**The initial parallel analysis suggests 9 factors, so I started from 9.**

General approach. I will start with running EFA with “Oblimin” rotation with all items and the specified number of factors. Later, I will trim out all the items that does not have good loading to any factor (loading < 0.3) to make sure the I get a clearer picture of all the item loadings. After obtaining a matrix of items loading, I will start inspecting items in each factor separately and see if they make sense conceptually.

9 factors model was abandoned because after inspection the6th, 8th, and 9th factors are not conceptually distinguishable: they mostly talk about treating other people equally. Therefore, I decided to drop the 9 factors model.

1. Factor 1
   1. 7,13,15,17,18,19,22,30,32,33,36,49,56,60,61,73,75
   2. This cluster of items are talking about **prosocial values**.
2. Factor 2
   1. 5,10,15,26,41,47,54,57,69,72,74
   2. This cluster of items are talking about **perception of reciprocal norms.**
   3. Note that the items in this cluster are mostly reverse coded items.
3. Factor 3
   1. 9,20,27,29,46,51,52
   2. This cluster of items are talking about **empathy**.
4. Factor 4
   1. 1,21,44,63,65
   2. This cluster of items are talking about **attitudes toward social problems**.
5. Factor 5
   1. 20,24,28,35,48,68
   2. This cluster of items are talking about **attitude toward prosocial spending (relates to effective altruism).**
6. Factor 6
   1. 45,67,70
   2. This cluster of items are talking about **fairness (categorical imperative).**
7. Factor 7
   1. 2,11,48
   2. This cluster of items are talking about **behavioral elements relating to prosocial spending** which seems to overlap with factor 5. Indeed, the 48 have good loading on both factors.
8. Factor 8
   1. 38,-57,-63,71,73,-74 (negative signs indicate negative loading)
   2. After inspection, this cluster seems to contain multiple elements such as **care for strangers (fairness which relates to factor 6)** and **disbelief about small effort combined to make big changes (relates to factor 2 and 4).**
9. Factor 9
   1. 6,-9,19,39,55
   2. This cluster seems to talk about **fairness and care for strangers (relates to factor 6).**

Looking at the results, I reckon that factors **6,8, and 9 seem to overlap** in their conceptual meaning. Therefore, I explore whether they could be combined into a more general laten variable. Through a 7 factors EFA model with “Oblimin” rotation, I want to see if all the variables relating to fairness will merge. In this model, I followed the same procedure that I indicated above (trimmed out loading <0.3). The result supported my thoughts: all items relating to fairness merge into single factor in the 7 factors model.

1. Factor 1 (ML1)
   1. 4,7,12,15,17,18,19,22,25,30,32,33,34,36,37, -38,49,55,56,57,58,59,60,61,-62,63,64,-71,74,75
   2. High loading items (>0.45): 4,12,15,17,19,22,32,33,49,56,60,75
   3. This cluster of items are talking about **prosocial values and seems to include 3 subcomponents, 1) urge to help (proactiveness), 2) selfless orientation, and 3) insight in helping**.
   4. Due to the complexity of this factor, I further explore its subcomponents using EFA and CFA later.
2. Factor 2 (ML7)
   1. 3,5,8,10,15,26,33,41,44,45,47,54,57,58,69,72,74
   2. High loading items (>0.45): 10,26,41,54,57,69,72
   3. This cluster of items are talking about **perception of reciprocal norms and seems to include 3 subcomponents: 1) reciprocal norms, 2) activists’ ideology, and 3) social norms.**
   4. Note that the items in this cluster are mostly reverse coded items.
3. Factor 3 (ML4)
   1. 9,27,29,46,51,52
   2. High loading items (>0.45): 9,27,29,46,51,52
   3. This cluster of items are talking about **empathy**.
4. Factor 4 (ML6)
   1. 1,21,44,63,65,69
   2. High loading items (>0.45): 1,21,44,65
   3. This cluster of items are talking about **attitudes toward social changes**.
5. Factor 5 (ML2)
   1. 20,24,28,35,-47,48,64,68
   2. High loading items (>0.45): 20,28,35,48,68
   3. This cluster of items are talking abou t **attitude toward prosocial spending (relates to effective altruism).** This cluster seems to contain **affective and reflective components**.
6. Factor 6 (ML5)
   1. 6,18,38,39,45,55,62,67,70,71
   2. High loading items (>0.45): 6,38,39,45,67,70,71
   3. This cluster of items are talking about **fairness (categorical imperative).**
7. Factor 7 (ML3)
   1. 2,5,11,13,48,66,73
   2. High loading items (>0.45): 2,11
   3. This cluster of items are talking about **behavioral elements relating to self-sacrifice.**

