JVR

PSYCHOMETRICS

SEED TECHNICAL MANUAL

Prepared by
Brett Gregory & Cobi Hayes

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

The Self Education Employability Device (Beukes, 2010), hereafter referred to as the SEED, is an assessment that aims to assist individuals in exploring and developing their employability skills. The tool was developed as a non-psychometric assessment to facilitate the identification and development of career skills.

USER QUALIFICATIONS

In line with the Health Professions Act (No. 56 of 1974), only registered psychology professionals are allowed to use measures of psychological constructs. Given that the SEED was developed to be a non-psychometric skills assessment, its use is not restricted.

APPROPRIATE USE

The SEED can be used for:

- Career counselling
- Career guidance
- Leadership development
- Personal development

ADMINISTRATION

The SEED can be administered either in paper-and-pencil format or online. The paper-and-pencil version of the assessment requires individuals to manually sum their question and scale totals, and compute percentages for the scales. The online version assists in automating this scoring procedure.

TRAINING

Currently, there is no certified training for the SEED. Support is available if needed by contacting the test developer directly through the contact page on www.employability.co.za.

METHODOLOGY

RESEARCH DESIGN

A cross-sectional research design was employed whereby all participants in the sample were assessed within a short period. This design allowed for multiple characteristics to be collected from participants (biographical and test-related information), while being highly cost-effective. Moreover, the study did not directly involve the manipulation of any variables in the study.

SAMPLE

Non-probability sampling, specifically convenience sampling, was used to collect data on the SEED for the current study. This sampling strategy does mean that there was a subjective component involved in recruiting participants, and thus there is a potential for some sampling bias. However, the method did allow for data to be collected from participants for whom the test was designed.

INSTRUMENTS

SELF EDUCATION EMPLOYABILITY DEVICE (SEED)

The SEED (Beukes, 2010) is a 60-item self-exploration instrument aimed at assisting individuals in exploring their employability skills. Each item in the assessment represents an employability skill. When clustered, the items provide feedback on 12 scales relating to employability, these being: (a) Attitude; (b) Service; (c) Creativity; (d) Barriers; (e) Foundation; (f) Core; (g) Sectors; (h) Career; (i) Work; (j) Goals; (k) Network; and (l) Transitions.

PROCEDURE

JVR Psychometrics was provided with participants' biographical and SEED data, which was then cleaned and collated for further statistical analysis. The dataset was analysed using both exploratory and inferential techniques to establish the SEED's psychometric properties.

ETHICAL CONSIDERATIONS

Standard ethical principles and guidelines were adhered to throughout the present research. Of importance, participants were briefed on what the present study entailed and what was expected of them should they volunteer to participate in the research. The data that was provided to JVR Psychometrics was stored within a secure data repository and only made available to those directly involved in the study. Moreover, all sensitive participant information was transformed into a scrambled version of itself through a process known as hashing. This was done to anonymise the dataset.

ANALYSIS

The main objectives of our analyses were to:

a) Explore the biographical and test-related variables;
b) Assess the underlying factor structure of the SEED through exploratory and confirmatory factor analysis;
c) Estimate internal consistency coefficients for the SEED scales;
d) Investigate whether subsets of participants grouped accordingly to their demographic variables obtained significantly different results on the SEED;
e) Perform Rasch analysis and interpret the results of item fit and differential item functioning analyses.

We should note that all *p*-values that are reported in the RESULTS section were adjusted to account for the familywise error rate (FWER). Specifically, we employed the k-FWER procedures using Holm's (1979) method as suggested by Wilcox (2017, p. 363).

RESULTS

DESCRIPTIVE STATISTICS

The section that follows provides an exploratory overview of the data involved in this research. To this end, we report on general descriptive statistics computed to describe both the set of demographic variables and participants' average performance on the various SEED scales.

In the first subsection titled SAMPLE OVERVIEW, we present the number and proportion of respondents in the sample that share similar demographic classifications. This is done using frequencies and percentages. In the SCALE DESCRIPTIVE STATISTICS subsection, we examine the central tendency for each of the twelve scales and the total score, in addition to providing the minimum, maximum and range of scores. Our goal here was to provide some general measure of location that reflects what a typical scale variable in the dataset looks like for the given sample. We decided to avoid standard arithmetic means to ensure that no small proportion of outliers in the tails of the distributions would severely influence our estimates (Wilcox, 2017). A more appropriate approach would be to utilise trimmed means in conjunction with trimmed standard deviations, both of which should facilitate eliminating the possibility of outliers severely shifting any mean estimations. We felt that a visual presentation of the SEEDs scales would also prove useful, and thus we do also present violin plots as they are capable of conveying a large amount of useful information. Lastly, we provide Pearson-moment correlations between the different scales and dimensions along with a heatmap of the correlations.

SAMPLE OVERVIEW

The sample for this research project consisted of 483 participants, each of whom completed the SEED. An overview of the primary demographics of the sample is provided in Table 1. It is clear from the results that representation across the different demographic groups was not entirely equal. What this means from a practical perspective is that some inferential analyses which involve grouping participants according to participants' demographics would be limited since certain groups may be deemed as being too small to successfully perform the computation procedures. At times this limitation was overcome by collapsing specific groups into a single category as will be seen later in the GROUP DIFFERENCES section of the report.

Table 1. Sample Overview

Group	*N*	%
Gender		
Female	381	78.88
Male	102	21.12
	483	**100**
Age		
17 Years or Younger	12	2.53
18-21 Years	194	40.84
22-25 Years	144	30.32
26-29 Years	68	14.32
30-33 Years	36	7.58
34 Years and Older	29	6.11
	483	**100**
Ethnicity		
Black African	439	90.89
Caucasian (White)	14	2.90
Indian	2	.41
Mixed Ethnic Origin	27	5.59
Other	1	.21
	483	**100**
Education Level		
NQF 4	332	68.74
NQF 5	18	3.75
NQF 6	16	3.31
NQF 7	12	2.48
NQF 9	5	1.04
Other	100	20.70
	483	**100**
Province		
Free State	21	4.35
Gauteng	18	3.73
KwaZulu-Natal	108	22.36
Mpumalanga	336	69.57
	483	**100**

Note. N: Frequency of participants belonging to the demographic classification; *%*: Proportion of the sample in respect to the demographic classification.

SCALE DESCRIPTIVE STATISTICS

We decided to report on the scale descriptives in two different ways. Our first approach was to present the trimmed means and standard deviations for the entire samples overall scale scores. This overview is provided in Table 2.

Table 2. Scale Descriptive Statistics

Scale	$\bar{X}_{trim}$	σ_{trim}	Min	Max	$Range$
Total	439.17	80.03	238	700	462
Attitude	30.45	5.72	12	47	35
Service	21.93	5.26	10	36	26
Creativity	37.74	7.88	20	58	38
Barriers	43.49	9.65	18	72	54
Foundation	32.17	6.66	16	48	32
Core	91.79	16.99	47	136	89
Sectors	25.15	6.96	12	48	36
Branding	28.44	6.85	14	48	34
Work	27.42	6.91	13	48	35
Goals	43.08	10.06	18	72	54
Network	28.31	6.89	13	48	35
Transitions	28.48	6.60	14	48	34

Note. $\bar{X}_{trim}$: 10% trimmed mean; σ_{trim}: 10% trimmed standard deviation; Min: Minimum score; Max: Maximum score; $Range$: Range of scores (i.e. $Max - Min$).

Our second approach was to provide a visual summary of the overall distribution of the scale data and its probability density. This was done using violin plots, which is a combination of a general box plot and a density plot. These are presented in Figure 1. We should point out that we decided to convert participants' scores into percentages in our graph. Since the scales in the SEED do not all have the same total score, we decided to rather rely on scale percentages as it would allow the scales to be more easily comparable to one another.

The first thing that we can observe in Figure 1 is the overall distribution of scale scores. The general observable pattern is that the peak of the distribution is lower than 50%. This indicates that a large proportion of the sample obtained a scale score that was lower than half of the total obtainable score for a given scale. This pattern remains prevalent, for the most part, as the sample is partitioned according to Gender and Age. The median for each scale, indicated by the white dot on the spine of each violin, further corroborates this pattern since it appears that it generally hovers under 50%.

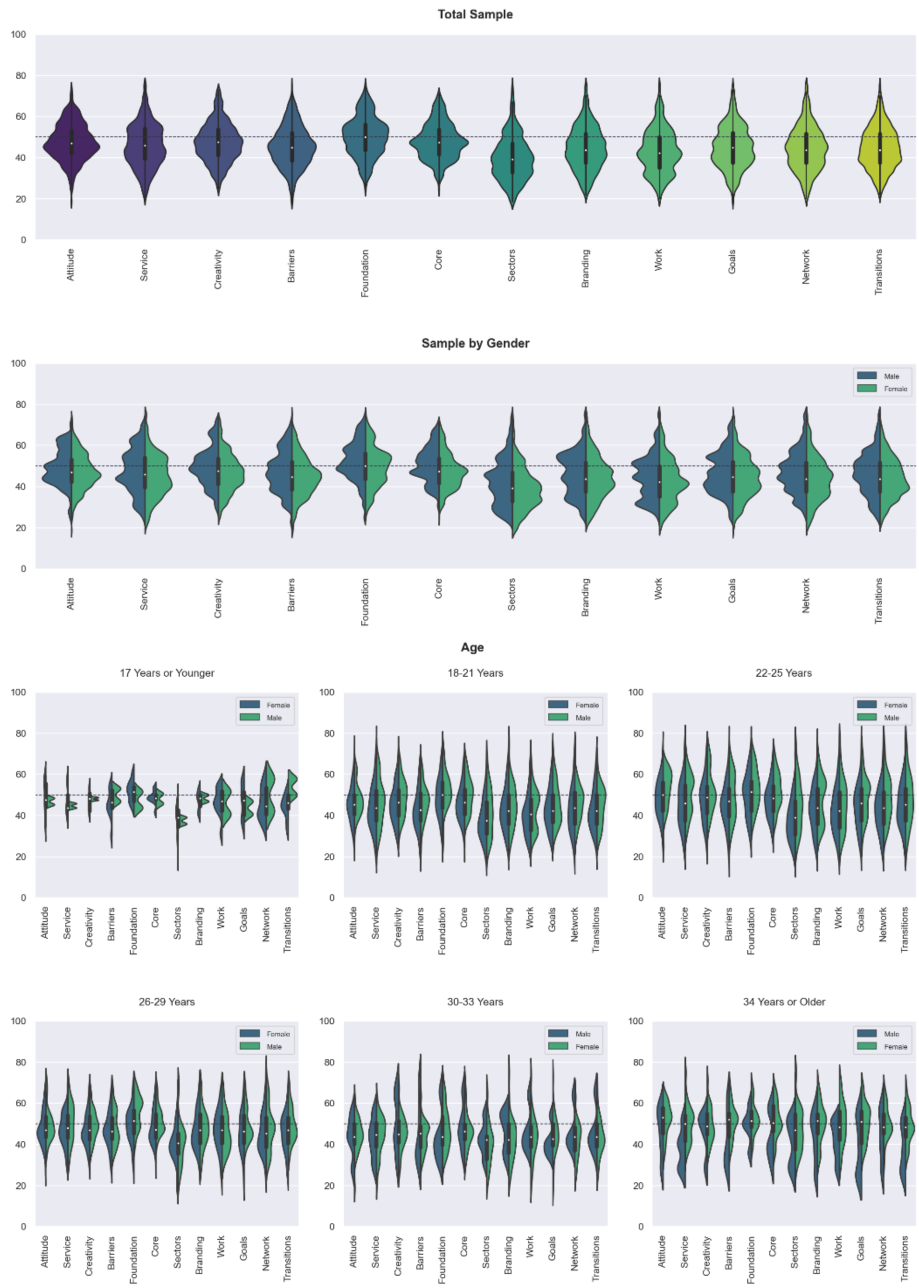


Figure 1. Violin plots for the SEED scales.

SCALE CORRELATIONS

We computed Pearson product-moment correlations between the different SEED scales. These correlations are presented in Table 3 and visually presented in the form of a heatmap in Figure 2. Correlations help us to understand how variables linearly relate to one another. Moreover, many of the statistical procedures used throughout this report rely on certain correlational assumptions not being violated.

From the correlation table, one can observe that the correlations among the variables are all positive and fairly strong. The strongest relationship was .80 between Goals and Branding. The weakest relation was found to be .47 between Sectors and Attitude.

Table 3. Pearson product-moment correlation matrix

	Attitude	Service	Creativity	Barriers	Foundation	Core	Sectors	Branding	Work	Goals	Network	Transitions
Attitude	1.00											
Service	.66	1.00										
Creativity	.70	.70	1.00									
Barriers	.60	.64	.73	1.00								
Foundation	.54	.51	.65	.68	1.00							
Core	.63	.62	.73	.76	.78	1.00						
Sectors	.47	.54	.58	.65	.55	.71	1.00					
Branding	.55	.58	.64	.71	.63	.78	.77	1.00				
Work	.50	.50	.56	.64	.54	.71	.71	.76	1.00			
Goals	.55	.54	.60	.68	.62	.75	.72	.80	.74	1.00		
Network	.50	.51	.60	.65	.56	.70	.69	.77	.72	.76	1.00	
Transitions	.56	.54	.62	.67	.60	.72	.67	.72	.70	.77	.78	1.00

Figure 2. Heatmap of Pearson product-moment correlation matrix.

DIMENSION CORRELATIONS

Each item in the SEED is given an item total score. These totals are computed by summing responses given for that item across three dimensions. Table 4 presents the results of the tests of associations that were done to assess that the three dimensions were linearly related to each other. All tests were significant, meaning that there is indeed some degree of relationship across the item dimensions.

Table 4. Test of associations between three dimensions

	r	*Confidence Intervals*		*T*	*df*	*p*
Knowledge-Importance	.56	.55	.57	114.97	28978	<.001
Knowledge-Experience	.64	.63	.64	140.56	28978	<.001
Importance-Experience	.49	.48	.50	95.93	28978	<.001

Note. r: Pearson-moment correlation; *T*: *T-statistic*; *df*: degrees of freedom; *p*: *p*-value.

Additionally, we constructed three heatmaps that depict the correlations for the item responses according to each dimension. We decided against including the actual correlation matrices given their large dimensions. What is evident from the figures is that the diagonal of the matrix is a lighter colour compared to the rest of the matrix elements. This means that correlations between the same item across two dimensions had a much stronger linear relationship than with different items. We also can infer that participants would generally respond the same across the three dimensions rather than responding differently.

