GrowthConsulting



Results of 5DGM Forced Choice Assessment Trials

PREPARED BY

Ian Dyason

Founder of Growth Consulting Asia

Creator of 5 Dimensions of the Growth Mindset

NOVEMBER 2021

Background Research

- Xiao, Y., Liu, H., Li, H., “Integration of the forced-choice questionnaire and the Likert scale: A simulation study,” May 2018, Frontiers in Psychology
- Brown, A. and Maydieu-Oliveres, A., “Item Response Modelling for Forced Choice Questionnaires,”
- Brown, A., “Personality Assessment, Forced Choice”, Dec 2015, International Encyclopedia of the Social and Behavioural Sciences, 2nd edition
- Brown, A., Bartram, D. Doing less but getting more: Improving forced-choice measures with IRT , 2009

Construct of Forced-Choice Assessment

- 40 groups of 3 statements
- Total number of statements = 120
- There are 24 statements for each of the 5 dimensions
- Each dimension will have 12 positively-stated statement and 12 negatively-stated statements
- There are 15 sets of two positive and one negative statement, 15 sets of two negative and one positive statement, 5 sets of all positive statements, and 5 sets of all negative statements
- For each group, users are asked to rate which of the statements is most like them, which is least like them, and which is in between.
- Users are required to rate the statements even if they do not 100% apply to them.
- The algorithm takes most of these inputs to calculate the output, but not all responses.
- The algorithm also uses these responses to assess consistency from within.

Research

GCA invited 1,000 individuals in its database and also online to whomever would be interested to help trial the new growth mindset algorithm. A total of 55 individuals signed up for the trial, and 31 ultimately did it. They were from Singapore, Malaysia, Vietnam, India and the US.

GCA has a current, operational assessment system (Version1) for the 5 dimensions of the growth mindset (5DGM). This is an ipsative assessment, where individuals assess themselves on a scale of 1 to 6 (1 = absolutely not like me; 6 = absolutely like me) for 40 questions. This system then plots the profile based on its own algorithm (algo1).

Ipsative assessments are known to be prone to distortion due to conscious or unconscious bias. This is overcome using forced choice assessments (see background research above).

This trial was conducted to determine if our algorithm for forced choice assessment (algo2) can create the same output as algo1, given the same input conditions. Responses were collected on the 40 sets of statements using Google Form, and the results were input into algo2 to give the 5DGM profile. The output of algo2 was created to be the same as the output for algo1. See figure 1 for the outputs of algo1 (on the left) and algo2 (on the right).

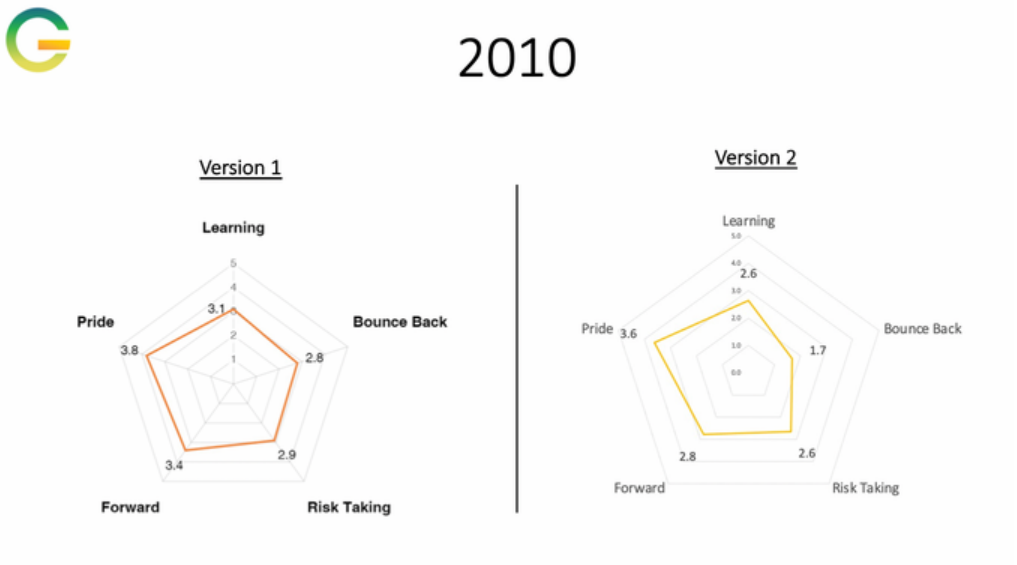


Figure 1: 5DGM profiles obtained from algo1 against algo2

Research (Con't)

The 55 volunteers were divided at random into two groups. Group 1 was asked to do Version 1 assessment (algo1) first, followed by Version 2 assessment (algo2) one week later. Group 2 did Version 2 first, followed by Version 1 later.

Volunteers were asked to answer Version 1 assessment as truthful as possible. The output from this assessment will be used as a baseline to test the accuracy and validity of the output of Version 2.