7 factor solution seems to provide a good conceptual as well as statistical illustration for the factor structure in the do-gooder scale (CFI = 0.93, RMSEA=0.04). Therefore, I adopted this model. However, when we attempt to create a more concise model using higher filtering criteria for factor loadings (>0.45), the 7th factor seems to no longer hold strongly, as it only have two strong items left. Looking back at the

**Higher and lower level factors:**

To figure out if there is upper or lower-level structure for the 7 factors model, I ran some more EFAs and CFAs.

In particular, for exploring the upper-level factors, I looked at the correlation between 7 factors.

Text

Description automatically generated

No factors correlate strongly with others (by strongly I mean r>0.5). Therefore, I did not further explore the upper-level structure since the correlation matrix indicates it may not exist.

For the lower-level factors, I ran some EFA in each latent factor to see if they are unidimensional.

Factor 1 (ML1)

1. Sub-factor 1 (ML1)
   1. 12,15,17,18,30,33,34,57,58,61,63,74
   2. This one seems to **reflect proactiveness and urge to do good**.
2. Sub-factor 2 (ML3)
   1. 4,7,19,22,25,32,37,49,55,56,59,64,75
   2. This one reflects **insight in helping**.
3. Sub-factor 3 (ML4)
   1. 17,32,36,56,60
   2. This one reflects **selfless orientation**.
4. Sub-factor 4 (ML2)
   1. 38,62,71
   2. This one reflects **care for strangers** which overlaps with factor 6.Also, considering the low loading (<0.4) in the original 7 factors EFA, I decided to drop this sub-factor.

**Concise version:**

Also, depending on the requirements of conciseness, I decided to create a more accurate and brief version of the DGS by trimming out factors with relatively low factor loadings or making little conceptual sense for each latent variable.

Factor 1 (ML1)

* 1. High loading items (>0.45): 4,12,15,17,19,22,32,33,49,56,60,75
  2. This cluster of items are talking about **prosocial values and seems to include 3 subcomponents, 1) urge to help (proactiveness), 2) selfless orientation, and 3) insight in helping**.
  3. Due to the complexity of this factor, I further explore its subcomponents using EFA and CFA later.