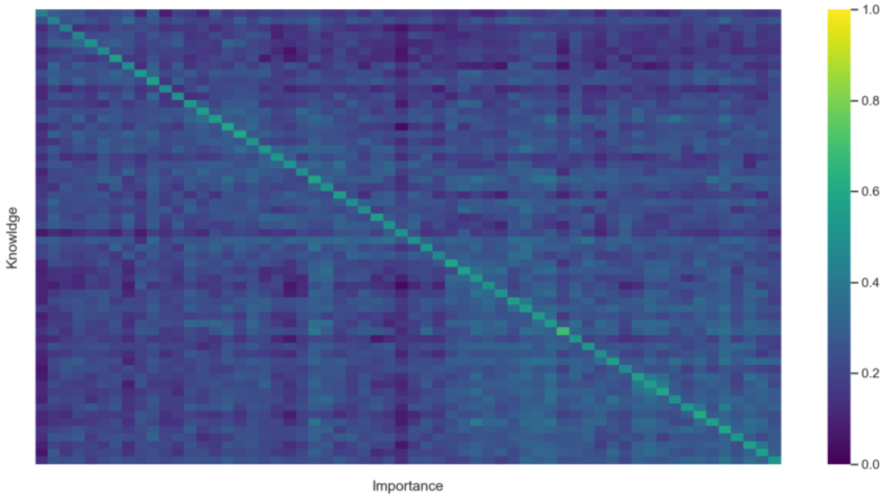


Figure 3. Heatmap of Pearson-moment correlations between items across Knowledge and Importance dimensions.

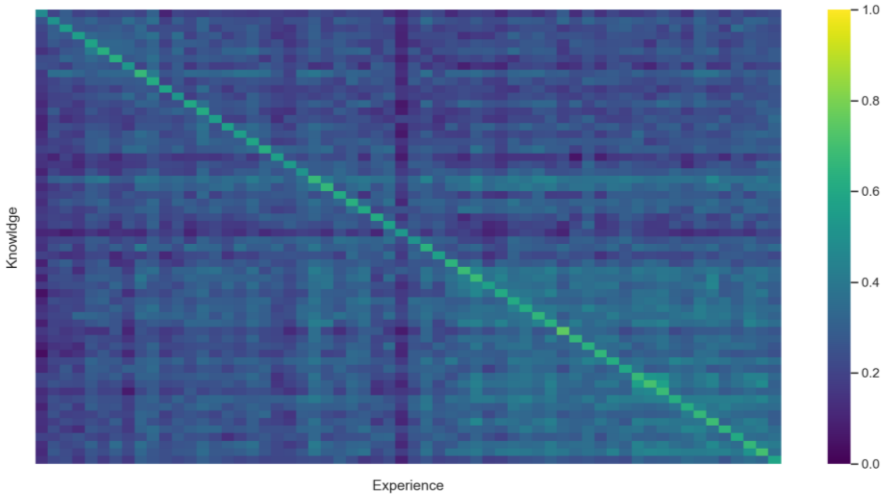


Figure 4. Heatmap of Pearson-moment correlations between items across Knowledge and Experience dimensions.

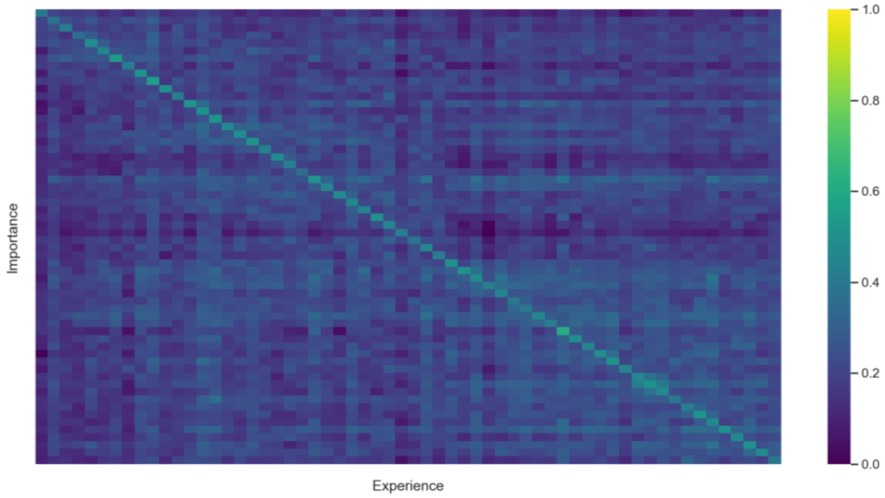


Figure 5. Heatmap of Pearson-moment correlations between items across Importance and Experience dimensions.

RELIABILITY

Internal consistency of the SEED was evaluated using both Cronbach's (1951) coefficient alpha (α) and McDonald's (1999) coefficient omega (ω). We checked two primary assumptions before estimating internal consistency coefficients, these being the assumption of tau-equivalence and the assumption that no outlying observations exist in the data. Violations in these two assumptions are known to greatly impact alpha and omega estimates (Zhang & Yuan, 2016). We found that six of the twelve scales violated the assumption of tau-equivalence and there were multiple outlying observations among the variables.

To ensure that our estimations were as accurate as possible, we employed Zhang and Yuan's (2016) method where they propose computing robust *M*-estimators of alpha and omega, which partly involves down-weighting outlying observations during the estimation process. The internal consistency results for the SEED are reported in Table 5 below.

Table 5. Internal consistency estimates

Scale	N_{items}	$\hat{\alpha}$	SE	$\hat{\omega}$	SE
Attitude	4	.77	.02	.77	.02
Service	3	.75	.02	.75	.02
Creativity	5	.78	.02	.79	.02
Barriers	6	.85	.01	.85	.01
Foundation	4	.80	.02	.80	.02
Core	12	.91	.01	.91	.01
Sectors	4	.86	.01	.86	.01
Branding	4	.84	.01	.84	.01
Work	4	.78	.02	.79	.02
Goals	6	.89	.01	.89	.01
Network	4	.84	.01	.84	.01
Transitions	4	.82	.01	.82	.01

Note. N_{items}: Number of items in the scale; $\hat{\alpha}$: Estimated Cronbach's alpha coefficient; $\hat{\omega}$: Estimated McDonald's omega coefficient; SE: Standard Error.

The estimated alpha and omega coefficients were on average found to be .82 and .83 respectively. If we assume that internal consistency estimates above .7 are acceptable, then no scales have low internal consistency metrics.

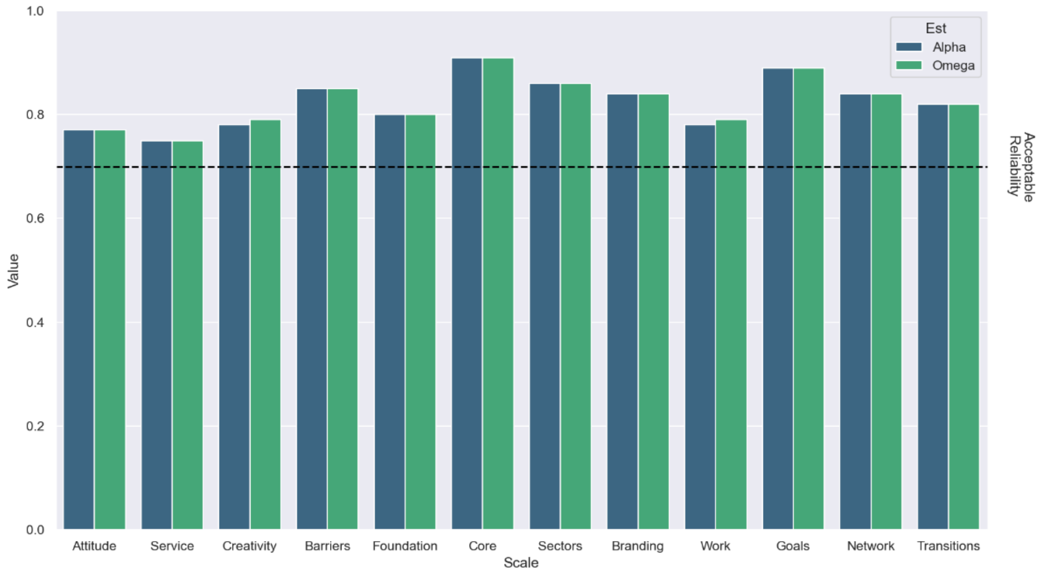


Figure 6. Reliability of SEED scales presented as a bar plot.

Reliability for the SEED has previously been established by Taylor and Beukes (2019) on a sample of 68 individuals, however, the SEED has undergone slight changes in its item mappings since then. With that being said, these changes were not drastically different, and we thus could examine the extent to which the internal consistency estimates reported here are similar to those previously found.

Our estimates were higher than those computed by Taylor and Beukes (2019). To justify this, we computed the arithmetic mean of the absolute value of all possible differences between the estimates given by Taylor and Beukes (2019) and the ones presented in this report. We did this by calculating the mean absolute difference Δ defined as

$$\Delta = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i - y_j\right|,$$

where $x_i \in X$ and $y_j \in Y$ are the estimates of either Cronbach Alpha $\hat{\alpha}$ or McDonald's Omega $\hat{\omega}$ presented by Taylor and Beukes (2019) and the ones provided in this report respectively. We estimated that the mean difference in estimates was $\Delta_1 = .099$ and $\Delta_2 = .094$ for $\hat{\alpha}$ and $\hat{\omega}$ respectively. This reveals the approximate distance was a .10 difference between previous reliability estimates and the ones presented here. Further examination revealed that the difference was heavily one-sided, in that the new reliability coefficients tended to be higher.

FACTOR ANALYSES

We performed both an exploratory and confirmatory factor analysis, abbreviated to EFA and CFA respectively, to assess the underlying factor structure of the SEED. The results of these are presented in the sections that follow.

EVALUATION FRAMEWORK

Judging the overall model-data fit in factor analysis is not straightforward. Broadly speaking, researchers generally review the results of inferential tests of exact fit or assess a set of goodness-of-fit indices based on a collection of acceptable cut-off ranges specific to those individual indices. In the former case, the most popular approach is to evaluate the *p*-value that is obtained by comparing the normal theory likelihood ratio test statistic T_{ML} to the χ^2 distribution to determine whether the structures from the observed data are reproduced from the model implied structures, signalling a good *exact* fit (i.e. $p > .05$) (McNeish, 2020). In the latter case, the goodness-of-fit indices are examined to determine whether they are located within some predetermined range which signals a good *approximate* data-model fit.

The supporters of either approach offer their own evidence for why their methodology should be the preferred way of judging the fit of a model, but no one approach is agreed to be better than the other in a general sense. Arguably, however, both approaches suffer from a similar problem in that they can lead to an increased risk of making a Type I error when the underlying sample size is small to moderate, which, from a practical point of view, means that good-fitting models can potentially be incorrectly rejected on the grounds of seemingly poor-fitting indices (McNeish, 2020). Why this happens is primarily due to the tendency of T_{ML} to become inflated when sample sizes are not sufficiently large, which can also impact goodness-of-fit indices since their computations often make use of T_{ML} (McNeish, 2020). This inflation of T_{ML} can also be a result of non-normal distributions among the observed variables, although this can be overcome to an extent by using robust alternatives in the estimation of T_{ML} or applying post hoc corrective procedures (Li, 2016).

Given the above, we felt we needed to evaluate whether this study's sample size of 483 is classified as being small/moderate relative to the theoretical model specification. We adopted Herzog and Boomsma's (2009) definition which defines a small sample as being one in which the $N{:}df$ ratio is smaller than three. We computed that this ratio was .27, leading us to believe that the sample size is small and that the aforementioned issues relating to T_{ML} may arise, although this cannot be guaranteed.

We decided that our interpretation would still incorporate the traditional methods of factor analysis interpretation but that we would also supplement our analysis with two alternative methods of interpretation in the hope to refine our evaluation of the SEED model. The first method, proposed by McNeish (2020), involves comparing the test statistic with a more appropriate theoretical distribution, that being the *F*-distribution, as this can lead to more desirable performance with smaller samples. The second method involves applying a multiplicative post hoc corrective procedure, specifically a Swain

(1975) correction, to reduce T_{ML} (McNeish, 2020). This correction reduces the estimated test statistic so that it follows the χ^2 distribution more reliably (McNeish, 2020). Importantly, we only applied this post hoc corrective procedure in our EFA since we used the standard maximum likelihood estimator as no robust alternatives are available in the statistical packages. For the CFAs, a robust maximum likelihood estimator was used, and thus we argued that no post hoc corrections would be required given that the point of the robust variant is to reduce test statistics that may be inflated.

Since our interpretation still involved examining the goodness-of-fit indices, we should note that we adhered to the traditional guidelines (Cangur & Ercun, 2015; Hu & Bentler, 1999) in determining good model fit:

- Comparative Fit Index (CFI): A value of above .95 is generally considered a good fit.
- Tucker-Lewis Index (TLI): Greater values for TLI generally indicate a better fit for the model. Values above .95 are considered to be acceptable by most researchers.
- Standardised Root Mean Squared Residual (SRMR): Although values below .08 are considered to be acceptable, only values below .05 are an indication of a good fit.
- Root Mean Square Error of Approximation (RMSEA): It is preferred if models have RMSEA values below .05, though values between .05-.08 can also be said to be a good fit. Values that fall within the range of .08-.10 are neither good nor bad.

EXPLORATORY FACTOR ANALYSIS

We began by assessing the correlations between the SEED item totals. Our objective was to determine that there were indeed relationships between the variables (see SCALE CORRELATIONS section) and that our correlation matrix was not computationally similar to the identity matrix of similar size. We tested this hypothesis using Bartlett's (1951) test. The results indicated that the correlation matrix was not computationally similar to the identity matrix, $\chi^2(1770) = 18\,133.68, p < .001$.

We also evaluated the sampling adequacy using the Kaiser-Meyer-Olkin measure (KMO; Kaiser, 1970). The KMO statistic was estimated to be .98 which is sufficiently large (Hutcheson & Sofroniou, 1999). Lastly, we evaluated multicollinearity by determining whether the determinant of the correlation matrix was smaller than 1×10^{-6} (Field, 2013). We found that the determinant was smaller than this cut-off point. This indicated that correlations among our item-level variables were highly linearly related and did mean that computational difficulties could arise when using them in our analyses.

Given that we were exploring the underlying factor structure of the SEED, we were interested in estimating the number of factors to extract. To this end, we constructed a scree plot that included the outputs of a parallel analysis. The plot can be seen in Figure 7 below. The two classical decision rules (i.e. Kaiser's (1960) rule and parallel analysis) suggest retaining seven and three factors respectively.

Given these results, we decided to perform two separate models retaining the specified factors. In addition, we examined a third model whereby the theoretical 12 factors were retained.

