|  |
| --- |
| University of the West of Scotland |
| Validation of the NHS Scotland Employee Engagement Index |
| Final Report |
|  |
| **Dr Austyn Snowden & Dr Ewan MacArthur** |
|  |



|  |
| --- |
| This is the final report for the external validation of NHS Scotland Employment Engagement Index (NHSEEI). It entails in depth analysis of 1193 responses to the questionnaire. It describes a four-factor structure to the questionnaire and details the item fit. It analyses the response pattern of a typical NHS Scotland staff member and concludes that this questionnaire provides an excellent metric within which to measure improvement in staff engagement. Recommendations are detailed within the body of the report. |

Table of Contents

[Executive summary 4](#_Toc351201873)

[Introduction 6](#_Toc351201874)

[Background 6](#_Toc351201875)

[Method 9](#_Toc351201876)

[Data 9](#_Toc351201877)

[Analytic procedures 10](#_Toc351201878)

[Results 12](#_Toc351201879)

[Descriptives: participants 12](#_Toc351201880)

[Descriptives: responses 13](#_Toc351201881)

[The NHSScotland Employee Engagement Index (NHSEEI) 15](#_Toc351201882)

[Analysis 17](#_Toc351201883)

[Factor analysis 17](#_Toc351201884)

[IRT & Rasch Analysis 21](#_Toc351201885)

[Item person map with Likert data 22](#_Toc351201886)

[Detailed analysis of each question. 25](#_Toc351201887)

[Summary synthesis of item analysis. 25](#_Toc351201888)

[Is engagement related to role? 27](#_Toc351201889)

[How reliable is this finding? 27](#_Toc351201890)

[Discussion 28](#_Toc351201891)

[What this study found 29](#_Toc351201892)

[Context and theory: why these findings are valid 29](#_Toc351201893)

[A justification for factor analysis 30](#_Toc351201894)

[A justification for Rasch analysis 30](#_Toc351201895)

[Factor analysis and Rasch analysis: complementary or contradictory? 30](#_Toc351201896)

[Where are the weak points? 31](#_Toc351201897)

[Item 29: A global metric? 31](#_Toc351201898)

[What did NHSScotland staff think of the NHSEEI? 32](#_Toc351201899)

[Recommendations 32](#_Toc351201900)

[References 33](#_Toc351201901)

[Appendix 1. NHSEEI 35](#_Toc351201902)

**Figures**

[Figure 1. Drivers of staff engagement 8](#_Toc351020501)

[Figure 2. Sample by profession 12](#_Toc351020502)

[Figure 3. National NHSScotland staff population breakdown 13](#_Toc351020503)

[Figure 4. Summary of all responses by Likert Category 14](#_Toc351020504)

[Figure 5. Mean NHSEEI (%) by profession 15](#_Toc351020505)

[Figure 6. NHSEEI score by frequency 16](#_Toc351020506)

[Figure 7. Responses to item 29: Overall within my organisation… 16](#_Toc351020507)

[Figure 8. Four-factor solution 20](#_Toc351020508)

[Figure 9. Item person map 21](#_Toc351020509)

[Figure 10. Item person map with Likert parameter thresholds 23](#_Toc351020510)

[Figure 11. Bubbleplot of all 28 items. 24](#_Toc351020511)

[Figure 12. Bubbleplot with reversed items removed 26](#_Toc351020512)

**Tables**

[Table 1. Principal component analysis with reversed items removed 19](#_Toc351130768)

[Table 2. Item difficulty estimates by rank 22](#_Toc351130769)

[Table 3. ANOVA in profession 27](#_Toc351130771)

[Table 4. Correlation between NHSEEI and item 29 27](#_Toc351130772)

[Table 5. ANOVA by profession using item 29 28](#_Toc351130773)

## Executive summary

*Background*

The NHS Scotland Employee Engagement Index (NHSEEI) is a summary score presented as a percentage, derived from a 28 item questionnaire. In line with current policy[[1]](#footnote-1) it is designed to inform individuals, teams and organisations about the degree of staff engagement. Built in conjunction with NHS Scotland staff, Scotland is the only country in the world to be developing such a systematic and systemic measure in this inclusive manner.