All respondents were anonymized and reports only bore their volunteer number which is known only to the test administrator.

At the end of both rounds of results taking, the data was tabulated as such for all respondents:

5DGM Trial Data

	Ipsative Assessment					Forced Choice Assessment				
	Learning	Bounce Back	Risk Taking	Forward	Pride	Learning	Bounce Back	Risk Taking	Forward	Pride

Figure 2: Sample of tabulation of data

Expected Outcomes

To show that the new algorithm is reliable and produces similar results as the first algorithm, we need to see that there are similar outputs between the two assessment systems, despite the different algorithms and constructs.

While is it not expected that the results be identical, the second algorithm will be said to be consistent in its output compared to the first if:

Condition 1

Profiles show little distortion between first and second (in other words, both profiles have similar shape)

Condition 2

Algo1 and algo2 scores that are within 15% of each other

It is also not expected that all the two profiles will be consistent. This comes down to two variables driven by the volunteer:

- Gaming – hoping to show better-than-normal result in Version 1, and
- Unconscious bias – thinking that one was truthful in responding in Version 1 but which turned out to be otherwise

These two limitations of Version 1 were the reasons to create Version 2.

Hence, the presence of one set of profiles that meets the above consistency criteria shows that the algorithm works. Needless to say, the presence of more than one set of the consistent profiles increases the validity of the Version 2 algorithm.

Process

To calculate Condition 1, we first normalized the data by dividing algo2 results with algo1 results for each dimension, and for each volunteer. A result of 0.85 to 1.15 meets Condition 2 above.

After normalization, we calculated the mean of these normalized scores for each individual, and the standard deviation of these scores from the mean.

Finally, we divided the standard deviation by the mean normalized score to obtain the Coefficient of Variation for each set of algo1 and algo2 results. The lower the score, the more consistent it is, meeting Condition 1 expectations.

Results

No. 01

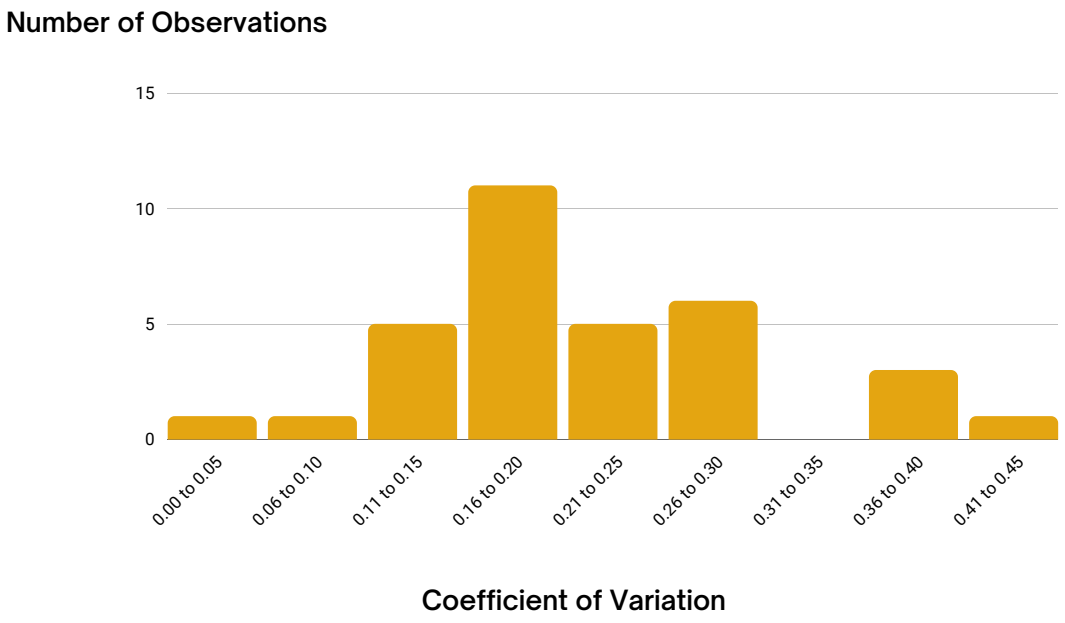
Results from algo2 average between 2.12% to 13.54% lower than algo1 across all dimensions. On average, this meets Condition 2 above (within 15%)

No. 02

47.1% (73 out of 155) of algo2 outputs met Condition 2 above.

