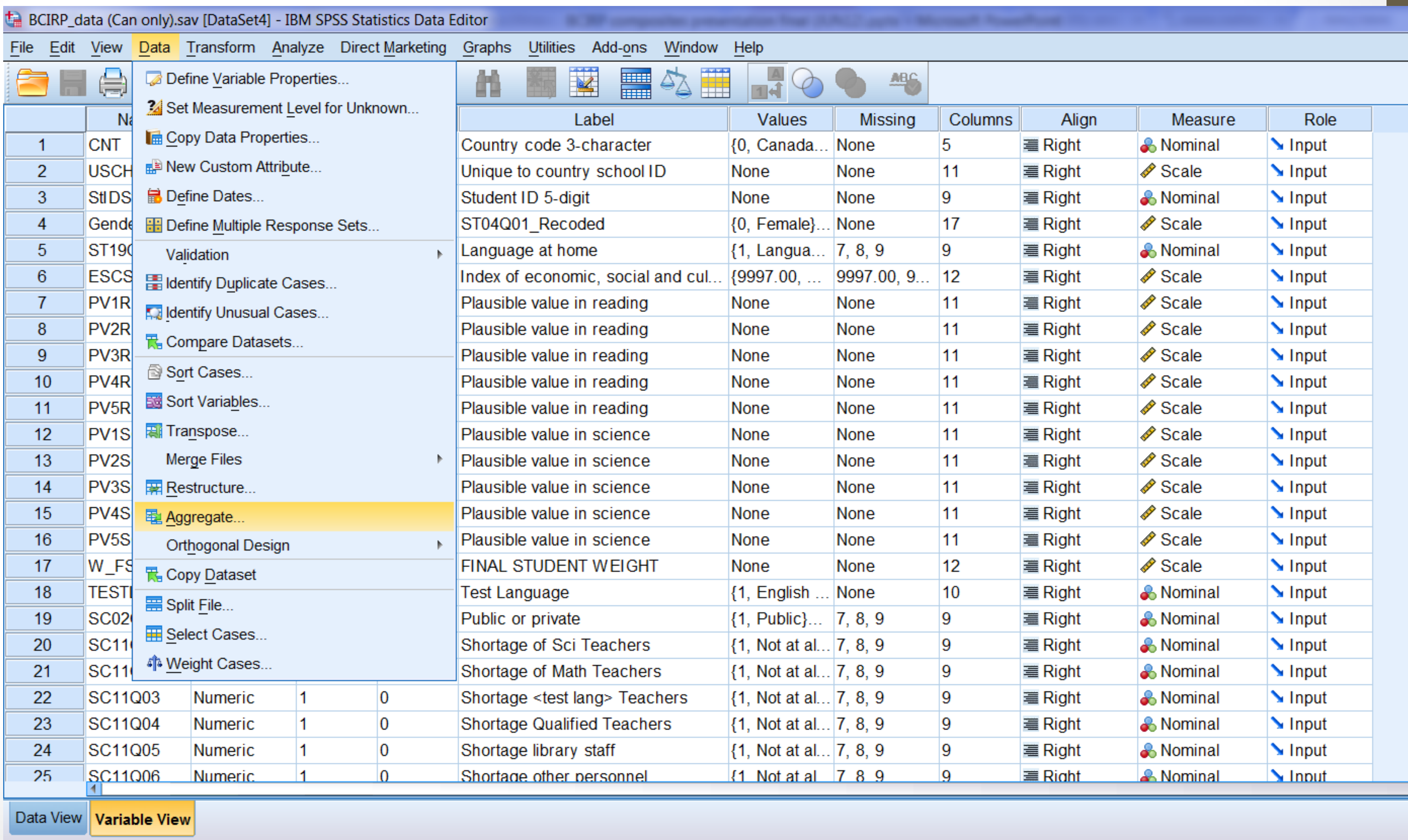
= 0.236948 / 0.84047

= 0.28

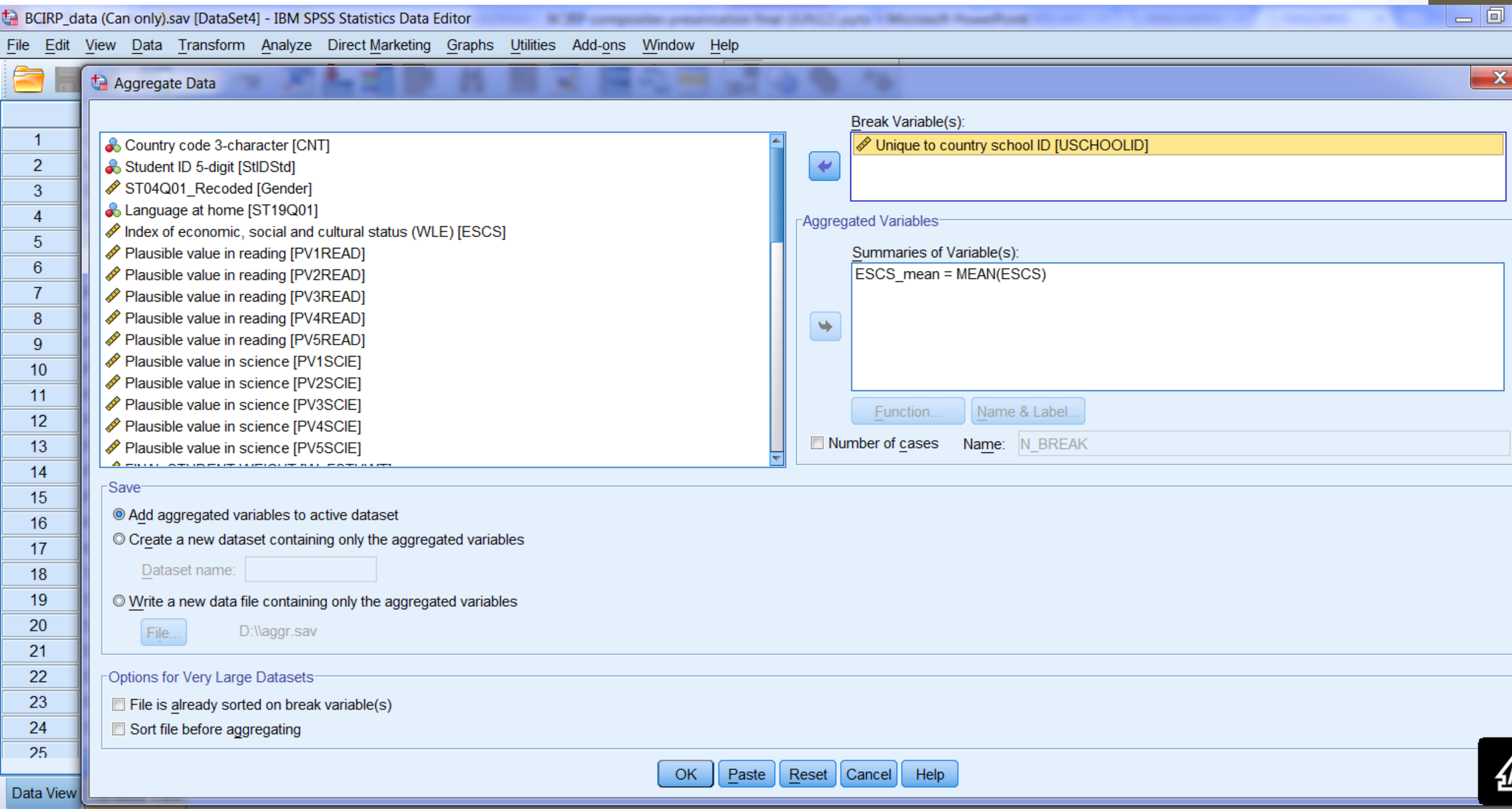
Unconditional MLM Results…

- Results were very similar to the ANOVA
- ESCS is associated with group membership (e.g., the school), and we can say that about 28% of variance in ESCS was related to school membership
- The reliability of the means is acceptable, at 0.75, so when we take into consideration students nested in their programs it reduced the reliability slightly from 0.77.
- The large F-values and p-values indicated that the schools in this sample could be differentiated using ESCS .

