Greg Johnson

Exam Part II: Psychometric Analysis of ACS Scale

Walkthrough of Analysis: Response Distributions

FlexMIRT doesn’t run if a particular response category within an item has less than 10% of total participant’s responses for that item. A criterion of less than 10.5% was adopted when collapsing response categories. Response categories were always collapsed into the adjacent response category with the smaller response distribution e.g. for item 2, the ‘somewhat disagree’ response had 8.3% of total responses and so what was collapsed into ‘disagree’ (11.5%) rather than ‘neutral’ (22.5%).

No one item remained on a full, seven-point Likert-type scale. Most were reduced to 5 categories; some to 4; a few to 6. The full details are in the excel sheet output.

Walkthrough of Analysis: CEFA

The ACS scale was assessed for unidimensionality by factor-analyzing the correlation matrix of its items, derived from 435 sets of responses on the scale. The following criteria were considered (in descending order of importance and consideration) when finding the optimal factor solution:

1. Variance explained
2. Fit Indices
3. Kaiser criterion
4. Quality of factor loadings
5. Scree Plot
6. Interpretability of factors

First eigenvalue/number of items = 9.938/15 = 66.3% of the variance explained. Suggests that the one-factor solution is adequate.

Varimax orthogonal rotation was used with one exception, explained below. Given the weak theoretical background for a multi-factor solution (Adlerian theory from the 1920s and possible biased response from questionnaires about intrapersonal motivations for surgery), Varimax orthogonal rotation was used to maximize the interpretability of the factors (if more than one was extracted) despite the strong assumption of uncorrelated factors.

Fit Indices:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Factors |  |  |  |  |  |  | C.I. |  |  |  |
| 0 | 16.14 | 7008.14 | 105 | .000 | .000 | .389 | (.38,.40) |  |  | 16.22 |
| 1 | 3.47 | 1505.07 | 90 | .000 | .000 | .190 | (.18,.20) |  |  | 3.61 |
| 2 | 1.70 | 736.77 | 76 | .000 | .000 | .142 | (.13,.15) |  |  | 1.90 |
| 3 | .674 | 292.39 | 63 | .000 | .000 | .092 | (.08,.10) |  |  | .936 |
| 3\* | .674 | 292.38 | 63 | .000 | .000 | .092 | (.08,.10) |  |  | .936 |
| 4 | .464 | 201.22 | 51 | .000 | .000 | .082 | (.07,.10) |  |  | .782 |
| 5 | .266 | 115.51 | 40 | .030 | .000 | .066 | (.05,.08) |  |  | .635 |
| 6 | .151 | 65.56 | 30 | .390 | .000 | .052 | (.04,.07) |  |  | .566 |

\*3-factor solution with oblique rotation was used to try replicate the original paper. An identity matrix based on Table 1 from the paper was supplied to CEFA.

Fit indices appeared to be pulling towards a +6 factor solution. Parsimony suffered accordingly. The 3-factor model replicating the model detailed in the original paper was as unsatisfactory as the other factor models detailed here.

Eigenvalues: 9.938, 1.288, 0.8556, 0.5470, 0.3854, 0.3362, 0.2942, 0.2721, 0.2167, 0.2025, 0.1699, 0.1526, 0.1372, 0.1236, 0.08098

Kaiser criterion: 9.938, 1.288. 2-factor solution suggested.

Quality of factor loadings: adequate until 4-factors or more which is when solutions begin to give signs of being over-factored. Highlighted in CEFA output.

The Scree Plot suggested a 1- or 2-factor model.

Interpretability of factors: this criterion was pretty much ignored because it is incredibly subjective.

Considering the different criteria (especially the variance-explained), a one-factor solution was considered adequate for IRT analysis and a better solution than any multi-factor solution, including the 3-factor solution put forth by the original ACS paper. The eigenvalues suggest possibly 2 factors; a third factor would explain very little variance. There are an infinite amount of rotations so it is not impossible that the authors of the ACS paper milked their factor analysis (which was actually PCA, a data reduction technique) until it squeezed out factor loadings that supported their theory.