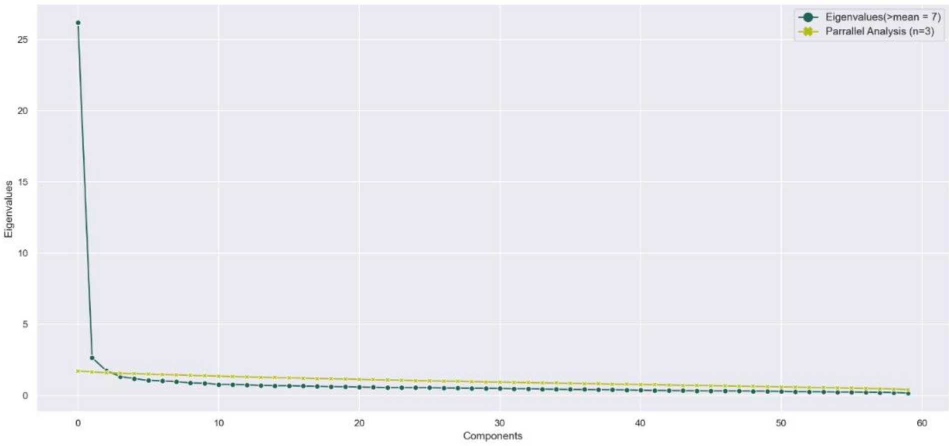


Figure 7. Scree plot and parallel analysis

We performed each factor analysis using maximum likelihood and applied an oblique rotation (*direct oblimin*) to help discriminate between the factors. All models converged successfully. The results of the EFAs are reported in Table 6 below.

Table 6. Fit indices for the EFA models

Model	$N_{Factors}$	χ^2	*df*	*p*	p_S	p_F	TLI	RMSEA	Lower$_{RMSEA}$	Upper$_{RMSEA}$
EFA_1	3	2998.20	1770	<.001	<.001	<.001	.90	.04	.04	.05
EFA_2	7	2016.25	1770	<.001	<.001	<.05	.95	.03	.03	.03
EFA_3	12	1370.07	1770	<.001	1.00	.99	.97	.02	.02	.03

Note. $N_{Factors}$: Number of factors extracted; χ^2: Chi-Squared statistic; *df*: degrees of freedom; *p*: *p*-value; p_S: Swain (1975) corrected *p*-value; p_F: *p*-value obtained by comparing $\hat{T}_{ML}$ to the *F*-distribution; TLI: Tucker-Lewis Index; RMSEA: Root Mean Square Error of Approximation.

All models were found to be significant ($p < .001$) when we compared the estimated maximum likelihood test statistic $\hat{T}_{ML}$ to the χ^2 distribution, suggesting that the models have a poor exact fit. However, the *p*-values obtained by applying a Swain (1975) correction and by comparing $\hat{T}_{ML}$ to the *F*-distribution suggest that the third model EFA_3 has satisfactory *p*-values. Interestingly, if one were to apply *p*-value corrective procedures to the set p_F-values, the EFA_2 is no longer is significant, and thus may be considered to have a satisfactory exact fit.

Another interesting observation to take note of is how the goodness-of-fit indices seem to improve as more factors are retained in the models. Moreover, all goodness-of-fit indices were within the acceptable cut-off ranges, except for the TLI of EFA_1 which was below .95.

The factor loadings for EFA_3 were also examined to observe if the variables were loading correctly onto their respective scales. These can be viewed in Table 20 in the APPENDIX. It can be seen that many items did not load correctly onto their respective scales and there was significant cross-loading among the items.

Overall, the EFAs performed seemed to support the existence of the theoretical structure underpinning the SEED. The most interesting observation is that the EFA models tended to improve as the number of factors retained in the model increased closer to the theoretical twelve factors. The results indicate that a twelve-factor structure is viable but given that a seven-factor model has arguably satisfactory fit indices, we felt that both models should be investigated further with CFA.

CONFIRMATORY FACTOR ANALYSIS

The *a priori* structural relationships that exist between the different SEED variables were investigated using confirmatory factor analysis. We checked all primary parametric assumptions before running the analyses, with a specific focus on assessing heteroskedasticity. Heteroskedasticity was evaluated using the Breusch Pagan test (1979), which indicated that variance was not equal across our parameters. Furthermore, some observed variables in the model followed a non-normal distribution.

We thus opted to use robust maximum likelihood (MLR) to estimate the unknowns in our model. We felt maximum likelihood would be appropriate given that our observed variables followed a continuous and, for the most part, a multivariate normal distribution (Li, 2016). However, due to the slight deviations in normality for some of the variables in the model, we chose the robust variant of maximum likelihood as it has shown to provide more accurate estimations under such conditions (Li, 2016). Given our choice of estimator, we did not apply any multiplicative post hoc corrections to our test statistic. We do, however, compare the test statistic to the *F*-distribution and evaluate the corresponding *p*-value as an alternative method for judging the exact fit of the models.

The first model we examined was the seven-factor model that was identified with our EFA. We constructed the model according to the item loadings. The second model we examined was the prescribed SEED factor structure. The theoretical model underpinning the SEED is hierarchical whereby the total SEED score acts as a latent variable being manifested by the twelve scales, which in turn are manifested by the scale's respective items. The results of the analysis including the goodness-of-fit indices are presented in Table 7 below. We have also provided path diagrams (items omitted) in Appendix B: Path Diagrams for CFA Models to aid in visualising the CFA models under investigation.

Table 7. Fit indices for item-level CFA models

Model	χ^2	*df*	*p*	p_F	TLI_r	CFI_r	$SRMR_r$	$RMSEA_r$	$Lower_{RMSEA}$	$Upper_{RMSEA}$
CFA_1	2294.73	1317	<.001	<.001	.91	.91	.05	.04	.04	.05
CFA_2	3271.85	1698	<.001	<.001	.88	.88	.05	.05	.05	.05

Note. χ^2: Chi-Squared statistic; *df*: degrees of freedom; *p*: *p*-value; p_F: *p*-value obtained by comparing $\hat{T}_{MLR}$ to the *F*-distribution; TLI_r: Robust Tucker-Lewis Index; CFI_r: Robust Comparative Fit Index; $SRMR_r$: Robust Standardised Root Mean Squared Residual; $RMSEA_r$: Robust Root Mean Square Error of Approximation.

The results suggest that both models had a poor fit. The first indication of this is the significant *p*-values obtained by comparing the test statistic to the χ^2 distribution and *F*-distribution, respectively. Moreover, both models had TLI_r and CFI_r values below the acceptable ranges. Even though the values for $SRMR_r$ and $RMSEA_r$ suggested a good approximate fit, we are hesitant to conclude off the basis of these two values alone in relation to the other metrics that the model has a good fit overall.

CONCLUDING REMARKS

Although the results of the EFA suggested that the underlying hypothesised model for the SEED is present, the factor loadings obtained in the EFA and the CFA results appear to show that there is a moderate to a large amount of model misspecification. This misspecification could also be the reason why we see a high number of items incorrectly loading or cross-loading onto other scales when performing an EFA. It is also why the overall fit of the model in the CFA was poor. Based on the current model specification, we were able to identify the following changes that could be made that might help in improving the fit of the model: (a) specifying item 55 belonging to the Network scale to instead load onto the Sectors scale; (b) specifying item 8 belonging to the Creativity scale to instead load onto the Sectors scale; and (c) specifying item 50 belonging to the Goals scale to instead load onto the Creativity scale. These adaptations are only suggestions and have not been thoroughly tested in the current analysis. An alternative adaptation could be to revisit the question dimensionality as this does add a layer of complexity to the model.

GROUP DIFFERENCES

One of our objectives was to explore whether participants grouped according to different demographic classifiers obtained significantly different SEED scores. We were primarily concerned with testing for differences between gender and age groups. We decided against testing for differences for ethnicity on the basis that group sizes would not have been sufficiently large enough for our analyses.

We performed two separate *multivariate analysis of variance* tests, hereafter referred to as MANOVAs, to evaluate whether participants differed in their scale scores. A MANOVA extends the general ANOVA to situations involving two or more measures and thus is appropriate given that participants were scored according to twelve different scales (Wilcox, 2017).

We can formally state the intentions of a MANOVA as follows (Wilcox, 2017, p. 321): The participants in the sample each had p measures taken, where $p = 12$, in this case, corresponding to the twelve scales in the SEED. Furthermore, participants were grouped into J independent groups, where J was determined by the number of categories the demographic variable contained. For example, $J = 2$ when we examined differences in gender since we were working with two groups, namely male and female. For a group j, we compute p trimmed means which we denote by $\mu_j = (\mu_{tj1}, \ldots, \mu_{tjp})$. Our goal here was to test the null hypothesis

$$H_0: \mu_1 = \cdots = \mu_j.$$

In other words, we are testing the hypothesis that the collection of trimmed scale score means for each group are equal. What is evident from our objective is that we are putting forward a multivariate hypothesis. Thus, any significant findings would suggest that potential differences exist between a linear composite of the SEED scales rather than looking at the variables individually as is done with ANOVAs or direct comparisons. We decided to take this route given that we are primarily concerned with answering the question "*Do participants differ on the SEED as a whole?*".

With regards to the present analyses, we set the trimming amount Υ to be .10 (i.e. 10% trimming). We avoided setting $\Upsilon = .2$ as is typically done since the number of observations in certain groups was fairly small, and we wanted to limit the number of observations being excluded in the mean estimations.

To assess our null hypothesis H_0, we performed a robust MANOVA based on the extension of Johansen's (1980) method to trimmed means (Wilcox, 2017). Robust methods were chosen to limit the amount of statistical power lost due to unequal group sizes and to assist in controlling the probability of a Type I error which is higher given the unequal group sizes and non-normal underlying distributions (Wilcox, 2017). Furthermore, two primary parametric assumptions for a standard MANOVA were violated as will be discussed shortly.

After obtaining a significant MANOVA result, our goal was then to determine where these differences may potentially lie by testing a set of linear contrasts in the context of multivariate data (Wilcox, 2017). In general, we were testing whether our linear contrasts were significantly different from zero using a percentile bootstrap method.

PRELIMINARY ANALYSES

Besides the general parametric assumptions that are typically assessed, we were particularly interested in testing for multivariate normality and the assumption of homogeneity of covariance matrices (Field, 2013). Multivariate normality was assessed using the Shapiro-Wilks (1965) test of normality specifically meant for multivariate data. This assumption was violated for the different subsets formed by grouping participants into groups according to their demographic variables. The assumption of homogeneity of covariance matrices was checked using Box's (1949) M-test for homogeneity of covariance matrices. The test was significant, indicating that the data did indeed violate this assumption.

Although MANOVA assumptions were violated, utilising robust methods rather than standard parametric procedures should help mitigate any issues regarding bias for the most part.

GENDER DIFFERENCES

Participants were grouped into two subsets, male and female respondents respectively. The results were that there was a non-significant effect of participants' gender on their scores across the twelve scales, $\hat{F} = 23.67$, $p = .053$. Given these results, linear contrasts were omitted.

AGE DIFFERENCES

We initially attempted to perform a MANOVA with the original age groups presented in Table 1. However, the system was computationally singular, with a reciprocal condition number of 1.50758 x 10^{-18}. We thus decided to collapse participants who were aged 17 years or younger with the group of participants aged between 18-21 years. This solved the previous issue, allowing us to continue with our analysis as planned.

The results suggested that there was a significant effect of a participant's age on their scores across the twelve scales, $\hat{F} = 95.64$, $p < .01$. Given these results, we decided to perform linear contrasts as our post hoc analyses to determine where these differences may lie.

The results were that no linear contrasts were significantly different from zero, and thus we suspect that if any age effects do indeed exist, these effects are most likely small and can be considered to be negligible. We can interpret both sets of results as follows: there is potentially an effect of age on participants SEED scores, however, these differences are not prominent enough to be able to determine between which pairs of age groups these differences may lie.

An alternative possibility is that age may be acting as a moderating variable, directly impacting the relationship between the SEED scores and some other underlying cause that has not been considered or yet identified. For example, it may be that age and work experience are directly proportional to each other, meaning that as a person's age increases so too do we expect their work experience to increase by a proportional amount, and the factors that come to the fore may differ compared to younger people just entering the workforce. This possibility is a potential topic for future research projects.

RASCH ANALYSIS

The Rasch model was developed by Georg Rasch in 1960, which is oftentimes referred to as the 1-parameter logistic model. This model assumes that the probability of correctly responding to an arbitrary item within an assessment depends on: (a) an estimate of item difficulty; and (b) an estimate of the overall ability level of the respondent to solve the item (von Davier, 2014). Stated more formally, given estimates of item difficulty and overall ability level $\hat{b}_i$ and $\hat{\theta}_j$ respectively, the probability of a correct response to an item is defined as

$$P_{ij}(\hat{\theta}_j, \hat{b}_i) = \exp(\hat{\theta}_j - \hat{b}_i)\left(1 + \exp(\hat{\theta}_j - \hat{b}_i)\right)^{-1}.$$

Such equations are relatively simple to solve using specialised computer software that utilises techniques that involve taking logarithms and applying conditional probability routine (Waugh, 2011). The use of Rasch models is commonly used in psychometric test development (Irribarra, 2018). These models allow us "to determine what is measurable on a linear scale, how to determine what data can be used reliably to create a linear scale, and what data cannot be used in the creation of a linear scale" (Waugh, 2011, p. 822).

The sections that follow report on the results of item fit and differential item functioning analyses that were performed for the SEED. Before the results are discussed, we briefly discuss what item fit and differential item functioning analyses entail.

ITEM FIT

Item fit is the degree to which responses conform a logical pattern (Green & Frantom, 2002). Items within an assessment vary in how predictable they are. At times, items may be too predictable or too unpredictable compared to the model. Evaluating fit indices is crucial as it provides evidence as to which items need to be considered for removal or adaptation to allow a better fit to the overall model. The results presented for item fit include the item statistics along with graphical representations for each SEED scale.

Three primary statistics are important to understand for the interpretation of this section, these being:

- Item Location: This refers to how easy it is for a participant to correctly respond to an item in an assessment. The greater the magnitude of the negative values are, the easier the questions. On the other hand, the greater the magnitude of the positive values, the more difficult the item is perceived to be.
- Mean-Squared Statistic (MNSQ): These values indicate how well each item fits against the predictions of the Rasch model. It is expected that these values be close to 1.0, though values ranging between .70 and 1.35 (Lincare, 2015) are generally considered to be an indication of good fitting items. Items with MNSQ values that fall below this cut-off range might be considered to be redundant in that they are measuring similarly to other items or are not adding any additional information. Items with values that exceed the cut-off range might be measuring different constructs than what the item was intended to measure.
- Standardised Fit Statistics (ZSTD): These values are the *z*-score outputs of *t*-tests used to determine how well the data fit the Rasch model. These values should be centralised around zero. Scores that are smaller than zero indicate too much predictability, that is, that there is not enough variance in response patterns. Scores that are higher than zero indicate a lack of predictability which suggests that the item did not function as the Rasch model predicted. Items that have $|ZSTD| > 2.0$ are flagged for further investigation.