The NHSEEI has been fine tuned over a series of pilots and validation events through 2011-2013. A sample of 160 NHS employees completed pilot 1. A sample of 247 completed pilot 2. This final report details the responses from pilot 3, consisting of responses from 1,271 staff from 3 boards: NHS Golden Jubilee, NHS Forth Valley and NHS Tayside. As a result of this excellent sample size we can be confident the results presented here should be generalisable to all NHS Scotland employees.

*The NHSEEI*

The NHSEEI is generated from 28 items relating to staff engagement (Appendix 1). Each item has six Likert type responses: ‘Strongly Agree’, ‘Agree’, ‘Slightly Agree’, ‘Slightly Disagree’, ‘Disagree’, ‘Strongly Disagree’. These responses were translated to a score for statistical analysis, with 6 for ‘Strongly Agree’ down to 1 for ‘Strongly Disagree’. The sum total of these 28 item responses gives a range of 28 – 168. The NHSEEI is then presented as a percentage, representing the proportion of actual score in relation to the maximum possible score. NHSEEIs can subsequently be generated for any subsample, for example professions, or teams, or board areas. This report has data for professions and so that is utilised as the unit of analysis here. There is also a 29th item: ‘*Overall working within my organisation is a: very good experience* (10 points)*…very poor experience’* (0 points). Response entails an 11 point sliding scale.

*Methods*

The questionnaire was given to 2300 staff in electronic and paper form in Feb 2013. Descriptives, factor analyses and Rasch analyses were conducted to analyse response patterns and psychometric properties of NHSEEI. Three focus groups were run in the participating boards to ascertain feedback.

*Results*

* 1271 staff responded to the survey, a return rate over 56%. This is more than twice the return rate on the 2010 National Staff Survey
* According to both mean NHSEEI (%) and item 29 the most engaged profession were the managers (77% on NHSEEI). The least engaged were support services but they still scored average 62%. Mean overall response equated to 69%, consistent with the Likert category ‘Agree’.
* The four main factors established through principal component analysis mirrored the structure of the questionnaire:
1. My experience as an individual
2. My experience with my direct line manager
3. My experience with my team
4. My experience with my organisation
* Factor analysis is therefore best explained by the item placement within the questionnaire. Because of substantial variance in the response patterns and the strong theoretical co-construction of the NHSEEI this is interpreted as a function of the questionnaire’s reliability.
* Rasch analysis showed there was a good range of item difficulty, which means the NHSEEI is a broad and inclusive measure of staff engagement.
* The majority of the 28 items fit the model of staff engagement very well with the following exceptions:
* The reversed items did not fit the model, and six items were a borderline fit. Factor analysis accounted for the borderline items but not the reverse items. These reverse items should therefore be amended.
* Overall the 28 items show enough variation to contribute uniquely to the overall construct and measurement of staff engagement.
* Focus group feedback was very positive, particularly in relation to the time taken to complete, the pertinence of the items and the layout. Only minor comments emerged.

**Conclusions**

The NHSEEI is a robust, reliable, valid and popular measure of staff engagement. It is also an excellent tool to measure improvement in staff engagement.

**Recommendations**

Remove the reversal of items 4, 16 and 22, otherwise leave the 28 items as they are. Add a timescale to specify the response period.

Keep item 29 but reverse the scale to go from 0 – 10 instead of 10 – 0. Responses to item 29 should be utilised as a snapshot for crosschecking levels of staff engagement and ascertaining response logic. They should not be used as a standalone metric.

Keep sample sizes comparably high in subsequent iterations to maintain the level of credibility achieved here.

Feedback results to staff. Their engagement in the co-construction of scale development has enhanced credibility and generalisability. In order to maintain this validity staff must remain engaged in the process of dissemination, feedback and action.

Test NHSEEI nationally. Rollout the NHSEEI simultaneously with any proposed continuation of staff survey to: establish construct validity between the two measures and to ascertain which measure staff prefer.

## Introduction

This is the final report pertaining to validation of the NHS Scotland Employee Engagement Index. Specifically it relates to objectives 2 & 3, italicised below, which are situated within the following aim and objectives.

*Aim*

To validate the proposed questions for an NHS Scotland Employee Engagement Index (NHSEEI), and make recommendations on the most appropriate way of constructing that Index.