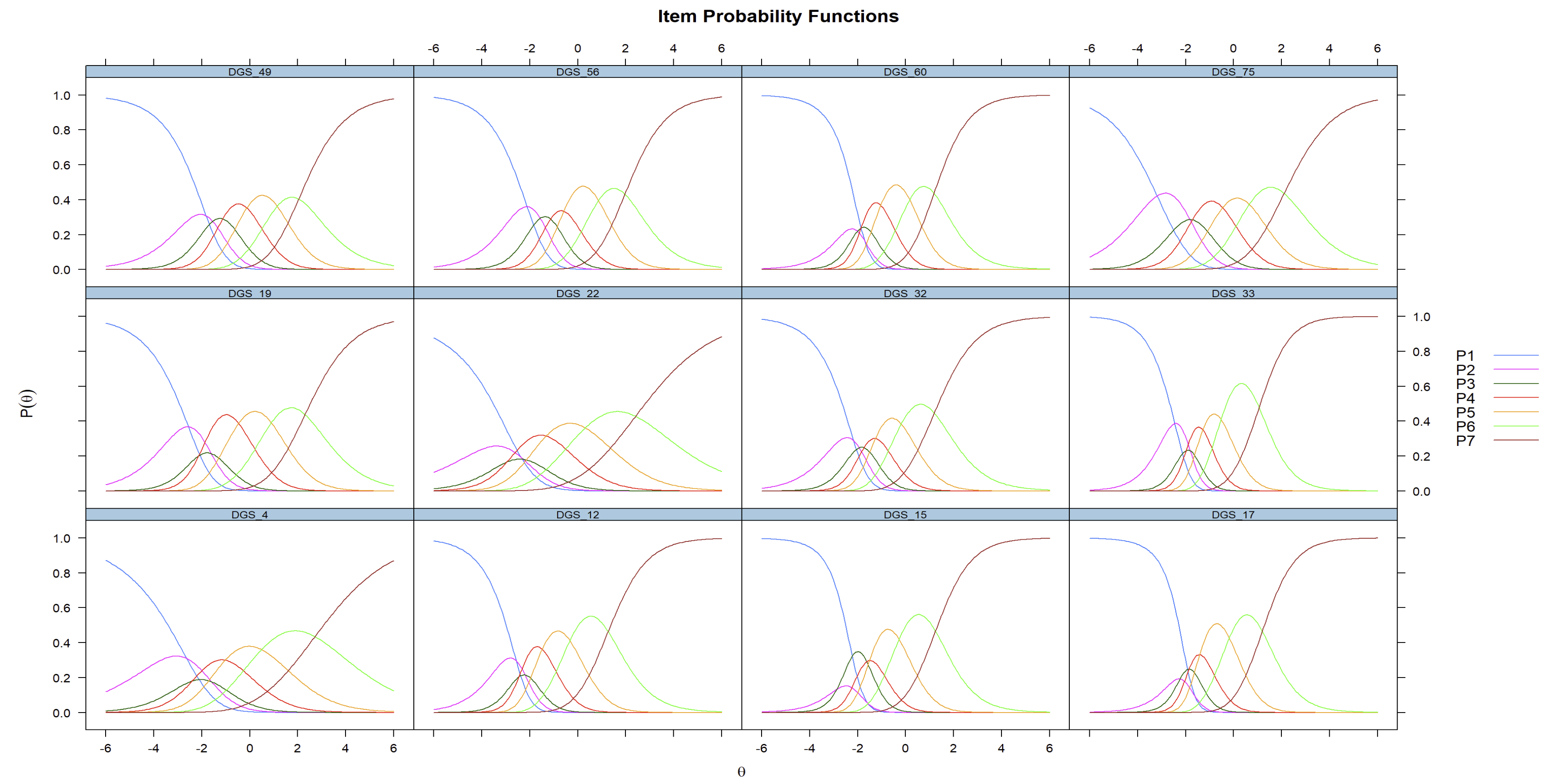
1. Factor 2 (ML7)
   1. High loading items (>0.45): 10,26,41,54,57,69,72
   2. This cluster of items are talking about **perception of reciprocal norms and seems to include 3 subcomponents: 1) reciprocal norms, 2) activists’ ideology, and 3) social norms.**
   3. Note that the items in this cluster are mostly reverse coded items.
2. Factor 3 (ML4)
   1. High loading items (>0.45): 9,27,29,46,51,52
   2. This cluster of items are talking about **empathy**.
3. Factor 4 (ML6)
   1. High loading items (>0.45): 1,21,44,65
   2. This cluster of items are talking about **attitudes toward social changes**.
4. Factor 5 (ML2)
   1. High loading items (>0.45): 20,28,35,48,68
   2. This cluster of items are talking about **attitude toward prosocial spending (relates to effective altruism).** This cluster seems to contain **affective and reflective components**.
5. Factor 6 (ML5)
   1. High loading items (>0.45): 6,38,39,45,67,70,71
   2. This cluster of items are talking about **fairness (categorical imperative).**
6. Factor 7 (ML3)
   1. High loading items (>0.45): 2,11+ 13,48,73 (was added in for better conceptual fit; they all have loadings >0.3)
   2. This cluster of items are talking about **behavioral elements relating to self-sacrifice.**