INFIT

Infit statistics refer to a weighted fit that is not influenced by specific outliers in the data and is more sensitive to the pattern of responses for a specific sample on the test items.

OUTFIT

Outfit statistics are sensitive to outlying data points. This statistic is influenced by data points that fall outside the expected response pattern (i.e. extremely low scores on specific questions due to time constraints where many participants could not answer the question). It is a less robust measure of item fit, but still gives valuable information about the outlier data points that warrant further investigation.

DIFFERENTIAL ITEM FUNCTIONING

The probability that respondents that are equal in ability level might not have similar response patterns for specific items within a measure based on one or more of their population specifications is referred to as differential item functioning (or DIF for short). Significant DIF values are an indication that certain items in an assessment may be unfair to certain population groups. It is thus crucial that these items not only be identified, but also evaluated as this informative data provides the basis on which judgements can be made regarding possible future item adaptations or removals from the assessment (De Beer, 2004; Strobl, Kopf, & Zeileis, 2011).

We measured bias between different gender and age groups. For these analyses, three statistics that are common in DIF were interpreted, these being:

- DIF Contrast: Indicates the difference between item difficulties for the two groups being compared. A negative DIF contrast value suggests that the item was easier for the first group. In other words, participants from the second group were less likely to get the item correct. A positive DIF value indicates that the item was easier for the second group and that they were more likely to get the item correct. Items with a DIF contrast greater than an absolute value of .50 are identified for further investigation. The significance of the DIF was considered by exploring the Rasch-Welch and Mantel-Haenszel probabilities.
- Rasch-Welch: – The Rasch-Welch test is a *t*-test that estimates a Rasch difficulty for the item for each group in the DIF comparison. The Rasch-Welch test allows for missing data in the dataset.
- Mantel-Haenszel: The Mantel-Haenszel (1959) test is a chi-square estimate of item difficulty differences. The Mantel-Haenszel statistic is the industry standard for reporting DIF in psychometric instruments, but at times cannot be estimated due to small sample sizes.

ATTITUDE SCALE

ITEM FIT

Table 8 provides an overview of how well the items belonging to the Attitude scale of the SEED fit the Rasch model. Item difficulty was relatively spread out across the scale. Moreover, all infit and outfit MNSQ metrics were within their respective ranges. Item 2 had a ZSTD metric outside of the expected ranges, but because the MNSQ metrics for the item were within acceptable ranges, this is not a concern.

Table 8. Item Statistics: Attitude Scale

Item	Item Location	Model S.E.	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
1	-.37	.04	.98	-.20	1.00	.00
2	.04	.04	.85	-2.50	.85	-2.50
3	.30	.04	1.07	1.00	1.06	.90
4	.03	.04	1.06	.90	1.06	.90
M	.00	.04	.99	-.20	.99	-.20
P.SD	.24	.00	.09	1.40	.09	1.40

DIFFERENTIAL ITEM FUNCTIONING

GENDER

Figure 8 gives a graphic representation of how each item performed based on item difficulty for the participants grouped according to their gender classification.

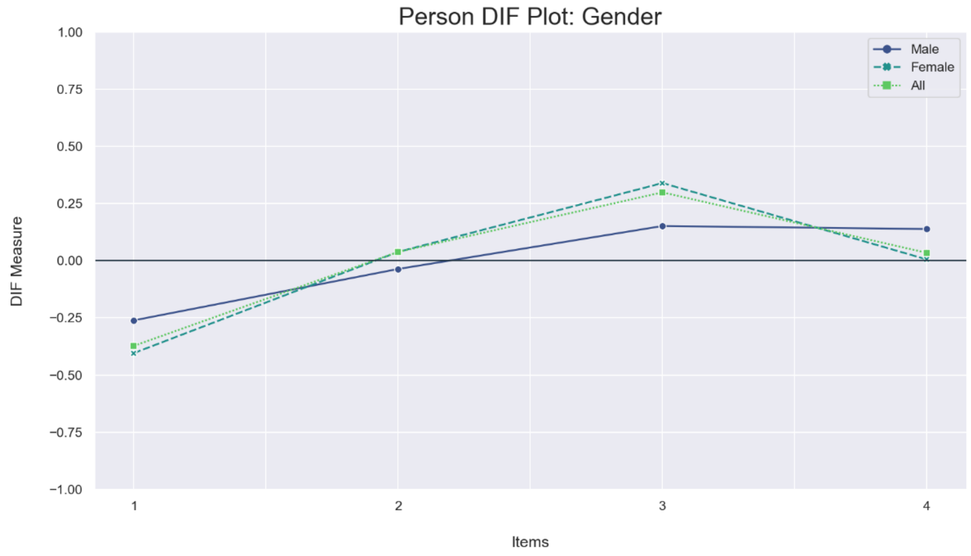


Figure 8. Person DIF plot for the Attitude scale.

No noticeable visual differences seem to be present across the four items in the scale between the two gender groups. *T*-tests revealed no significant differences in item location between the two gender groups, and thus corroborate the graphical representation in the figure.

AGE

Figure 9. Person DIF plot for the Attitude scale. gives a graphic representation of how each item performed based on item difficulty for the three age groups.

Figure 9. Person DIF plot for the Attitude scale.

Three of the items in the scale showed statistically significant differences between the three age groups. Item 1 showed significant differences between the *30 years or older* group and both of the other two age groups. Item 2 had a significant difference between the *21 years and younger* group and the *22 – 29 years* group, while item 3 showed a significant difference between the youngest and oldest age group. None of these differences were, however, greater than the .50 cut-off, so these items are not flagged for further investigation.

SERVICE SCALE

ITEM FIT

Table 9 provides an overview of how well the items belonging to the Service scale of the SEED fit the Rasch model. Although there are only three items in the Service scale, we still see a spread amongst the item location with two items having negative location metrics and one item having a positive location metric. When looking at the mean Item Location, we see this metric is .00 as one would hope. All infit and outfit metrics were within their respective ranges.

Table 9. Item Statistics: Service Scale

Item	Item Location	Model S.E.	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
1	-.14	.04	.92	-1.20	.92	-1.30
2	-.13	.04	.92	-1.30	.92	-1.30
3	.27	.04	1.12	1.80	1.10	1.50
M	.00	.04	.98	-.30	.98	-.40
P.SD	.19	.00	.09	1.40	.08	1.30

DIFFERENTIAL ITEM FUNCTIONING

GENDER

Figure 10 gives a graphic representation of how each item performed based on item difficulty for the participants grouped according to their gender classification.

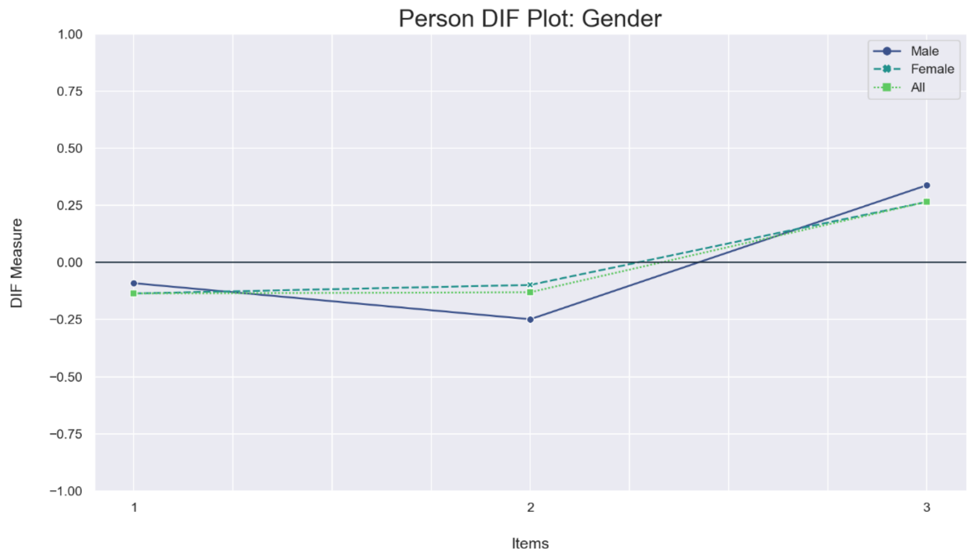


Figure 10. Person DIF plot for the Service scale.

There were no statistically significant differences in how the items functioned for the two gender groups. We see that the DIF plots also follow a fairly unidirectional pattern.

AGE

Figure 11 gives a graphic representation of how each item performed based on item difficulty for the three age groups.

Figure 11. Person DIF plot for the Service scale.

The second item in the Service scale showed statistically significant differences between the *21 years and younger* group and both other age groups. These differences were small (< 0.50) and probably due to the sample, rather than actual group differences on the item.

CREATIVITY SCALE

ITEM FIT

Table 10 provides an overview of how well the items belonging to the Creativity scale of the SEED fit the Rasch model. We see a good spread in the item locations. All infit and outfit MNSQ metrics were within their respective ranges. There were three items where the ZSTD metrics were outside of the expected ranges, but based on the MNSQ metrics, these items are not flagged for further investigation.

Table 10. Item Statistics: Creativity Scale

Item	Item Location	Model S.E.	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
1	-.57	.03	.87	-2.10	.86	-2.20
2	.38	.03	1.09	1.50	1.08	1.30
3	.02	.03	1.02	.30	1.01	.10
4	.27	.03	1.23	3.40	1.23	3.40
5	-.10	.03	.86	-2.30	.88	-2.00
M	.00	.03	1.01	.20	1.01	.10
P.SD	.33	.00	.14	2.20	.14	2.10

DIFFERENTIAL ITEM FUNCTIONING

GENDER

Figure 12 gives a graphic representation of how each item performed based on item difficulty for the participants grouped according to their gender classification. The plots follow a unidirectional pattern, and no statistical differences were found.

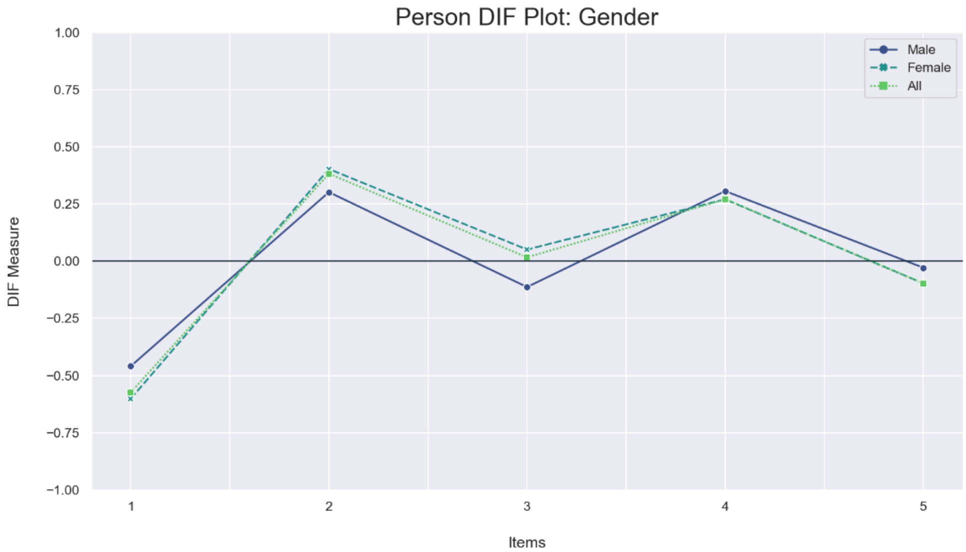


Figure 12. Person DIF plot for the Creativity scale.

AGE

Figure 13. Person DIF plot for the Creativity scale. gives a graphic representation of how each item performed based on item difficulty for the three age groups. We see that all of the DIF plots followed a unidirectional pattern. There were two items with statistically significant differences (items 1 [between the youngest and oldest groups] and 2 [between the oldest group and both other groups]), but the effect of these differences was small (DIF contrast < 0.50), so these significant results are likely to be sample-specific.

Figure 13. Person DIF plot for the Creativity scale.

BARRIERS SCALE

ITEM FIT

Table 11 provides an overview of how well the items belonging to the Barriers scale of the SEED fit the Rasch model. We see a relatively good spread among the item locations with the mean Item Location, being .00. We see two items with ZSTD metrics outside of the expected ranges, but all of the items in the Barriers scale showed acceptable MNSQ metrics.

Table 11. Item Statistics: Barriers Scale

Item	Item Location	Model S.E.	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
1	.44	.03	1.15	2.30	1.14	2.10
2	-.27	.03	.94	-1.00	.93	-1.20
3	-.03	.03	1.14	2.10	1.14	2.20
4	.11	.03	.94	-1.00	.93	-1.10
5	-.20	.03	1.03	.50	1.02	.40
6	-.05	.03	.89	-1.80	.89	-1.90
M	.00	.03	1.01	.20	1.01	.10
P.SD	.23	.00	.10	1.60	.10	1.60

DIFFERENTIAL ITEM FUNCTIONING

GENDER

Figure 144 gives a graphic representation of how each item performed based on item difficulty for the participants grouped according to their gender classification. One of the items (item 1) showed a statistically significant result, but the DIF contrast for this item was below 0.50, suggesting that the difference is due to the sample, rather than an actual difference between how men and women perform on the item.

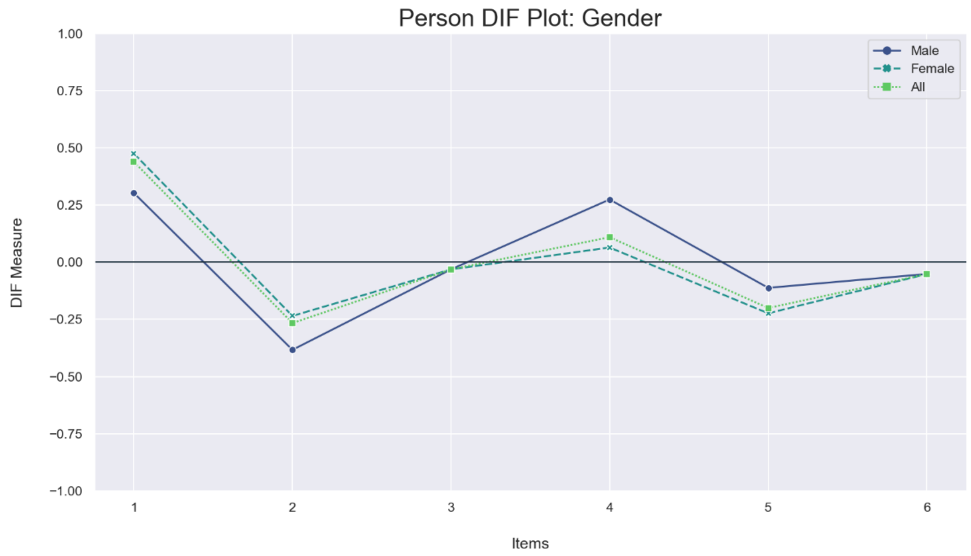


Figure 14. Person DIF plot for the Barriers scale.