To aggregate or not to aggregate?



Aggregating in SPSS



Using the Aggregate Score…

- Multilevel model that examined two levels running regression equations simultaneously at the student level and school level

- **Student Level:**
  - Dependent Variable: PISA PV Science Achievement Scores
  - Independent Variables: gender and ESCS

- **School Level:**
  - School size
  - School mean ESCS (newly calculated)
  - Teaching composite score (newly created)

Results…

- We applied an intercepts- and slopes-as-outcomes model to these data…
- The average score on the science achievement was 511.71 with a standard error of 4.64
- There were statistically significant differences among schools in their average science achievement scores

Results…

- As the average ESCS score per school increased, so did their scores on science achievement, by about 67.49 points $p < 0.001$
- Also, within each school, students with higher ESCS did better than other students in their school by about 10 points $p < 0.05$.
- Reliability estimate for the full model was at 0.71…still unexplained variance $p < 0.001$…more work to do!

DISCUSSION? OTHER IDEAS OR COMMENTS?

stephanie.mckeown@ubc.ca

scott.emerson@ubc.ca

References & resources

- Bliese, P. D. (2000). *Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis.* In K. J. Klein & J. S. Kozlowski (Eds.), Multilevel theory, research, and methods in organizations (pp. 349-381). San Francisco: Jossey-Bass.
- Borden, V. M., & Young, J. W. (2008). Measurement validity and accountability for student learning. *New Directions for Institutional Research*, (1), 19-37.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
- DiStefano, C., Zhu, M., & Mîndrilă, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation, 14*(20).
- Field, A. (2009). *Discovering statistics using SPSS*. Sage publications.
- Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: An introduction*. Sage.
- Griffith, J. (2002). Is Quality/Effectiveness An Empirically Demonstrable School Attribute? Statistical Aids for Determining Appropriate Levels of Analysis. *School Effectiveness and School Improvement, 13*, 1, 91-122.
- Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist*. London: Sage.

- Kane, M. (2006). *Validation*. In R. Brennan (ed.), Educational Measurement. (4th ed.). Westport, Conn.: American Council on Education/Praeger.
- LeBreton, J. M. and Senter, J. L. (2008). *Answers to 20 Questions About Interrater Reliability and Interrater Agreement.* Organizational Research Methods, 11, 4, 815-852.
- Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of Learning Environments: How to Use Student Ratings of Classroom or School Characteristics in Multilevel Modeling. *Contemporary Educational Psychology, 34, 2, 120–131*.
- Pike, G. R.,. (2006). The convergent and discriminant validity of NSSE scalelet scores. *Journal of College Student Development, 47* (5), 551-564.
- Zumbo, B. D. (2007). Validity: Foundational issues and statistical methodology. *Handbook of statistics: Psychometrics*, *26*, 45-79.
- Zumbo, B. D. (2009). Validity as contextualized and pragmatic explanation, and its implications for validation practice. *The concept of validity: Revisions, new directions and applications*, 65-82.

Other useful resources

- Allison, P. D. (2001). *Missing Data*. Thousand Oaks, CA: Sage Publications.
- Little, R. J. A. & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York, Wiley.
- Organisation for Economic Co-operation and Development (2009). *PISA 2006 technical report*. Paris: Author.
- Organisation for Economic Co-operation and Development (2010). *PISA 2009 results: Executive summary.* Paris: Author.
- Organisation for Economic Co-operation and Development (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: Author.