**IRT model**

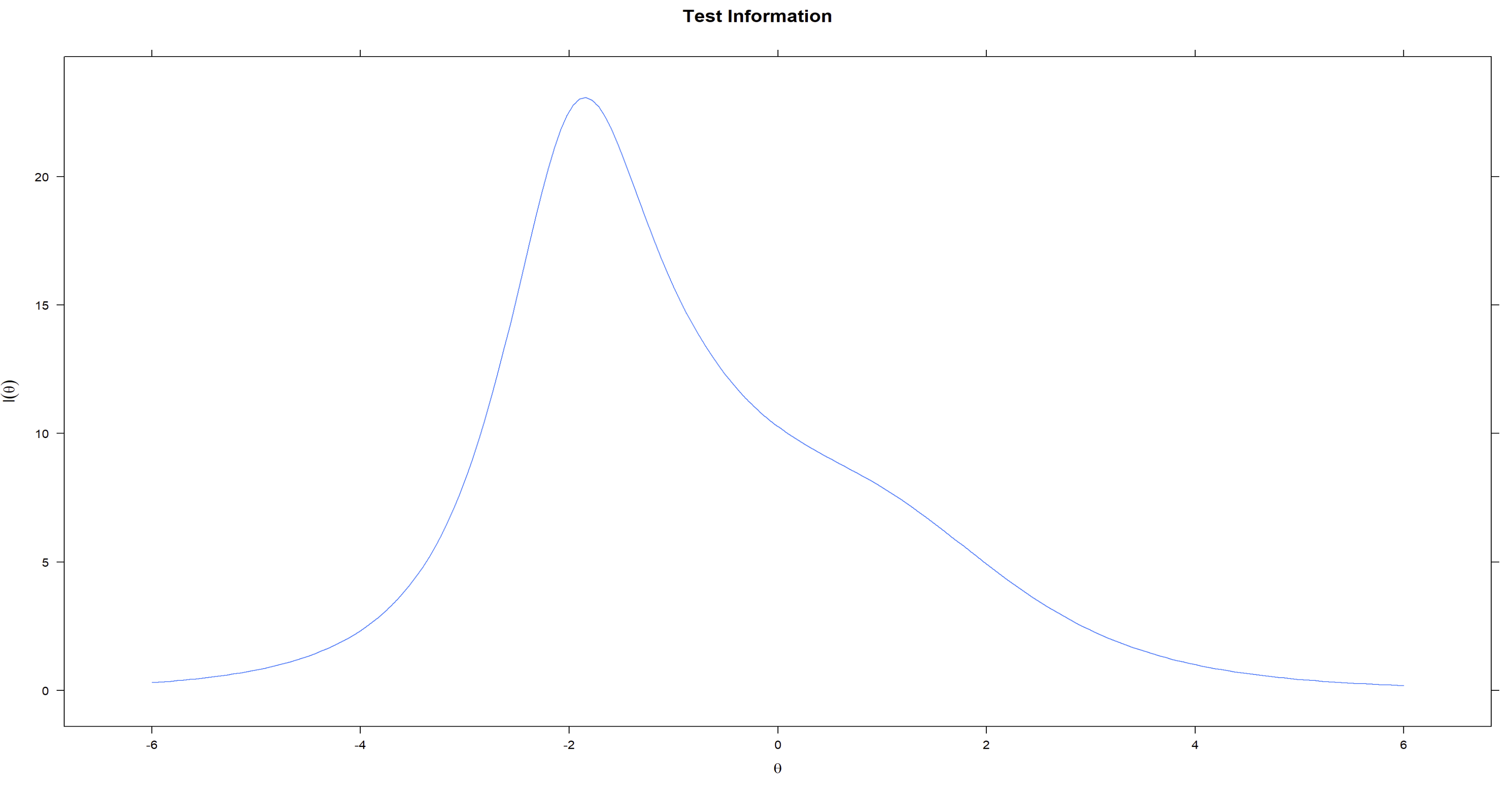
To ensure the sound psychometric property of the DGS scale, I decided to run an Item Response Theory model (IRT). Based on the data type, I went with a generalized partial credit model (GPCM). This model reveals information about 1) the chance of endorsing an option (i.e., 1-7) in one item by individuals with the latent trait; 2) The ability of an item to distinguish between people with different levels of the latent trait; 3) The range of latent trait that the item is good at demonstrating. Record one of the goals of DGS is to find people with extremely moral/altruistic ideologies. IRT allows us to examine how much clarity that the item have in discerning people with high score in the latent trait. We can think of it as evaluating the difficulty of an exam item. A good exam question for identifying the top students should have relatively high difficulty so that only people who have deep understanding of the subject matters can answer the question correctly. In our case, the ideal situation is that only people with extreme altruism can answer in certain way in the DGS. We cannot test for difficulty because there is no right or wrong answer in rating scale.