AGE

Figure 15 gives a graphic representation of how each item performed based on item difficulty for the three age groups. All of the DIF plots follow a unidirectional pattern, and no items showed any significant differences for the three age group.

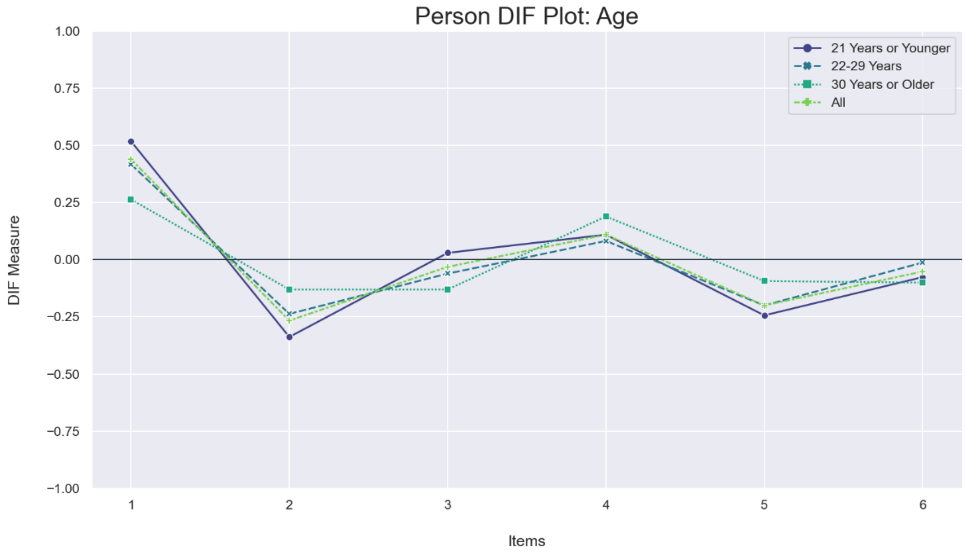


Figure 15. Person DIF plot for the Barriers scale.

FOUNDATION SCALE

ITEM FIT

Table 12 provides an overview of how well the items belonging to the Foundation scale of the SEED fit the Rasch model. We see items with both negative and positive item location metrics and the mean item location metrics indicates an overall fair spread between items. All of the items had MNSQ metrics within the expected ranges, with two items that had ZSTD metrics outside of the expected ranges. These items, both, showed acceptable MNSQ metrics, so they were not flagged for further investigation.

Table 12. Item Statistics: Foundation Scale

Item	Item Location	Model S.E.	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
1	-.10	.04	.98	-.30	.97	-.40
2	.13	.04	1.16	2.40	1.17	2.60
3	-.05	.04	.86	-2.20	.86	-2.20
4	.02	.04	.95	-.70	.95	-.80
M	.00	.04	.99	-.20	.99	-.20
P.SD	.09	.00	.11	1.70	.11	1.80

DIFFERENTIAL ITEM FUNCTIONING

GENDER

Figure 16 gives a graphic representation of how each item performed based on item difficulty for the participants grouped according to their gender classification. The DIF plots do not follow a unidirectional pattern as expected, but when looking at the differences we do not see any statistically significant differences between how the items functioned in the two gender groups.

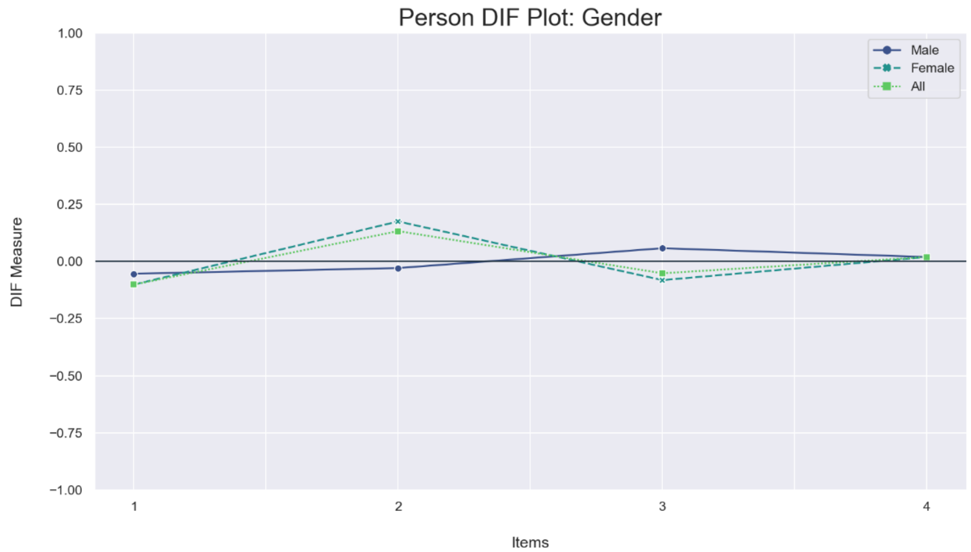


Figure 16. Person DIF plot for the Foundation scale.

AGE

Figure 17 gives a graphic representation of the DIF plots for the three age groups we investigated. There were no statistically significant differences between how the three groups performed on the scale items. This is confirmed by the unidirectional pattern of the DIF plots.

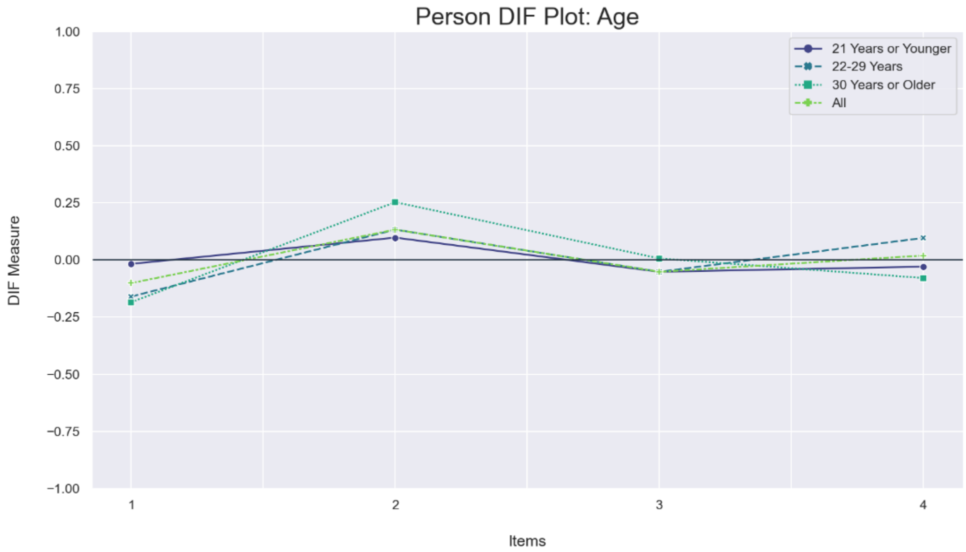


Figure 17. Person DIF plot for the Foundation scale.

CORE SCALE

ITEM FIT

Table 13 provides an overview of how well the items belonging to the Core scale of the SEED fit the Rasch model. Item difficulty was relatively spread out across the scale. Items 9 and 12 had a high ZSTD value above the acceptable 2.00. Additionally, the infit and outfit MNSQ values for item 9 were below the 0.70 cut-off point. These results suggest that item 9 is potentially redundant within the scale. For item 12, the infit MNSQ fell within the acceptable range, but due to the high ZSTD metric, we wonder if this item functions as predicted for the current sample. With samples larger than 300 participants, we often find that the t-statistic of the ZSTD metric can be 'too sensitive' so if the item is not flagged anywhere else, we are not concerned with these metrics.

Table 13. Item Statistics: Core Scale

Item	Item Location	Model S.E.	Infit		Outfit	
			MNSQ	ZSTD	MNSQ	ZSTD
1	.55	.03	1.04	.60	1.03	.50
2	.42	.03	1.05	.80	1.04	.70
3	-.33	.03	1.09	1.40	1.11	1.60
4	.13	.03	.85	-2.50	.84	-2.70
5	.27	.03	1.00	.10	1.00	.00
6	-.46	.03	1.03	.50	1.04	.60
7	-.32	.03	.94	-1.00	.95	-.80
8	-.17	.03	1.28	4.10	1.30	4.30
9	-.34	.03	.69	-5.50	.70	-5.40
10	.15	.03	1.03	.60	1.03	.50
11	-.30	.03	.87	-2.10	.94	-.90
12	.39	.03	1.27	4.00	1.26	3.90
M	**.00**	**.03**	**1.01**	**.10**	**1.02**	**.20**
P.SD	**.34**	**.00**	**.16**	**2.60**	**.16**	**2.50**

DIFFERENTIAL ITEM FUNCTIONING

GENDER

Figure 18 gives a graphic representation of how each item performed based on item difficulty for the participants grouped according to their gender classification. There appear to be small differences in DIF between the male and female groups by looking at Figure 18. No differences, however, were found to be significantly different for any of the items.

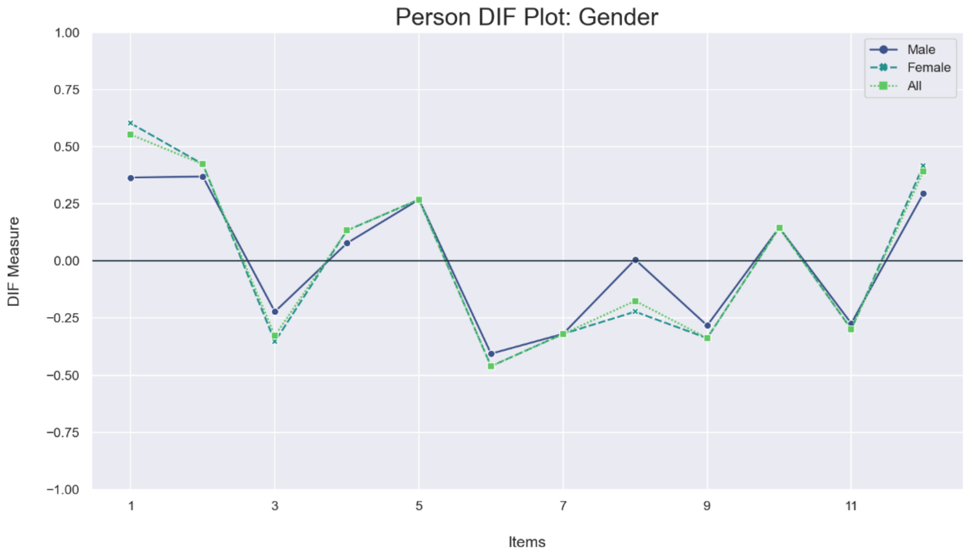


Figure 18. Person DIF plot for the Core scale.

AGE

Figure 19 provides an overview of how the items in the scale performed for the three age groups we investigated. As we can see from the figure, the DIF plots follow a fairly unidirectional pattern, with a few items where we see larger differences. Item 1 and 5 showed statistically significant differences between how the *21 years and younger* group and the '30 and above' group responded. Item 5 and 10 showed statistically significant differences between the responses of the *22 – 29 years* and *30 years or older* groups. None of these differences were large enough (< 0.50) to warrant further investigation.

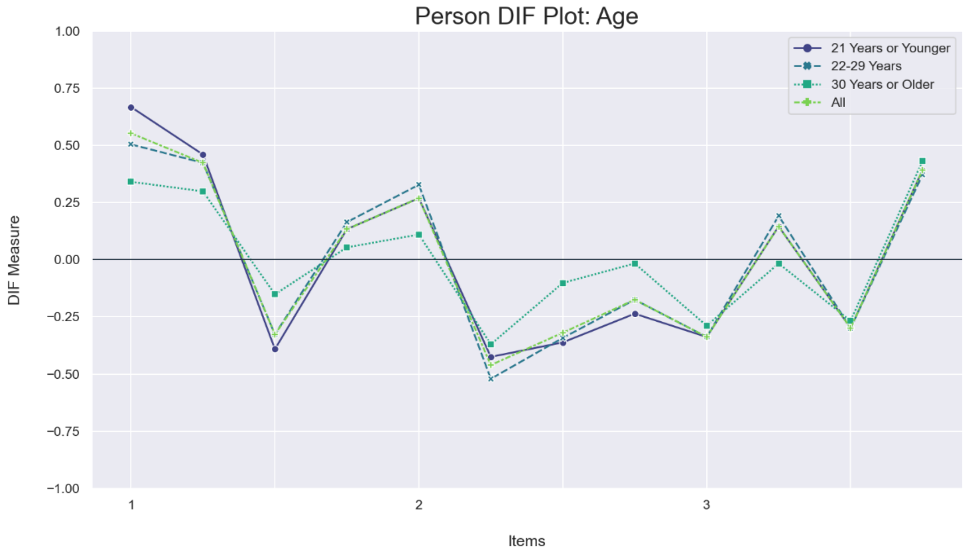


Figure 19. Person DIF plot for the Core scale.

SECTORS SCALE

ITEM FIT

Table 14 provides an overview of how well the items belonging to the Sectors scale of the SEED fit the Rasch model. Item difficulty was relatively spread out across the scale. All items in the scale had acceptable values across the different metrics.

Table 14. Item Statistics: Sectors Scale

Item	Item Location	Model S.E.	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
1	.10	.04	1.12	1.80	1.10	1.50
2	-.06	.04	.89	-1.70	.88	-1.80
3	.01	.04	.99	-.10	1.01	.10
4	-.05	.04	.94	-1.00	.92	-1.30
M	.00	.04	.99	-.30	.98	-.40
P.SD	.06	.00	.09	1.30	.08	1.30

DIFFERENTIAL ITEM FUNCTIONING

GENDER

Figure 20 gives a graphic representation of how each item performed based on item difficulty for the participants grouped according to their gender group.