*Objectives*:

1. Carry out psychometric testing on each question with the aim of determining whether it is appropriate for verifying the subject it is intended for. This was completed with an Interim report1 on 4th January 2013.
2. *Conduct a factor analysis using data from the January pilot to determine, which, if any, questions are correlated and the implications for deriving an Employee Engagement index. This should be completed by 1st March 2013.*
3. *Make recommendations for an Employee Engagement Index from the above findings. This should be completed in a Final Report by 1st March 2013.*

### Background

The purpose of the overall project is to measure staff engagement within NHS Scotland. The importance of staff engagement has recently been highlighted in the Francis report2. It is a fundamental principle of partnership as articulated by the revised Staff Governance Standards 3. Staff engagement therefore sits at the centre of all current health strategy4 5 and legislation6 and the evidence for its importance is explicit in Scotland’s Healthcare Strategy7 (p6):

*We already know from extensive research about the links between staff engagement and enhanced organisational performance.*

This translates to the following Healthcare Quality Outcome as articulated in NHS Scotland Chief Executive’s annual report 2011/128 (p9) :

**Staff feel supported and engaged**

*Staff throughout N.H.S.Scotland , and by extension, their public and third sector partners, feel supported and engaged, enabling them to provide high quality care to all patients, and to improve and innovate.*

The timeliness of this project, particularly in the wake of the Francis report2 cannot be overstated. A robust metric of staff engagement generated by the staff themselves is entirely coherent with the recommendations of that report. The NHS Scotland Employee Engagement Index (NHSEEI) has been designed to provide empirical data in support of the above agenda. This evaluative aspect of the project was tasked with utilising psychometrics to establish to what extent the latest version of the NHSEEI actually measured staff engagement.

Early indications were very promising. In December 2012 we delivered an interim report1 based on feedback and factor analyses of previous pilot data. From this analysis a five-factor structure emerged congruent with the underpinning theory of the questionnaire. That is, it was built from staff governance standards, the Macleod enablers, KSF core dimensions and the Health Care Quality ambitions (see figure 1). That these principles had been translated into meaningful questionnaire items by participating staff and the project team was a sign that

1. the embedding principles were present as factors within the questionnaire, and
2. the questionnaire items themselves were meaningful components of these principles to the staff that filled in the questionnaire

However, in the interim report1 we observed that some NHSEEI items did not fit into these factors as well as others. Because of the relatively small sample size we were unable to recommend excluding these less well fitting items as poor fit could have been an artefact of the particular staff sample. Therefore the only key recommendation for change that we made in the interim report was for some items to be reversed to avoid participants simply entering the same response for all items. The purpose of this was to offer a check on data quality whilst leaving the putative factor structure intact for further testing.

The other recommendation taken up from the interim report was the inclusion of a single scale measure associated with the following stem: ‘*overall, working within my organisation*’ followed by an 11 point scale ranging from ‘*is a very good experience*’ (10) to ‘*is a very poor experience*’ (0). The purpose of this was to ascertain any correlation between this one scale item and the overall score on the engagement index. If a correlation was found it could provide a useful standalone metric of engagement following completion of the questionnaire.

In order to examine these issues in a more robust manner a larger sample was required. This report analyses the results from this larger sample.



##### Figure 1. Drivers of staff engagement

## Method

### Data

The pilot 3 version of the NHSEEI (Appendix 1) was distributed electronically and paper copies to 2300 NHS Scotland staff across three boards: NHS Forth Valley, NHS Tayside and NHS Golden Jubilee. Participants completed the questionnaire online. Below is a list of the NHSEEI items in numbered order:

|  |
| --- |
| 1. I am clear what my duties and responsibilities are
 |
| 1. I get the information I need to do my job well
 |
| 1. I am given the time and resources to support my learning and growth
 |
| 1. I do not have sufficient support to do my job well
 |
| 1. I am confident my ideas and suggestions are listened to
 |
| 1. I am confident my ideas and suggestions are acted upon
 |
| 1. I feel involved in decisions relating to my job
 |
| 1. I am treated with dignity and respect as an individual
 |
| 1. I am treated fairly and consistently
 |
| 1. I get enough helpful feedback on how well I do my work
 |
| 1. I feel appreciated for the work I do
 |
| 1. My work gives me a sense of achievement
 |
| 1. I feel my direct line manager cares about my health and wellbeing
 |
| 1. My direct line manager is sufficiently approachable
 |
| 1. I have confidence and trust in my direct line manager
 |
| 1. I do not feel involved in decisions relating to my team
 |
| 1. I am confident performance is managed well within my team
 |
| 1. My team works well together
 |
| 1. I would recommend my team as a good one to be part of
 |
| 1. I understand how my role contributes to the goals of my organisation
 |
| 1. I feel my organisation cares about my health and wellbeing
 |
| 1. I do not feel senior managers responsible for the wider organisation are sufficiently visible
 |
| 1. I have confidence and trust in senior managers responsible for the wider organisation
 |
| 1. I feel involved in decisions relating to my organisation
 |
| 1. I am confident performance is managed well within my organisation
 |
| 1. I get the help and support I need from other teams and services within the organisation to do my job
 |
| 1. I would recommend my organisation as a good place to work
 |
| 1. I would be happy for a friend or relative to access services within my organisation
2. Overall, working within my organisation…(10= ‘is a very good experience’…0=’a very poor experience’)
 |