Walkthrough of Analysis: GPC vs NRM Model

|  |  |  |
| --- | --- | --- |
|  | GPC Model | |
|  | CRC | IIF |
| Item 1 | Neutral is doing very little. | Narrow range; nice, high peak |
| Item 2 | All 5 categories are functioning nicely. | All categories providing nice info. |
| Item 3 | The middle categories have high conditional probabilities for tiny intervals of theta yet the CBD is huge. | The 4 category boundaries are close to each other and have steep discriminations, leading to very high info in the mid-range of theta. |
| Item 4 | No category is really slacking even if they are again packed near theta of 0. | Item info is nearly identical to that of Items 1 and 2. |
| Item 5 | This time the lowest category (strongly disagree in this case) is doing very little. | Nice peaked info again but this time shifted over to around -0.5 theta. |
| Item 6 | Extremely high CBD for the item. | Info is off the charts, suggesting an artefact. |
| Item 7 | This is one of two items that retains 6 categories and it shows: some categories never have a section of theta in which they have the highest conditional probability. | A few categories are either giving low info or are giving redundant info in relation to the adjacent categories. |
| Item 8 | Nice CBD. Again centered around theta of 0. | Info goes off the graph. |
| Item 9 | ‘Disagree’, ‘somewhat disagree’ and ‘neutral’ have low peaks in their response curves. | The category infos for the aforementioned items are very low. |
| Item 10 | CRCs are awful. CBD is rock bottom. | Info flatlines. This is the first truly bad item. |
| Item 11 | All of the categories in the middle appear to be functioning quite badly. | Info for middle items are all the same, low height in the same range of theta. Redundancy here. |
| Item 12 | Another item where the extreme categories are functioning the best. | Only good info is coming from strongly disagree and (agree and strongly agree). |
| Item 13 | Neutral and (somewhat agree, agree, strongly agree) are functioning poorly. | Getting nice info from all but neutral. |
| Item 14 | (Disagree and somewhat disagree) are functioning poorly. | All category infos are very low. The only reason the item info is somewhat large is the number of categories. |
| Item 15 | Categories are again packed around theta of 0 but the CBD is very steep. | Off the charts info around theta of 0. |
| Test | High, peaked information around a little above theta of 0. Drops off very sharply so that there is very little info for those below a theta of -1 and above a theta of 2. | |

The GPC Model seems to produce very, very nice information at the category, item, and test levels. However the NR Model should also be applied to examine the variability of CBDs within items because if the variability of CBDs is large, the GPC Model is washing out the high and low within-item CBDs.

|  |  |  |
| --- | --- | --- |
|  | GPC Model | |
|  | CRC | IIF |
| Item 1 | Neutral is doing very little here, as evidenced by the CBDs that it forms with it’s adjacent categories. | Little to no information from neutral but the information from the other categories is nicely spread out, giving the whole item nice spread. |
| Item 2 | The neutral and agree categories have very shallow CBDs with each other. | The CBD from strongly disagree and (disagree & somewhat disagree) is very sharp, giving a lot of information below theta of 0, which was lacking in the GPC analyzed models. |
| Item 3 | Neutral is the category with the highest conditional probability in a tiny range of theta yet if has strong CBDs. | The category with the worst info is (agree & strongly agree) because its one CBD is very shallow. |
| Item 4 | The agree categories again form poor CBDs. | Info peaks over -1 theta and skews to the right due to the poor CBDs from the agree categories. |
| Item 5 | Like the previous two items, the CBDs decrease in steepness as one moves up the categories. | Info again peaks over -1 theta and skews to the right. |
| Item 6 | This is a 4-category item and all of its CBDs are incredibly steep | Info is off the charts over a wide range of theta. |
| Item 7 | Interestingly, strong CBDs sandwich the weak CBDs in the middle i.e. the neutral category. | As a result, the item information dips in the middle. |
| Item 8 | A 4-category item with very strong CBDs. | Great info, off the charts. |
| Item 9 | Looks like the somewhat disagree category and neutral are functioning poorly and are responsible for the horrendous .83 CBD. | Item info skews off to the right because the higher categories have little infos. |
| Item 10 | This is a very poor item. Visual inspection indicates very low CBDs. One is actually negative, meaning that strongly agree and (agree and somewhat agree) actually trade places in terms of the theta needed to endorse them. Strange. | Item info is essentially a flat line. No amount of category collapsing will save this item. It should be scrapped. |
| Item 11 | There are some redundant categories here. Based on the CBDs, it looks neutral may have to be collapsed into the agrees. | The item info is getting most of its height from the two most extreme categories. |
| Item 12 | Almost looks identical to Item 11. Looks like a candidate for dichotomization. | Most of the item info seems to be coming from strongly disagree. |
| Item 13 | Disagree seems to be functioning poorly. | Disagree has poor item info but the overall item info is spread from -1 to+2 theta. |
| Item 14 | The best CBD is between strongly disagree and (disagree & somewhat disagree). The middle CBDs are very poor. | The infos for the middle categories are awful. The strongly disagree info is great, giving peak to the total item info. The (agree & strongly agree) category has a low but very spread out info. |
| Item 15 | Looks like a legitimate item. The CBDs are similar enough that the Wald test may find that the GPCM may be a better fit. | Great info which dips in the middle. |
| Test | Info doesn’t peak as much as the GPCM test info but it is more spread out and gives more information below a theta of 0. The info does drop off at +/- 2 theta. | |