No. 03

The Coefficient of Variation (standard deviation divided by mean) of the normalized score (algo2 result divided by algo 1 result) for all the dimensions ranged from 0.100 to 0.433. The number of observations obtained are as follows:

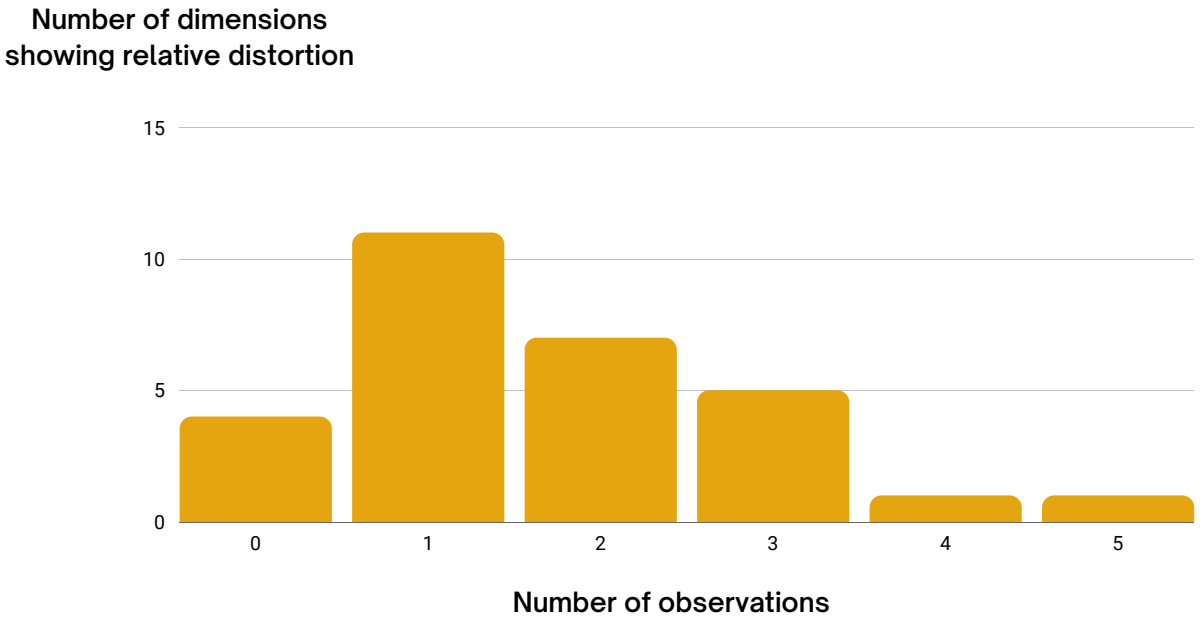


More than half (18 out of 31) of the results were 0.2 or below. A further 11 results were below 0.3. Only 4 results were above 0.3.

Results (Con't)

No. 04

There has been some relative distortion of profile from algo1 to algo2. Relative distortion refers to the change in shape of the 5DGM profile from algo1 to algo2. There is no relative distortion if the shape of the profile remains the same from algo1 to algo2, notwithstanding the amount of change. There is a distortion if any of the dimensions moves in a different direction, or in a relative amount larger or smaller than the other dimensions. The summary of relative distortion is as follows:



No. 05

There is correlation between the Coefficient of Variation and the relative distortion. All 22 pairs of profiles up to two dimensions of relative distortion had a Coefficient of Variation of 0.2 or less.

Analysis

01

Algo2 can create the same outputs as Algo1

6 sets of profiles were within Coefficient of Variation of 0.15 or less, and 4 sets had zero relative distortion. This shows that algo2 can provide the same outputs as algo1. However, there has been an average decline of up to 14% in scores between algo1 and algo2. This means that scores for algo2 for the different zones - fixed, transition and growth - will need to be moderated downwards by 10%. Hence, the growth zone will be from 3.1 onwards, transition zone will be between 2.2 to 3.0 and fixed scores will be 2.1 and below.

02

Distortion in results likely due to individual effects

While it would have been more significant if we could have more sets of profiles that meet Condition 1 and Condition 2, this could have been due to effects created by individual biases rather than the algorithm limitations. As it is widely known, ipsative assessments are prone to gaming or biases. These biases can be unconscious, even though the individual thinks he/she is answering them in a totally transparent and truthful manner. It is for this reason that the background research had identified that forced choice assessments are more valid and reliable than ipsative.

03

Language proficiency is important

It was also seen that non-native English-speaking volunteers had a larger relative distortion than native English speakers. This could be due to nuances in understanding in words used, even as care was taken to ensure that the statements used were unequivocal and simple. Now that the algorithm has been shown to deliver the right results, there is an opportunity to simplify the statements further without impact on the results. However, for cultural inclusivity, the only way is to translate the statements to the native language.

Conclusion



That the forced-choice construct was shown by prior research to provide unbiased, valid and reliable results makes Version2 a more robust system for assessing a person's growth mindset.

It is important to note that prior research on the validity of the 5 dimensions on predicting the growth mindset had been conducted, and hence this trial did not replicate that. What this trial did was to ensure that statements used and the algorithm deployed would yield consistent results between Version 1 and Version 2 assessments, and that has been achieved. However, more data would need to be gathered to ensure that consistency can be further achieved, with fewer relative distortion and smaller Coefficient of Variation. More work also needs to be done to ensure that congruence of results is achieved for other languages.

The trial was a success because it has shown that the results of algo2 can be consistent with algo1.