Questions 1-28 entailed six point Likert responses: ‘Strongly disagree’ ‘Disagree’ ‘Slightly disagree’ ‘Slightly Agree’ ‘Agree’ ‘Strongly Agree’. Question 29 entailed an 11-point scale from ten to zero. No identifiable data were requested or collected, but professional roles were obtained for sample descriptive and comparative purposes. Data were collated centrally by the project team and rechecked for the presence of any potentially identifiable data. Anonymised Microsoft Excel spreadsheets were then exported to the external validation team at UWS.

### Analytic procedures

The spreadsheet was imported into IBM SPSS Statistics (SPSS) version 20 for descriptive statistics and inferential analyses. All data were also converted to .csv files and .txt files, imported and recoded into WINSTEPS and jMetrik for item response analyses. Descriptive statistics entailed frequencies and summary graphical representations. The main methods of inferential analysis entailed factor analysis and Rasch analysis. They are described below.

Factor analysis is a particular approach to reducing the dimensionality of data. Using this study as an example, it can be seen that there is a minimum of 28 questions, excluding question 29, which is an overall rating question and therefore a summary of the whole rather than a specific dimension of it. Each one of these 28 items can be thought of as a particular response (dimension). Most humans have difficulty in viewing more than two dimensions, and therefore 28 dimensions are extremely difficult to make sense of. What could be done is to attempt to reduce the number of dimensions to make the overall response easier to understand. There are a number of ways of attempting to do this: Factor Analysis, Principal Component Analysis, Projection Pursuit, and Multidimensional Scaling. In SPSS there are three approaches offered; Factor Analysis, Correspondence Analysis, Optimal Scaling. Factor Analysis has been used widely, for example a standard tool used in Service Quality10, SERVQUAL, was established using Factor Analysis.

As a very simple explanation we reduce the problem as follows. Imagine that there were five questions to respond to, but a reduction was sought in this number. Suppose it was thought that there were two basic factors (in practice this is not known, and part of the tool set is to offer information on how many factors might be present). Thus we could think of the data as

Response 1 = aFactor 1 + bFactor 2 + Error

Response 2 = cFactor 1 + dFactor 2 + Error

:

:

Response 5 = iFactor 1 + jFactor 2 + Error

The a,b,c . . I,j are unknown constants and differ for each respondent; these are called the factor loadings. By some standard mathematics, we can identify the factors. Basically this produces an inverse of what we have above, giving

Factor 1 = mResponse 1 + nResponse 2 + pResponse 3 + qResponse4 + rResponse 5

Factor 2 = sResponse 1 + tResponse 2 + uResponse 3 + vResponse4 + wResponse 5

In practice it has been found that the picture usually becomes clearer if we make a rotation of the original dimensions at the start. The reason is that proceeding directly to factor analysis unrotated usually finds all the items loading onto the first couple of factors. Rotation avoids this and the outcome is that we usually can produce a marked reduction in dimensionality. Often we can “name” the new dimensions by considering how they are produced. Thus in Factor 1 above if m and p are large positive values and r is large and negative this may give us a way of understanding the “new” dimension. For example, this is what we did in the interim report to infer the presence of the factor structure illustrated in figure 1.

However, whilst very useful for identifying patterns within responses to items and hence drawing attention to underlying structure and correlations between items, factor analysis is limited in the amount of information it can give us in relation to individual item performance within a scale, or if the scale is the right one to be using for a particular sample. Item response theory helps in this regard.