In the section below, I created a summary of the item property:

**Factor 1:**



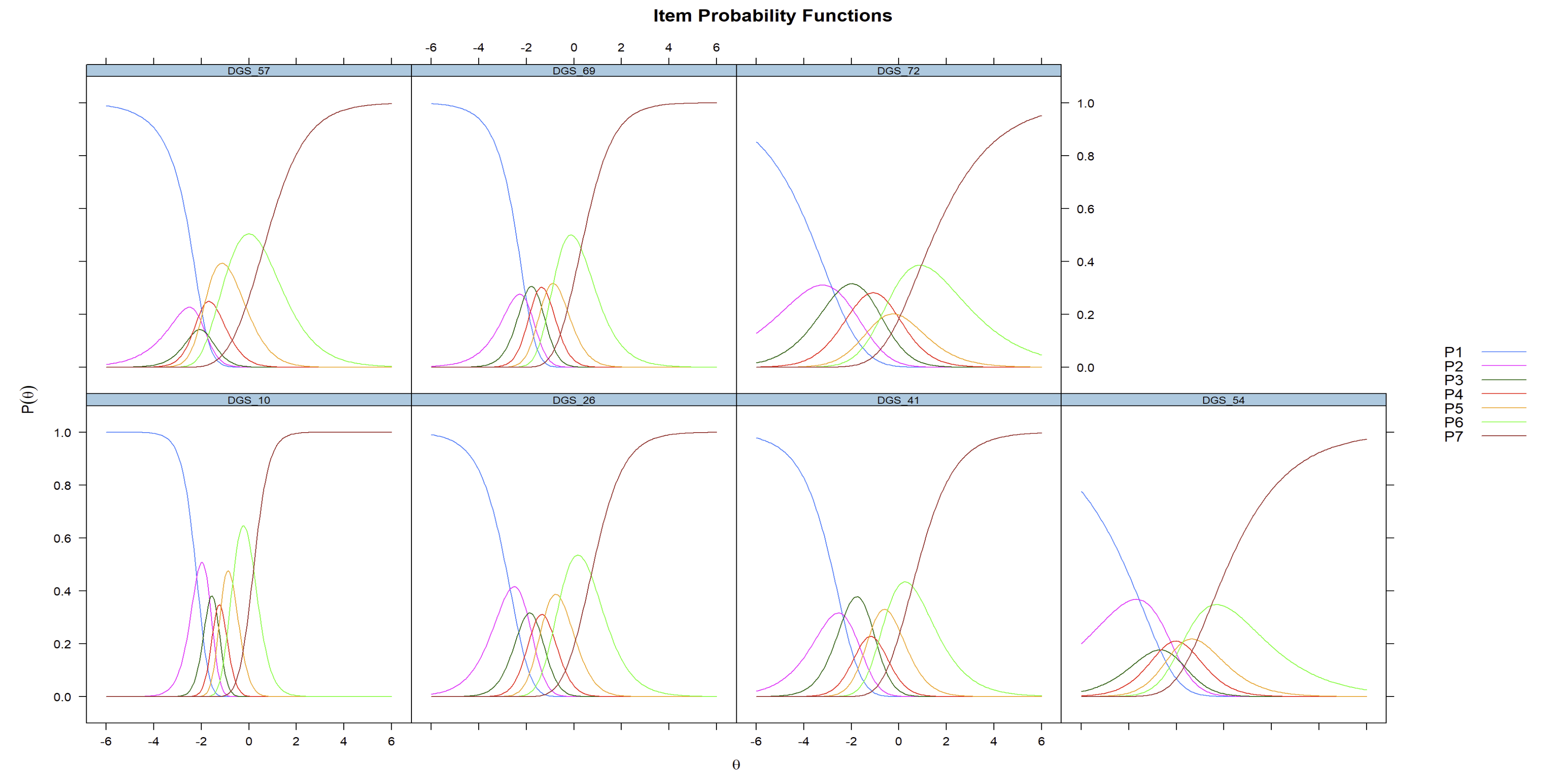
This figure reveal the “difficulty” and “discriminability” of the items for factor 1. The x axis represents the z score of participants score in the latent trait (i.e., factor). In other words, if they have x = 1, it means that they scored 1 SD above the means in latent trait. The y axis shows the likelihood of choosing specific answer. The 7 lines in each graph show the 7 different options (i.e., 1-7 scale). This graph shows the quality of each item in identifying people with different levels of the latent trait. It seems like DGS4 and DGS22 do not have good clarity in estimating latent trait from the given response. We can tell from their relatively flat curve, indicating that people with a large range of latent trait level can choose the same option. This is not what we want, so I will drop these items in the final scale.