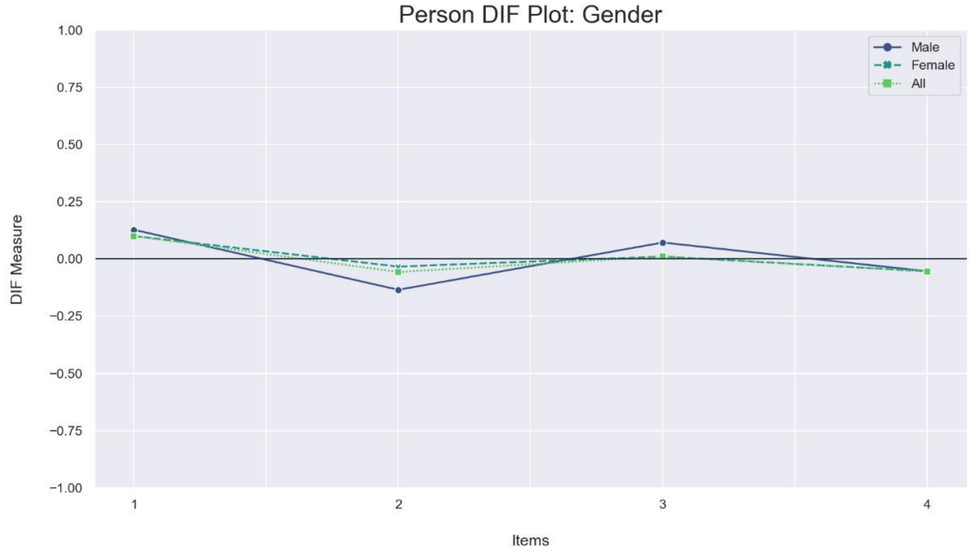


Figure 20. Person DIF plot for the Sectors scale.

A small difference in DIF can be visually seen for item 2. Male participants appear to have found this item easier to endorse as opposed to female participants. However, this difference was found to be not statistically different as indicated by the outputs of the *t*-statistics.

AGE

From Figure 21 we can see that items 3 and 4 of this scale performed differently for the *30 years or older* group than the other two age groups. These differences were, however, small (< 0.50) and not statistically significant.

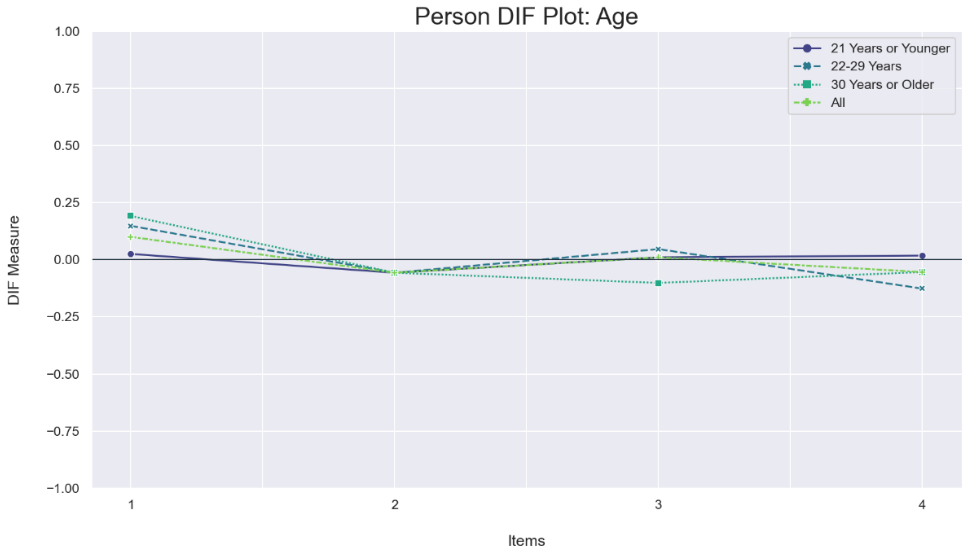


Figure 21. Person DIF plot for the Sectors scale.

BRANDING SCALE

ITEM FIT

Table 15 provides an overview of how well the items belonging to the Branding scale of the SEED fit the Rasch model. Item difficulty was relatively spread out across the scale. Additionally, all infit and outfit metrics were within their respective ranges.

Table 15. Item Statistics: Branding Scale

Item	Item Location	Model S.E.	Infit		Outfit	
			MNSQ	ZSTD	MNSQ	ZSTD
1	-.19	.04	1.02	.40	1.03	.50
2	-.29	.04	.84	-2.60	.84	-2.70
3	.05	.04	1.04	.60	1.02	.40
4	.43	.04	1.05	.70	1.03	.50
M	.00	.04	.99	-.20	.98	-.30
P.SD	.28	.00	.08	1.40	.08	1.30

DIFFERENTIAL ITEM FUNCTIONING

GENDER

Figure 22 gives a graphic representation of how each item performed based on item difficulty for the participants grouped according to their gender classification.

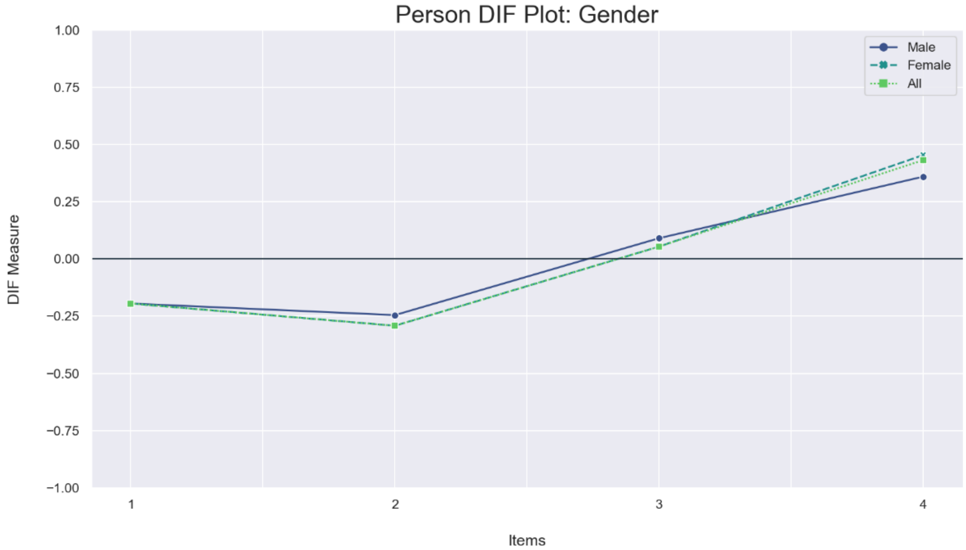


Figure 22. Person DIF plot for the Branding scale.

The figure suggests that there is some difference in item difficulty for item 2. Further analyses using *t*-tests revealed a non-significant effect.

AGE

Figure 23 provides an overview of how the items of the Branding scale functioned across the three age groups.

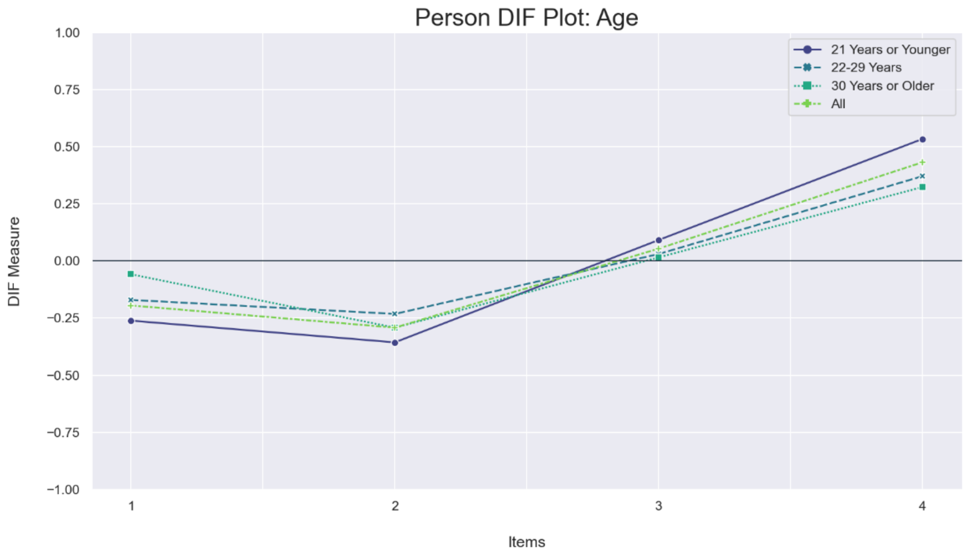


Figure 23. Person DIF plot for the Branding scale.

Only item 4 showed a statistically significant difference between two of the age groups (*21 years or younger* and *30 years or older*). The difference was below the recommended 0.50 cut-off, so the item is not flagged for further investigation.

WORK SCALE

ITEM FIT

Table 16 provides an overview of how well the items belonging to the Work scale of the SEED fit the Rasch model. Item difficulty was relatively spread out across the scale. Item 1 had infit and outfit ZSTD values above 2.00. Furthermore, the infit MNSQ values are close to the cut-off point of 1.35. It may be that item 1 is measuring a different construct. However, this cannot be said with complete confidence given that the MNSQ value was below 1.35 and the fact that ZSTD metrics are sensitive to sample size.

Table 16. Item Statistics: Work Scale

Item	Item Location	Model S.E.	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
1	-.06	.03	1.28	4.10	1.25	3.80
2	.05	.03	.98	-.30	.96	-.70
3	-.23	.03	.88	-1.90	.89	-1.70
4	.25	.03	.90	-1.60	.90	-1.70
M	.00	.03	1.01	.10	1.00	-.10
P.SD	.17	.00	.16	2.40	.15	2.20

DIFFERENTIAL ITEM FUNCTIONING

GENDER

Figure 24 gives a graphic representation of how each item performed based on item difficulty for the participants grouped according to their gender classification.

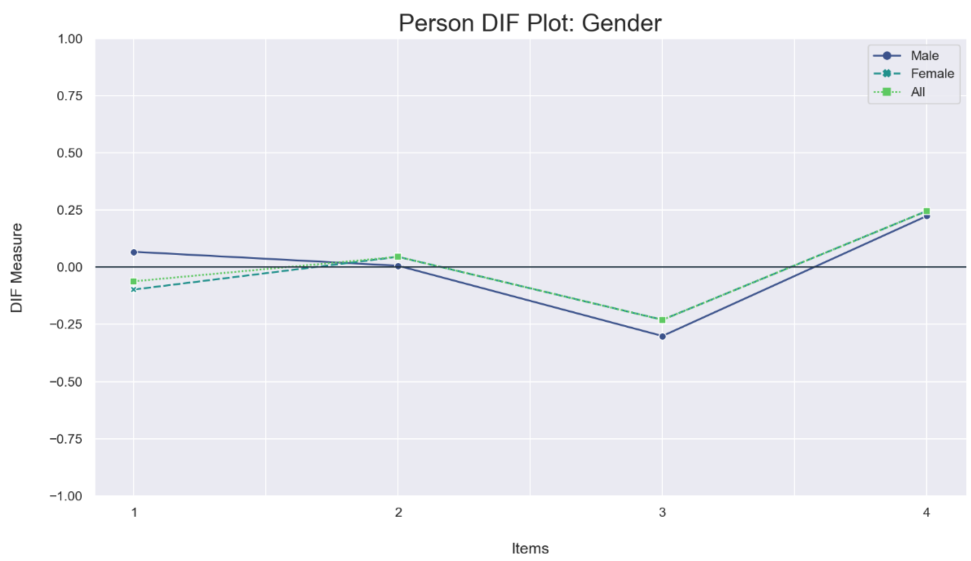


Figure 24. Person DIF plot for the Work scale.

Visually there appear to be differences in item difficulty for the two gender groups which are not unidirectional. These differences were not significant.

AGE

Figure 25 shows the item difficulty plotted for each of the three age groups. We see that there are some visible differences between the age groups, especially on item 1. For items 1 and 2 the youngest age group performed significantly different from both of the other age groups. We also saw that on item 3 the oldest age group performed significantly different from their two younger age groups counterparts. None of these differences were above the recommended cut-off and the items are not flagged for further investigation.

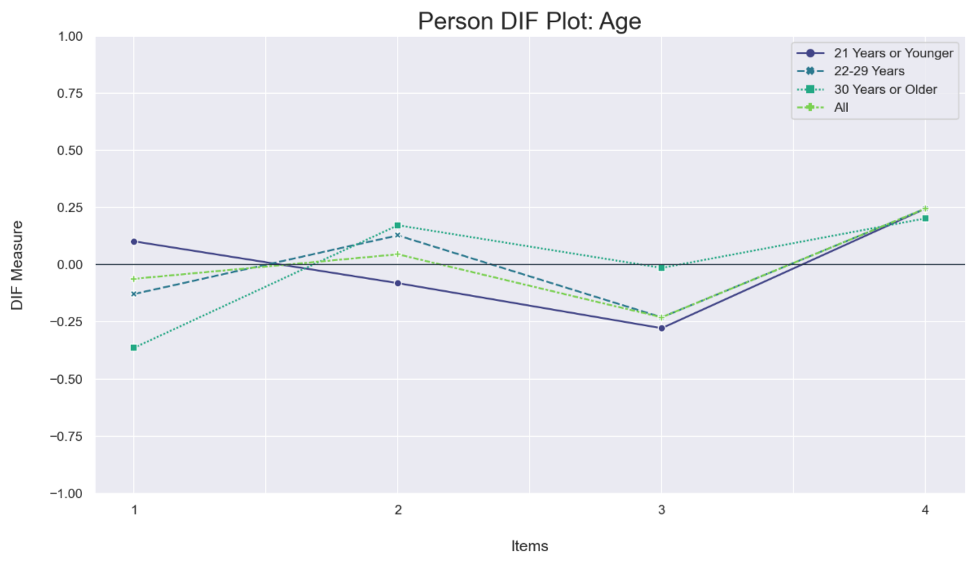


Figure 25. Person DIF plot for the Work scale.

GOALS SCALE

ITEM FIT

Table 17 provides an overview of how well the items belonging to the Goals scale of the SEED fit the Rasch model. Item difficulty was relatively spread out across the scale. All infit and outfit MNSQ metrics were within their respective ranges. Four items had ZSTD metrics outside of the acceptable ranges but based on their MNSQ metrics that were well within the acceptable ranges, these items weren't flagged for further investigation.

Table 17. Item Statistics: Goals Scale

Item	Item Location	Model S.E.	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
1	-.05	.03	1.14	2.10	1.16	2.50
2	-.46	.04	1.15	2.30	1.16	2.40
3	.08	.03	.90	-1.70	.89	-1.70
4	.41	.04	.85	-2.50	.84	-2.70
5	.21	.03	1.05	.80	1.04	.70
6	-.19	.03	.83	-2.80	.84	-2.50
M	.00	.03	.99	-.30	.99	-.20
P.SD	.28	.00	.13	2.10	.14	2.20

DIFFERENTIAL ITEM FUNCTIONING

GENDER

Figure 26 gives a graphic representation of how each item performed based on item difficulty for the participants grouped according to their gender classification. The only potential difference that one may suspect to exist from Figure 26 is for item 2. This difference was not significant.

Figure 26. Person DIF plot for the Goals scale.

AGE

Figure 27 provides an overview of how each of the three age groups performed on the items of the Goals scale.