All data were also imported into WINSTEPS and jMetrik for item response theory (IRT) analysis. The basic idea of IRT models is that there is an underlying trait that is reflected in the response to the test or survey items. This fits with the assumption that the NHSEEI is measuring the trait of staff engagement. Rasch analysis is a relative of IRT analysis11. In short, it uses the dataset of responses to construct probabilities on how likely a person is to get a particular item correct, and how likely individual items fit with the model constructed from this analysis. The endpoint is that the statistical technique calculates how ‘difficult’ particular questionnaire items are, and how well individuals perform on the overall test. Because both these measures are standardised to appear on the same scale we can then estimate the probability of a particular person endorsing a particular item based on their level of engagement. For example if the person ability and the item difficulty appear at the same point then this would indicate that that person has a 50/50 chance of endorsing that particular item.

Item response theory is also very useful for ascertaining how individual items perform within a questionnaire. Item characteristic curves (ICC) can be constructed within jMetrik to show the parameters at which a certain likelihood of Likert response should be expected. In other words it can illustrate whether the categories of Likert responses are a) necessary and b) how likely a certain participant would be to answer a particular question in a certain way (agree/strongly disagree etc). This information is very useful for identifying anomalies in both items and responses. The analysis section summarises this work.

In summary, descriptives, frequencies, correlations and principal component analyses were conducted on the dataset. Rasch analysis was conducted including item and person estimates. Item characteristic curves (ICC) were constructed to further ascertain the performance of individual items within the questionnaire. The purpose of these analyses was:

1. to describe and explain patterns within the responses to the questionnaire,
2. to ascertain whether the NHSEEI measured a unitary construct (staff engagement),
3. to identify and explain any factor structure underpinning the data,
4. to establish individual item ‘difficulty’ in order to identify the relative importance of the items,
5. to identify where items and/or respondents did not fit with the overall model,
6. to perform a close analysis on all individual items within the questionnaire.

From these analyses we can make a judgement on the validity of the NHSEEI in this sample.

## Results

### Descriptives: participants

1280 people completed the questionnaire in total, an excellent return of over 56% those staff targeted, and more than double the 27%[[2]](#footnote-2)\* return of the 2010 staff survey 12,13. If this return could be repeated nationally it would be a significant achievement in itself and considerably raise the credibility of the findings. Whilst all the sample areas deserve praise particular mention should go to Anne Benton and Forth Valley with a final return of 60%. The impact of this on the quality of the inferences we can make as a result is considerable. For example because of the large sample we could be robust in our treatment of missing data. In order to get the most coherent analysis we removed those responses where the questionnaire had only been partially completed or not completed at all. This left us a sample of 1193. The breakdown of this sample by profession is below in figure 2.

##### Figure 2. Sample by profession

The breakdown of the total sample is comparable with NHSScotland staff breakdown nationally (*Figure 3*). Although ISD do not collate figures on managerial and corporate staff their ratio of admin, medical, nursing and AHP staff in 2012 is comparable this sample14. Alongside the excellent response rate this adds further credibility to any generalisable claims arising from this analysis.

##### Figure 3. National NHS Scotland staff population breakdown

### Descriptives: responses

Response data were imported into IBM SPSS 20. For descriptive analysis data were coded as per box 1

|  |
| --- |
| *All NHSEEI items except 29* |
| 1 = Strongly Disagree2 = Disagree3 = Slightly Disagree4 = Slightly Agree5 = Agree6 = Strongly Agree |

#### Box 1. Item scoring

Summary of all responses per item, ranked by frequency of response to ‘strongly agree’ is in figure 4.

Figure 4. Summary frequency of all responses by Likert Category

### The NHSScotland Employee Engagement Index (NHSEEI)

In order to produce the NHSEEI measure for each individual we added the response scores for each item using the coding illustrated in box 2 then converted to a percentage. In brief this meant that the NHSEEI consisted of a representation of all the individual items scores, with the reverse items coded to incorporate their logical meaning; ie, not endorsing a negative was assumed to be positive.

|  |  |
| --- | --- |
| *All NHSEEI items except 4, 16, 22 & 29* | *NHSEEI items 4, 16 & 22 (reversed items)* |
| 1 = Strongly Disagree2 = Disagree3 = Slightly Disagree4 = Slightly Agree5 = Agree6 = Strongly Agree | 6 = Strongly Disagree5 = Disagree4 = Slightly Disagree3 = Slightly Agree2 = Agree1 = Strongly Agree |

Box 2. Logical scoring

Figure *5* below shows the mean NHSEEI by profession. It must be remembered that these groups are not numerically equivalent (figure 2). For example, although this graph looks good for managers (N=20) and less so for support staff (N=58) it is only a summary and requires further statistical analysis. This is returned to.