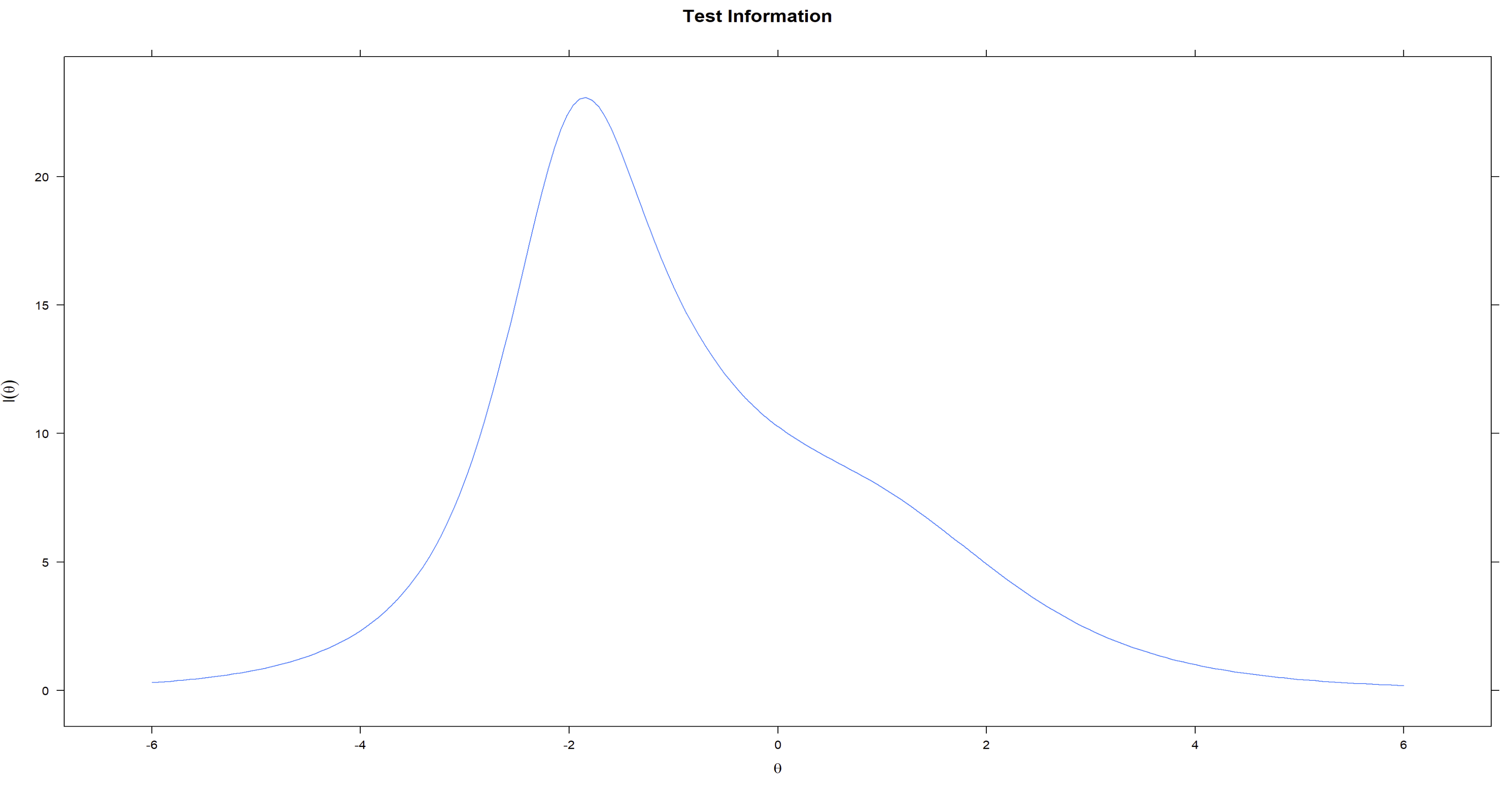
This figure is the test information graph. It shows the range of latent variables that the test is good at identifying. X and y axis are the same as above. In this graph, we can see that this subscale is best at identifying people with -2SD away from the mean in latent trait. This graph demonstrates that this subscale is good at identifying people that have very low in prosocial value (i.e. the topic of factor 1).