When CBDs vary within-item in the NR Model, we see that a lot of the first CBDs for most of the items are massive compared to the later CBDs (e.g. the CBD’s between strongly disagree and disagree/somewhat disagree are much steeper than those between neutral and agree or agree and strongly disagree). Adoption of the NR Model would allow the Wald test to empirically test which items should have the GPC Model restrictions and which should stay under the NR Model; this way, items with significant variance of within-item CBDs don’t have one omnibus CBD applied to all category boundaries, blunting the large CBDs and inflating the small CBDs (which may explain why some of the IIFs under the GPC Model appeared to be artefactually high and the test info is so high).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | -2LL | AIC | BIC | M2 | df | p | F0hat | RMSEA |
| Null | 19822.37 | 19940.37 | 20180.82 | 30653.99 | 1622 | 0.0001 | 70.4689 | 0.2 |
| GPCM | 14832.59 | 14980.59 | 15282.16 | 6082 | 1607 | 0.0001 | 13.9816 | 0.08 |
| NRM | 14557.95 | 14813.95 | 15294.84 | 5889.36 | 1563 | 0.0001 | 13.5388 | 0.08 |

In terms of model fit, the fit indices universally give the NR Model the upper hand. This is of course because the NR Model has less fixed parameters and so fits the data better. The most telling fit index is the RMSEA which should penalize the NR Model for losing parsimony (less df) unless it makes up for it in fitting the data. The RMSEA is the same for the GPCM and the NRM, suggesting that the NRM was not penalized for losing parsimony. Based on the fit indices for unilateral application of an IRT model, the NR Model seems to be the way to go.

Therefore, the NR Model should be applied to all items, then a Wald test can be performed for differential application of the NRM and GPCM. In addition, based on visual inspection of the CBD matrices for the items under the NR Model, certain categories should be collapsed due to the shallow CBDs they form. The criterion used was CBD ≤ 1, give or take .1 or so. The CBDs I decided would lead to category collapse are highlighted in the output containing the Model Fit Indices and Parameters for the NRM and GPCM.

The ACS Scale-II

\* The final response format of a particular item is bolded.

**Item 1**: It make sense to have minor cosmetic surgery rather than spending years feeling bad about the way you look.

Initial: SD D soD N soA A SA 7-cat

Frequency collapse: SD (D & soD) N soA (A & SA) 5-cat

CBD collapse: **SD (D & soD & N & soA) (A & SA)** **3-cat**

**Item 2**: Cosmetic surgery is a good thing because it can help people feel better about themselves.

Initial: SD D soD N soA A SA 7-cat

Frequency collapse: SD (D & soD) N soA (A & SA) 5-cat

CBD collapse: **SD (D & soD) N (soA & A & SA)**  **4-cat**

**Item 3**: In the future, I could end up having some kind of cosmetic surgery.

Initial: SD D soD N soA A SA 7-cat

Frequency collapse: **SD (D & soD) N soA (A & SA) 5-cat**

CBD collapse: No collapse because no poor CBDs.