##### Figure 5. Mean NHSEEI (%) by profession

Finally in this descriptive section, we collated data on item 29: ‘*Overall within my organisation: scale 0(very poor experience) – 10 (very good experience)’.* In order to show any relationship between the NHSEEI and responses to item 29 the two response spreads were illustrated together. The total spread of grouped responses to NHSEEI is illustrated in figure *6.* The total spread of responses to Q29 is in figure *7*. The samples were checked for normality[[3]](#footnote-3)#, an important assumption to ascertain for further analysis. Even with the naked eye it can be seen that the distributions in figure *6* and figure *7* are comparable. This suggests that the measures may correlate, and the significance of this will be discussed in the next section.

#####  Figure 6. NHSEEI score by frequency

##### Figure 7. Responses to item 29: Overall within my organisation…

## Analysis

All items were coded as box 2 in SPSS and Winsteps for inferential analysis. For Rasch analysis in jMetrik the coding pattern was the same but ran from 0 – 5 instead of 1 – 6. This was because Rasch analysis needs a zero baseline in jMetrik.

#### Box 1. Item scoring

### Factor analysis

A factor analysis was run in SPSS. KMO and Bartlett’s test of sphericity were calculated to test for suitability of factor analysis in this data. Once established (p=0.000) a principal component analysis was then conducted with a varimax rotation based on eigenvalues greater than 1. Coefficients below .4 were suppressed.

This analysis generated six factors with eigenvalues greater than 1, although it is important to recognise the sixth factor was only just greater than 1 (1.003) and accounted for less than 3.5% total variance. Details are available on request from the evaluation team. The best explanation for these factor structures lies in the grouping of the items within the questionnaire. That is, all the factors entail clustered groups of sequential items. Factor one entails items 3 to 12 (with the exception of item 4, the reversed item). Factor two entails items 20 to 28 (with the exception of item 22, the reversed item). Factor 3 entails items 13 to 15. Factor four entails items 17 to 19. Factor five entails items 1 and 2 plus a negative correlation with item 4, and factor six entails the two remaining reversed items 16 and 22.

This would suggest that the factors expressed are largely a function of the categorical placing of the items within the questionnaire. Because the only real exceptions to this were the reversed items we removed these items and ran the analysis as before excluding items 4, 16 and 22. This solution generated four factors (table 1), which means that the three reversed items appeared to account for the addition of two factors in the original analysis. This time the factor grouping of the items was even more strikingly associated with the grouping of the items within the questionnaire. Factor 1 entailed all the items except item 1 (2 to 12) from the section about ‘I/me’. Factor 2 entailed all the items about ‘my organisation’ (items 20 to 28). Factors 3 (items 13 to 17) and factor 4 (Items 1, 18 & 19) entailed the remaining questions (figure 8). In summary, the factors are almost entirely explained by the category of question the item relates to. The difference between factors 3 and 4 can thus be explained by the fact that factor 3 relates to questions about ‘my line manager’ and how the manager manages the team whereas factor 4 is more about how the individual feels about ‘my team’. Item one is the only exception to this pattern. This item about clarity of role appears to sit more closely with attitudes towards ‘my team’ than issues about ‘I/me’.

So, by running the principal component analyses with and without the reversed items it appeared that in both cases the factors could best be accounted for by the categorical groupings of the items. In order to see if this pattern would hold in a slightly different analysis we ran a third principal component analysis with all items included but this time we fixed the factors at five. The rationale was to remove the sixth factor from the initial analysis. Recall this only accounted for 3.5% variance and had an eigenvalue just greater than 1 (1.003). Again, coefficients below 0.4 were suppressed. This solution showed identical factor item relationships to the previous analysis with the exception that factors one and two were swopped round. Otherwise the pattern was identical, with the addition of a fifth factor entailing solely the reversed items. Solution and figure are also available on request

In summary then, the factor interpretation is largely consistent with the item grouping within the questionnaire. In a sense this means that these patterns of responding are logical. However, this conclusion would be more robust if the items had been spread out throughout the questionnaire instead of being grouped together, as the possibility cannot be excluded that these patterns could be explained by individuals merely ticking the same boxes on each screen. The project team may wish to consider ungrouping these items to see whether the pattern still holds in future samples.