Conceptually, I think it makes sense as this factor contain a lot of the highly social-desirable statements (e.g. “I will listen closely to the people that I help”). However, this pattern is not what we want if the goal is to find people with extreme prosocial value. The logic is that if everyone tend to choose high (5-7) in this subscale, then there is little clarity in term of who are actually the extreme altruists. This may or may not be a major concern and we should discuss more about it.

**Factor 2:**



This figure reveals the “difficulty” and “discriminability” of the items for factor 1. The x axis represents the z score of participants score in the latent trait (i.e., factor). In other words, if they have x = 1, it means that they scored 1 SD above the means in latent trait. The y axis shows the likelihood of choosing specific answer. The 7 lines in each graph show the 7 different options (i.e., 1-7 scale). This graph shows the quality of each item in identifying people with different levels of the latent trait. It seems like DGS57 and DGS54 do not have good clarity in estimating latent trait from the given response. We can tell from their relatively flat curve, indicating that people with a large range of latent trait level can choose the same option. This is not what we want, so I will drop these items in the final scale.



This figure is the test information graph. It shows the range of latent variables that the test is good at identifying. X and y axis are the same as above. In this graph, we can see that this subscale is best at identifying people with -2SD away from the mean in latent trait. This graph demonstrates that this subscale is good at identifying people that have very low in prosocial value (i.e. the topic of factor 1).

Conceptually, I think it makes sense as this factor contain a lot of the highly social-desirable statements (e.g. “I will listen closely to the people that I help”). However, this pattern is not what we want if the goal is to find people with extreme prosocial value. The logic is that if everyone tend to choose high (5-7) in this subscale, then there is little clarity in term of who are actually the extreme altruists. This may or may not be a major concern and we should discuss more about it.

Concise version of the scale

1. DGS\_4,DGS\_12,DGS\_15,DGS\_17,DGS\_19,DGS\_22,DGS\_32,DGS\_33,DGS\_49,DGS\_56,DGS\_60,DGS\_75
2. DGS\_10,DGS\_26,DGS\_41,DGS\_54,DGS\_57,DGS\_69,DGS\_72
3. DGS\_9,DGS\_27,DGS\_29,DGS\_46,DGS\_51,DGS\_52
4. DGS\_1,DGS\_21,DGS\_44,DGS\_63,DGS\_65
5. DGS\_20,DGS\_28,DGS\_35,DGS\_48,DGS\_68
6. DGS\_6,DGS\_38,DGS\_39,DGS\_45,DGS\_67,DGS\_70,DGS\_71
7. DGS\_2,DGS\_11,DGS\_13,DGS\_48,DGS\_73