**Item 4**: People who are very unhappy with their physical appearance should consider cosmetic surgery as one option.

Initial: SD D soD N soA A SA 7-cat

Frequency collapse: SD (D & soD) N soA (A & SA) 5-cat

CBD collapse: **SD (D & soD) N (soA & A & SA) 4-cat**

**Item 5**: If cosmetic surgery can make someone happier with the way they look, then they should try it.

Initial: SD D soD N soA A SA 7-cat

Frequency collapse: SD (D & soD) N soA (A & SA) 5-cat

CBD collapse: **SD (D & soD) N (soA & A & SA) 4-cat**

**Item 6**: If I could have a surgical procedure done for free I would consider trying cosmetic surgery.

Initial: SD D soD N soA A SA 7-cat

Frequency collapse: **SD (D & soD & N) soA (A & SA) 4-cat**

CBD collapse: No collapse because no poor CBDs.

**Item 7**: If I knew there would be no negative side effects or pain, I would like to try cosmetic surgery.

Initial: SD D soD N soA A SA 7-cat

Frequency collapse: **SD (D & soD) N soA A SA 6-cat**

CBD collapse: No collapse because no poor CBDs.

**Item 8**: I have sometimes thought about having cosmetic surgery.

Initial: SD D soD N soA A SA 7-cat

Frequency collapse: **SD (D & soD & N) soA (A & SA) 4-cat**

CBD collapse: No collapse because no poor CBDs.

**Item 9**: I would seriously consider having cosmetic surgery if my partner thought it was a good idea.

Initial: SD D soD N soA A SA 7-cat

Frequency collapse: SD D soD N (soA & A & SA) 5-cat

CBD collapse: **SD D (soD & N) (soA & A & SA) 4-cat**

**Item 10**: I would never have any kind of plastic surgery.

Initial: SD D soD N soA A SA 7-cat

Frequency collapse: **SD D soD N (soA & A) SA 6-cat**

CBD collapse: Item dropped due to poor CBDs and category reversal of A and SA.

**Item 11**: I would think about having cosmetic surgery in order to keep looking young.

Initial: SD D soD N soA A SA 7-cat

Frequency collapse: SD D soD N (soA & A & SA) 5-cat

CBD collapse: **SD D (soD & N) (soA & A & SA) 4-cat**

**Item 12**: If it would benefit my career I would think about having plastic surgery.

Initial: SD D soD N soA A SA 7-cat

Frequency collapse: SD D soD N (soA & A & SA) 5-cat

CBD collapse: **SD (D & soD & N) (soA & A & SA) 3-cat**

**Item 13**: I would seriously consider having cosmetic surgery if I thought my partner would find me more attractive.

Initial: SD D soD N soA A SA 7-cat

Frequency collapse: **SD D (soD & N) (soA & A & SA) 4-cat**

CBD collapse: No collapse because no poor CBDs.

**Item 14**: Cosmetic surgery can be a big benefit to people’s self-image.

Initial: SD D soD N soA A SA 7-cat

Frequency collapse: SD (D & soD) N soA (A & SA) 5-cat

CBD collapse: **SD (D & soD & N & soA) (A & SA) 3-cat**

**Item 15**: If a simple cosmetic surgery procedure would make me more attractive to others, I would think about trying it.

Initial: SD D soD N soA A SA 7-cat

Frequency collapse: **SD (D & soD) N soA (A & SA) 5-cat**

CBD collapse: No collapse because no poor CBDs.

In conclusion, the seven-point response format was terrible. The majority of final response formats for items were 3- or 4-categories. Disagree and somewhat disagree, and agree and strongly agree were consistently collapsed. Too much choice is given here. However a Likert scale appears to have been a good choice because no item became dichotomous and neutral fared pretty well, getting over 10% of the responses and avoiding collapse in most of the items.

The next step would be to decide what the collapsed categories become e.g. does the combination of somewhat disagree, neutral and somewhat agree just become neutral? Then the revised ACS Scale can be administered to a new sample so that IRT analysis can evaluate how these new collapsed categories are functioning.