There is also the issue of the reversed items. They did not appear to be answered particularly illogically but still appeared discordant with the pattern seen elsewhere. For example we would expect to find a negative association between item 4 and the factor that accounts for all the other items in that grouping (items 2 to 12) if the factors can be explained by the place of the items within the questionnaire as a whole. In actuality it is best accounted for by association with factor five in the first PCA, and by association with the other reversed items (16 & 22) in the fixed factor PCA. The reversed items therefore blur rather than add clarity. This will be explained in the Rasch analysis next. But to summarise, the factor structure in this sample can be best accounted for by four factors:

1. My experience as an individual
2. My experience with my direct line manager
3. My experience with my team
4. My experience with my organisation

The circularity of this interpretation could be viewed as a concern. It could simply mean that respondents had not really differentiated between items in any particular section and answered them all identically. In order to examine this hypothesis we went back to the raw data and checked for these patterns. We used heat maps in excel to easily identify any response repetition. Whilst there were clearly a few instances of this it was not routine. Also, as will be shown in the Rasch analysis each individual item within these sections contributes something unique to the questionnaire (see table 2 and figure 11). This would not be the case if all the questions had been answered in the same manner. A more rational explanation therefore has to account for population variance, which adds credibility to the factor analysis interpretation being a function of the intended measure rather than an anomaly of scoring groupings. The fact that the questionnaire has been built from the bottom up by staff and the literature to measure these specific aspects suggests the questionnaire is measuring exactly what it was designed to measure.

|  |
| --- |
| **Rotated Component Matrixa** |
|  | Component |
| 1 | 2 | 3 | 4 |
| I am clear what my duties and responsibilities are |  |  |  | .595 |
| I get the information I need to do my job well | .588 |  |  |  |
| I am given the time and resources to support my learning and growth | .651 |  |  |  |
| I am confident my ideas and suggestions are listened to | .693 |  |  |  |
| I am confident my ideas and suggestions are acted upon | .720 |  |  |  |
| I feel involved in decisions relating to my job | .713 |  |  |  |
| I am treated with dignity and respect as an individual | .601 |  |  |  |
| I am treated fairly and consistently | .612 |  |  |  |
| I get enough helpful feedback on how well I do my work | .652 |  |  |  |
| I feel appreciated for the work I do | .643 |  |  |  |
| My work gives me a sense of achievement | .432 |  |  |  |
| I feel my direct line manager cares about my health and wellbeing |  |  | .820 |  |
| My direct line manager is sufficiently approachable |  |  | .853 |  |
| I have confidence and trust in my direct line manager |  |  | .828 |  |
| I am confident performance is managed well within my team |  |  | .528 |  |
| My team works well together |  |  |  | .803 |
| I would recommend my team as a good one to be part of |  |  |  | .769 |
| I understand how my role contributes to the goals of my organisation |  | .434 |  |  |
| I feel my organisation cares about my health and wellbeing |  | .670 |  |  |
| I have confidence and trust in senior managers responsible for the wider organisation |  | .784 |  |  |
| I feel involved in decisions relating to my organisation |  | .737 |  |  |
| I am confident performance is managed well within my organisation |  | .786 |  |  |
| I get the help and support I need from other teams and services within the organisation to do my job |  | .647 |  |  |
| I would recommend my organisation as a good place to work |  | .707 |  |  |
| I would be happy for a friend or relative to access services within my organisation |  | .626 |  |  |
| Extraction Method: Principal Component Analysis.  Rotation Method: Varimax with Kaiser Normalization. |
| a. Rotation converged in 12 iterations. |

#### Table 1. Principal component analysis with reversed items removed



##### Figure 8. Four-factor solution

IRT & Rasch Analysis*.*

Data were imported into WINSTEPS in a text file in order to construct analysis of item difficulty and person ability. The purpose of this was to see which items were more difficult than others and to estimate the comparable person ability range for the entire sample. This can be illustrated by constructing a person item map within WINSTEPS as per figure *9*.