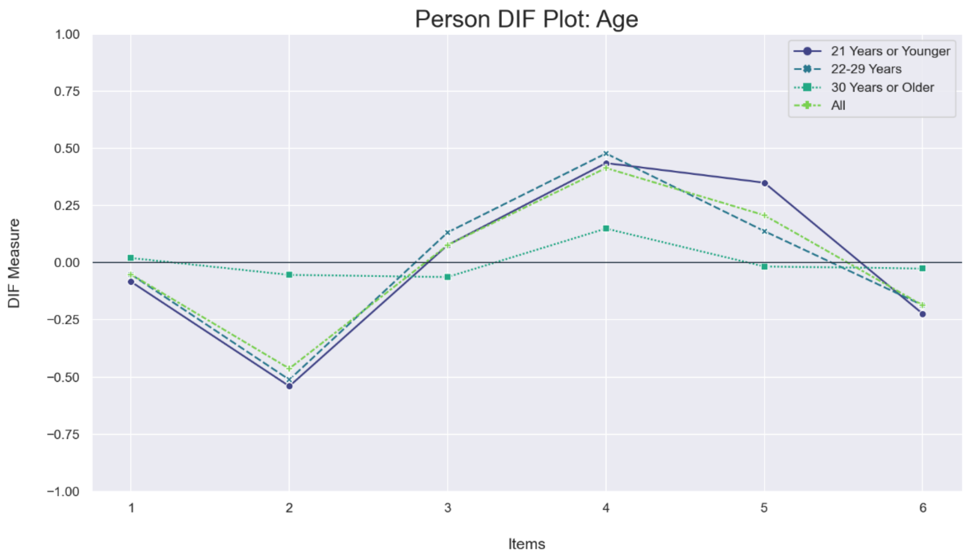


Figure 27. Person DIF plot for the Goals scale.

When looking at the DIF plots we can see that the oldest age group performed differently on the items than the other two age groups. On items 2 and 4 we see that the oldest age group yielded statistically significant results to their other age group counterparts. Item 2, specifically, also had a DIF contrast close to the 0.50 cut-off in both instances and should potentially be investigated further. On item 5 we also saw that the youngest age group performed significantly different from the other two groups, but these differences were small.

NETWORK SCALE

ITEM FIT

Table 18 provides an overview of how well the items belonging to the Network scale of the SEED fit the Rasch model. Item difficulty was relatively spread out across the scale. All infit and outfit MNSQ metrics were within their respective ranges. Although we saw that items 1 and 3 had ZSTD metrics beyond the acceptable ranges, these items still showed overall good fit statistics based on their MNSQ values. They were not flagged for further investigation.

Table 18. Item Statistics: Network Scale

Item	Item Location	Model S.E.	Infit		Outfit	
			MNSQ	ZSTD	MNSQ	ZSTD
1	-.25	.04	.85	-2.50	.83	-2.70
2	-.15	.04	.98	-.30	.95	-.70
3	.27	.04	1.14	2.20	1.15	2.20
4	.13	.04	.98	-.30	.97	-.40
M	**.00**	**.04**	**.99**	**-.20**	**.98**	**-.40**
P.SD	**.21**	**.00**	**.11**	**1.70**	**.11**	**1.80**

DIFFERENTIAL ITEM FUNCTIONING

GENDER

Figure 28 gives a graphic representation of how each item performed based on item difficulty for the participants grouped according to their gender classification. The *t*-test results indicated that there were no significant differences in item difficulty between the two gender groups.

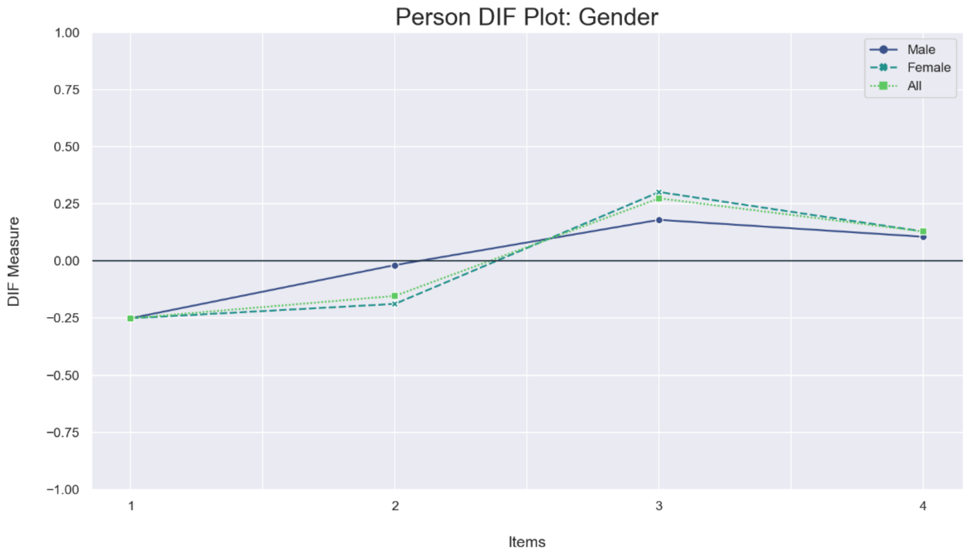


Figure 28. Person DIF plot for the Network scale.

AGE

Figure 29 provides an overview of the item difficulty for the three age groups on the Network scale. Although, from the graph, it seems as if the *30 years or older* group performed differently from the other two groups on the items, there were no statistically significant differences. On one item (item 1) we saw a statistically significant difference between how the *21 years or younger* and the *22-29 years* group responded, but the effect of this difference was small.

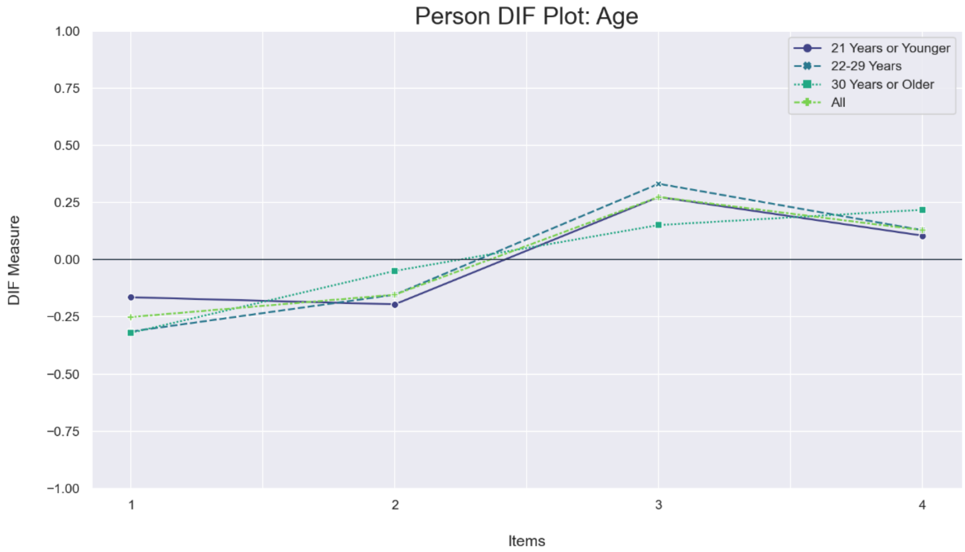


Figure 29. Person DIF plot for the Network scale.

TRANSITIONS SCALE

ITEM FIT

Table 19 provides an overview of how well the items belonging to the Transitions scale of the SEED fit the Rasch model. Item difficulty was relatively spread out across the scale. All infit and outfit metrics were within their respective ranges.

Table 19. Item Statistics: Transitions Scale

Item	Item Location	Model S.E.	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
1	-.33	.04	1.05	.80	1.05	.80
2	.26	.04	.90	-1.60	.88	-1.90
3	.20	.04	.99	-.20	1.00	.00
4	-.12	.04	1.00	.10	1.01	.20
M	.00	.04	.99	-.20	.99	-.20
P.SD	.24	.00	.05	.90	.06	1.00

DIFFERENTIAL ITEM FUNCTIONING

GENDER

Figure 30 gives a graphic representation of how each item performed based on item difficulty for the participants grouped according to their gender classification.

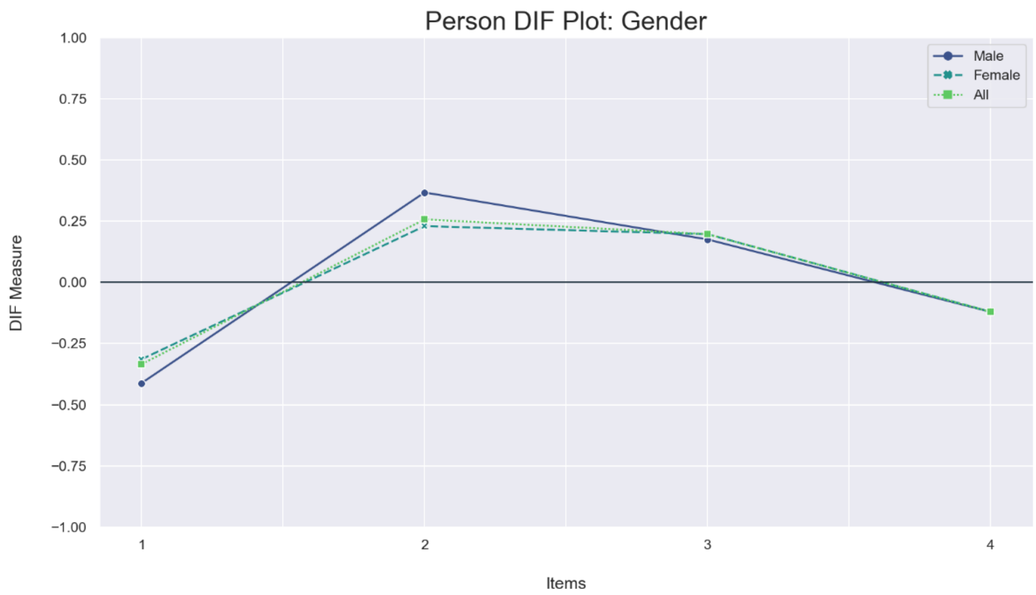


Figure 30. Person DIF plot for the Transitions scale.

One trend that is observable in Figure 30 is that there is a difference in item locations between male and female participants across all the items. None of these differences was significant.

AGE

An overview of how the different age groups endorsed the items of the Transitions scale is presented in Figure 31. The DIF plots for the three groups follow a fairly unidirectional pattern and when investigating if any differences existed between how the groups performed on the items, we found no statistically significant differences.

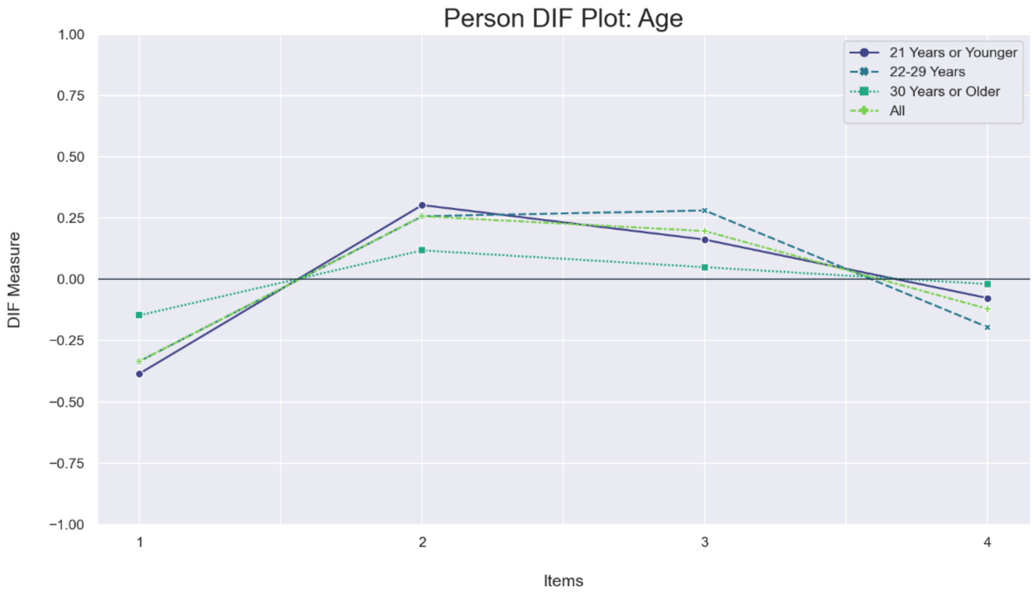


Figure 31. Person DIF plot for the Transitions scale.

CONCLUSION

This report focused on establishing the psychometric properties of the SEED on a sample of 483 participants who completed the assessment. After having explored the SEED scales, we noticed that some of the underlying distributions did not follow a normal distribution. This discovery did guide us in the additional analyses going forward. We also found that significant relationships existed among the theoretical dimensions (i.e. Knowledge, Importance and Experience) that individuals are assessed over, suggesting to us that participants generally responded consistently for each item.

The underlying model of the SEED was assessed using factor analyses. Although the EFAs hinted that the theoretical twelve-factor model would be appropriate for the assessment given that the model has the best fit indices, we found in the CFA that the theoretical model probably has some degree of misspecification. Adaptations may be the appropriate next step to take to improve the fit of the model. Given the purpose of the assessment, we feel that the SEED's use should not be greatly impacted by the results of our factor analyses, especially given that the SEED is an assessment that aims to assist individuals in exploring and developing their career skills. Since the items are ultimately measuring employability skills, interpretations on both a scale and possibly an item level may greatly assist these discussions.

The reliability of the SEED was also established and the scales all had acceptable reliability coefficient estimates. Additionally, no significant differences in SEED results between gender and age groups were apparent. We did note a potential age effect, but this effect may be too small to detect or it may be the case that age is acting as a potential moderator variable. The latter hypothesis is mere speculation on our part but could be an interesting topic for future research projects.

When investigating the results from the Rasch analyses, we saw that only a single item was flagged for potentially misfitting the model. This item was seen as redundant, i.e. not adding any additional information to the assessment, but because no other evidence existed to support the removal of this item, we decided to retain it. There was no evidence of potential bias across the items of the SEED for gender or age groups. Where we did see statistically significant differences between how different groups performed on the items, the differences were small and are probably sample-specific rather than an indication of actual group differences.

In conclusion, the SEED appears to be an appropriate measure to assist individuals in exploring and developing their career skills. We do advise that future research be dedicated to consistently improving the assessment, especially attempting to further examine the factor structure of the assessment so that future adaptions can help strengthen the underlying SEED model.

REFERENCES

Bartlett, M. S., (1951), The effect of standardization on a chi square Approximation in Factor Analysis, *Biometrika, 38*, 337-344. https://doi.org/10.1093/biomet/38.3-4.337

Beukes, C. J. (2010). *Self Education Employability Device*. Pretoria: African Skill Framework for Career Management.

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika, 36*, 317-346. https://doi.org/10.1093/biomet/36.3-4.317

Breusch, T. S., Pagan, A. R. (1979). A simple test for heteroskedasticity and random coefficient variation. *Econometrica, 47*(5), 1287–1294. https://doi.org/10.2307/1911963

Cangur, S. & Ercan, I. (2015). Comparison of model fit indices used in structural equation modeling under multivariate normality. *Journal of Modern Applied Statistical Methods, 14*(1), Article 14. https://doi.org/10.22237/jmasm/1430453580

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16* (3), 297–334. https://doi.org/10.1007/BF02310555

De Beer, M. (2004). Use of differential item functioning (dif) analysis for bias analysis in test construction. *South African Journal of Industrial Psychology, 30*(4), 52-58.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics: And gender and drugs and rock 'n roll (4th edition)*. Los Angeles, (etc.): Sage.

Frantom, C. G., & Green, K. E. (2002). *Survey development and validation with the Rasch model* [Paper presentation]. International Conference on Questionnaire Development, Evaluation, and Testing, Charleston, SC, November.

Health Professions Act, Number 56 (1974). Government Gazette, 433 (38816), Cape Town, 25 May.

Herzog, W., Boomsma, A., & Reinecke, S. (2007). The model-size effect on traditional and modified tests of covariance structures. *Structural Equation Modeling, 14*, 361-390.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65-70.

Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist*. London: Sage.

Irribarra, D. T. (2018). Rasch Model. In B. B. Frey (Ed.), *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* (pp. 1371-1374), SAGE Publications, Inc. http://doi.org/10.4135/9781506326139

Johansen, S. (1980). The Welch-James approximation of the distribution of the residual sum of squares in weighted linear regression. *Biometrika, 67*, 85-92.

Kaiser, H. F. (1960). The application of electronic computer to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Kaiser, H. F. (1970). A second-generation little jiffy. *Psychometrika, 35*, 401-415.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*, 936-949. https://doi.org/10.3758/s13428-015-0619-7

Linacre, J. M. (2015). *Winsteps Rasch Measurement Computer Program User's Guide*. Beaverton, OR: Winsteps.com.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*(4), 719-748. https://doi.org/10.1093/jnci/22.4.719

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.

McNeish, D. (2020). Should We Use F-Tests for Model Fit Instead of Chi-Square in Overidentified Structural Equation Models? *Organisational Research methods, 23*(3), 487-510. https://doi.org/10.1177/1094428118809495

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests*. Copenhagen, Denmark: Danish Institute for Educational Research.

Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3–4), 591–611. http://doi.org/10.2307/2333709

Strobl, C., Kopf, J., & Zeileis, A. (2011). A new method for detecting differential item functioning in the Rasch model. *Working Papers in Economics and Statistics, No. 2011-01*. Innsbruck, Austria: University of Innsbruck. Retrieved from http://hdl.handle.net/10419/73503

Taylor, N., & Beukes, C. J. (2019). A life design-related career development intervention for working adults in the manufacturing, engineering and related sectors. *African Journal of Career Development, 1*(1), a2. https://doi.org/10.4102/ajcd.v1i1.2

Tucker, L. R. (1951). A method for synthesis of factor analysis studies. *Personnel Research Section Report No.984*, Washington D.C.: Department of the Army.

Von Davier M. (2014) Rasch analysis. In: Michalos A.C. (eds) *Encyclopedia of Quality of Life and Well-Being Research*. Springer, Dordrecht.

Waugh, R. F. (2011). Rasch measurement model. In N. J. Salkind (Ed.), *Encyclopedia of Measurement and Statistics* (pp. 821-825), SAGE Publications, Inc. http://doi.org/10.4135/9781412952644

Wilcox, R. (2017). *Introduction to Robust Estimation and Hypothesis Testing* (4th Ed). Elsevier.

Zhang, Z., & Yuan, K. (2015). Robust coefficients alpha and omega and confidence intervals with outlying observations and missing data: methods and software. *Education and Psychological Measurement, 76*(3), 1-25. https://doi.org/10.1177/001316441559465

APPENDIX

Appendix A: EFA Factor Loadings

Table 20. EFA Factor Loadings

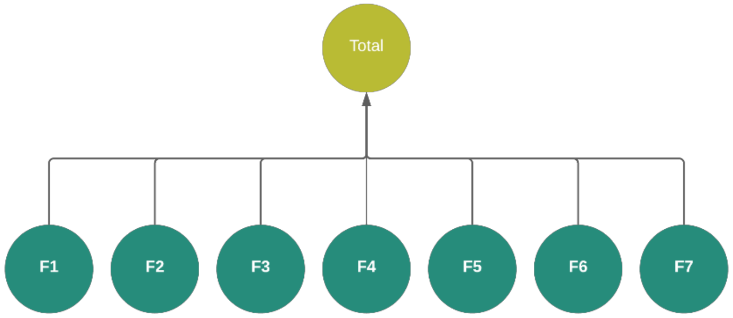
	Attitude	Goals	Sectors	Core	Network	Barriers	Branding	Foundation	Creativity	Transitions	Work	Service
Q1	0.55	0.06	-0.05	0.08	-0.03	0.03	-0.05	0.01	0.18	0.00	-0.01	-0.14
Q2	0.58	0.14	0.00	0.09	-0.01	0.05	-0.03	0.05	0.08	0.04	0.02	-0.11
Q3	0.52	0.07	0.12	-0.07	0.03	0.06	0.10	0.00	-0.01	0.03	0.01	-0.09
Q4	0.57	0.07	-0.10	-0.02	0.03	-0.04	0.09	0.00	0.06	-0.02	0.05	0.24
Q5	0.51	-0.01	0.06	-0.01	-0.02	0.01	0.11	0.03	-0.08	0.08	0.12	0.17
Q6	0.36	0.01	0.11	0.07	0.04	0.04	0.11	0.09	-0.05	-0.07	-0.01	0.29
Q7	0.47	-0.01	0.10	-0.12	0.01	0.16	0.09	-0.08	0.07	0.11	0.02	0.05
Q8	0.45	-0.08	-0.07	0.20	0.09	0.05	0.03	0.25	-0.16	0.10	-0.10	0.11
Q9	0.31	-0.02	0.25	-0.12	0.26	0.27	-0.02	0.12	0.05	-0.05	-0.12	-0.01
Q10	0.40	-0.10	0.08	0.18	0.17	0.06	0.01	0.02	0.08	0.06	0.03	0.01
Q11	0.12	0.04	0.06	0.05	-0.03	0.36	0.09	0.07	-0.08	0.07	0.06	0.08
Q12	0.29	-0.02	0.01	0.13	0.09	0.06	0.15	0.09	-0.07	0.10	-0.03	0.15
Q13	0.06	0.03	0.15	0.04	0.12	0.42	-0.03	0.02	0.12	0.02	0.01	0.02
Q14	0.10	0.10	-0.02	0.01	0.02	0.43	0.18	0.05	-0.04	0.09	0.03	0.06
Q15	0.10	0.11	-0.06	0.15	0.02	0.41	-0.01	0.04	0.10	0.04	0.12	0.02

Q16	0.05	-0.02	0.09	-0.17	0.01	0.23	0.13	0.25	0.12	0.06	0.14	0.20
Q17	0.13	0.12	0.00	-0.01	-0.02	**0.30**	0.07	**0.33**	-0.01	0.11	-0.04	0.05
Q18	0.03	-0.02	0.00	0.02	0.06	0.26	0.11	0.17	0.13	0.10	0.00	**0.35**
Q19	0.06	0.17	0.04	0.01	0.07	0.04	0.00	**0.57**	0.02	0.01	0.00	0.07
Q20	0.02	-0.05	0.17	0.00	-0.07	0.05	0.28	**0.36**	0.20	0.03	0.05	-0.15
Q21	-0.01	0.06	-0.07	0.21	-0.06	0.09	0.29	**0.32**	0.14	0.03	0.09	-0.12
Q22	-0.03	0.03	0.02	0.18	0.22	0.18	0.23	0.23	-0.03	-0.01	0.07	0.03
Q23	0.08	0.06	0.22	-0.01	-0.05	0.10	0.19	0.06	**0.37**	0.02	0.06	0.12
Q24	0.02	0.00	**0.31**	0.07	-0.07	0.13	0.03	0.22	0.05	0.09	0.27	0.05
Q25	0.06	0.09	0.00	0.06	-0.01	0.03	**0.69**	0.01	-0.05	0.03	0.02	0.04
Q26	0.13	0.07	0.12	0.24	0.08	0.03	0.17	0.05	0.21	-0.04	0.00	0.05
Q27	0.05	0.08	0.19	**0.34**	0.02	0.21	-0.07	-0.01	0.22	-0.04	-0.06	0.20
Q28	0.07	-0.09	-0.12	0.15	0.11	-0.02	0.19	0.28	**0.31**	0.07	0.00	0.03
Q29	0.10	-0.06	0.09	**0.44**	0.03	-0.07	0.26	0.15	0.00	0.16	-0.03	-0.03
Q30	-0.01	0.05	0.02	**0.57**	-0.08	0.13	0.10	-0.07	-0.05	0.02	0.11	0.05
Q31	0.08	0.04	-0.04	**0.40**	0.07	0.02	0.21	0.12	0.14	-0.04	0.09	0.03
Q32	0.07	0.14	0.10	0.07	0.13	0.22	0.19	0.04	**0.30**	-0.05	-0.04	-0.16
Q33	0.08	0.06	0.10	**0.45**	0.17	-0.04	-0.03	0.14	0.10	0.02	-0.07	-0.05
Q34	0.05	0.09	0.27	0.08	0.17	0.09	0.06	0.12	-0.04	-0.11	0.23	-0.01
Q35	0.00	0.07	**0.45**	0.22	0.08	0.22	-0.05	-0.11	0.01	0.02	0.13	0.03
Q36	0.02	0.13	**0.65**	0.06	0.00	0.02	0.05	0.06	0.01	0.10	-0.01	0.01
Q37	0.00	0.02	**0.40**	-0.13	0.17	0.02	0.03	0.14	0.07	0.03	0.22	0.10
Q38	0.05	0.15	**0.40**	0.05	-0.02	-0.04	-0.01	0.01	0.13	0.04	0.13	0.25

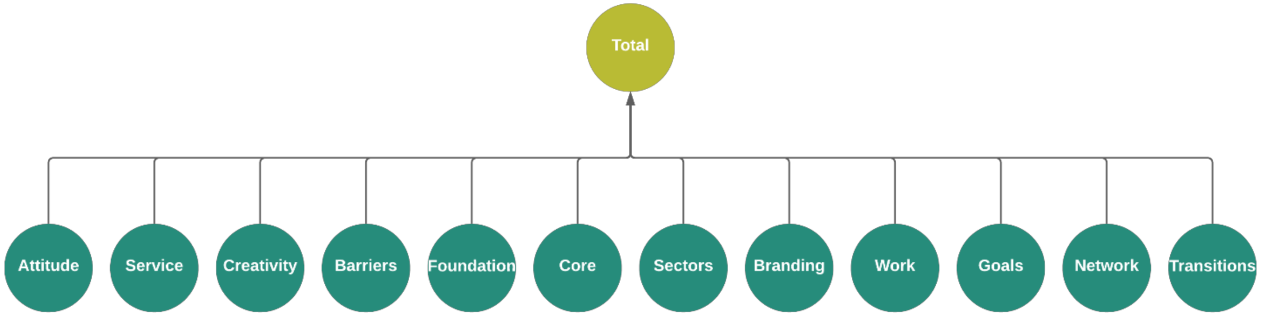
Q39	0.04	0.26	0.05	0.16	0.24	-0.07	-0.02	0.25	0.09	-0.09	0.07	0.17
Q40	0.03	0.18	0.08	0.13	0.22	0.01	-0.04	0.17	0.12	0.07	0.11	0.15
Q41	0.04	0.06	0.14	0.03	0.23	0.04	0.23	-0.04	0.06	0.04	0.14	0.22
Q42	0.00	0.22	0.17	0.01	0.07	0.15	-0.02	-0.06	0.10	0.12	0.22	0.21
Q43	0.07	0.10	-0.02	-0.01	0.01	0.01	-0.13	0.05	**0.56**	0.18	0.07	0.04
Q44	0.06	0.11	0.07	0.13	0.08	0.10	-0.04	-0.08	0.08	0.13	0.25	0.21
Q45	0.10	0.02	0.20	0.19	0.12	-0.19	0.03	0.09	0.11	0.20	0.25	-0.01
Q46	0.01	0.08	0.08	0.03	**0.36**	0.04	0.11	-0.13	0.22	-0.11	0.22	0.11
Q47	0.10	0.17	-0.05	0.00	0.19	0.03	0.19	-0.04	0.09	0.08	**0.34**	0.03
Q48	0.07	0.17	0.02	**0.31**	0.16	-0.02	0.10	0.09	-0.09	0.11	0.12	0.05
Q49	-0.03	**0.68**	-0.03	0.07	0.08	0.10	0.08	0.01	0.01	0.12	-0.15	0.08
Q50	0.05	**0.83**	0.04	-0.04	0.02	-0.02	0.04	0.02	0.03	-0.04	0.07	-0.05
Q51	0.02	**0.58**	0.12	0.05	-0.04	0.01	-0.07	0.06	0.03	0.09	0.10	0.03
Q52	0.05	0.22	0.06	0.01	0.19	0.04	0.10	0.09	-0.03	0.22	0.25	-0.05
Q53	-0.02	0.03	-0.05	0.02	**0.48**	0.16	0.07	-0.03	0.06	0.22	0.16	-0.06
Q54	0.04	0.10	0.03	0.00	**0.67**	0.00	-0.03	0.06	-0.03	0.07	-0.01	0.05
Q55	-0.01	0.18	**0.35**	-0.06	0.23	-0.07	0.20	-0.11	0.19	0.19	-0.17	0.07
Q56	-0.02	0.02	-0.01	0.13	0.28	0.14	-0.04	0.00	0.16	0.27	0.20	0.00
Q57	0.14	0.02	0.05	0.12	0.21	-0.11	0.09	0.09	0.10	**0.33**	0.05	-0.09
Q58	0.06	0.17	0.12	-0.03	0.08	0.01	-0.02	0.00	0.12	**0.45**	-0.01	0.13
Q59	0.03	**0.30**	0.09	0.04	-0.02	0.08	0.02	0.01	0.05	**0.34**	0.08	0.04
Q60	0.06	0.06	0.02	-0.02	0.13	0.21	0.16	0.01	0.05	**0.38**	0.02	-0.09

Appendix B: Path Diagrams for CFA Models

Model 1: Seven-Factor Model



Model 2: Twelve-Factor Model



PSYCHOMETRICS