Items 6 CATS 1.0.0

--------------------------------------------------------------------------------

 Persons MAP OF Items

 <more>|<rare>

 3 . +

 . |

 . |

 . |

 . |

 . |

 . |

 2 . +

 . T|

 # |

 ## |

 .### |

 .### |

 .##### S|T

 1 .####### + 24

 .############ | 16 4

 ########## |

 .########## |S 23 25

 .############ M| 10 3 6

 .######## | 11 21 7

 .####### | 26 27 5

 0 .####### +M 17

 .###### | 22

 .#### S| 28 8 9

 #### | 12 15 18 2

 .## |S 13 19 20

 .## |

 .# | 14

 -1 .# T+

 . |T

 . |

 . | 1

 . |

 . |

 . |

 -2 . +

 . |

 |

 |

 |

 |

 |

 -3 +

 <less>|<frequ>

 EACH '#' IS 10.

##### Figure 9. Item person map

The vertical axis is scaled in logits, which in the Rasch model represent a standardised scale of difficulty such that the difference between -2 and -3 is equivalent to the difference in difficulty between +1 and +2 and so on. This scale also applies to person ability. Figure 9 therefore shows that the items range in difficulty from -1.47 (relatively easy item 1) to +1.06 (relatively difficult item 24). Table 2 below shows item difficulty estimates by rank, with the most difficult item at the bottom.

|  |  |  |
| --- | --- | --- |
| Item |  | Item difficulty (easy to hard) |
| 1 | I am clear what my duties and responsibilities are | -1.47 |
| 14 | My direct line manager is sufficiently approachable | -0.8 |
| 20 | I understand how my role contributes to the goals of my organisation | -0.59 |
| 19 |  I would recommend my team as a good one to be part of | -0.55 |
| 13 | I feel my direct line manager cares about my health and wellbeing | -0.52 |
| 2 | I get the information I need to do my job well | -0.5 |
| 12 | My work gives me a sense of achievement | -0.47 |
| 18 | My team works well together | -0.44 |
| 15 | I have confidence and trust in my direct line manager | -0.43 |
| 8 |  I am treated with dignity and respect as an individual | -0.35 |
| 9 | I am treated fairly and consistently | -0.23 |
| 28 | I would be happy for a friend or relative to access services within my organisation | -0.23 |
| 22 | I do not feel senior managers responsible for the wider organisation are sufficiently visible | -0.14 |
| 17 | I am confident performance is managed well within my team | 0.06 |
| 27 |  I would recommend my organisation as a good place to work | 0.07 |
| 5 |  I am confident my ideas and suggestions are listened to | 0.16 |
| 26 | I get the help and support I need from other teams and services within the organisation to do my job | 0.19 |
| 21 | I feel my organisation cares about my health and wellbeing | 0.34 |
| 7 | I feel involved in decisions relating to my job | 0.35 |
| 11 | I feel appreciated for the work I do | 0.35 |
| 10 | I get enough helpful feedback on how well I do my work | 0.4 |
| 6 | I am confident my ideas and suggestions are acted upon | 0.42 |
| 3 |  I am given the time and resources to support my learning and growth | 0.49 |
| 23 |  I have confidence and trust in senior managers responsible for the wider organisation | 0.55 |
| 25 |  I am confident performance is managed well within my organisation | 0.62 |
| 16 | I do not feel involved in decisions relating to my team | 0.83 |
| 4 |  I do not have sufficient support to do my job well | 0.84 |
| 24 | I feel involved in decisions relating to my organisation | 1.06 |

#### Table 2. Item difficulty estimates by rank

Figure 9 shows that items are nicely clustered around zero with some items harder and some easy. This is a good spread of item difficulty. The person spread is also clustered below zero suggesting that this questionnaire is a good test of engagement in this sample.

However, this analysis doesn’t account very well for *degree* of response. In order to summarise all this data for both person ability and item estimates including Likert thresholds we need to use jMetrik.

### Item person map with Likert data

Figure 10 summarises all the data according to the Rasch model. It is similar to figure 9 but contains more information in relation to degrees of response. It is a snapshot of person ability within this sample contrasted with the likelihood of a particular Likert response. The person density on the left is equivalent to the information given in figure 9. The item frequencies on the right of the diagram represent the step parameters (item difficulty + category threshold) in the Likert scales. The category threshold is the point where the item characteristic curves for two adjacent categories intersect. It is where responding in adjacent categories is equally likely. In the NHSEEI there are five thresholds (five step parameters) because the questionnaire has six-category items. The item frequencies represent the frequency of the step parameters. The light blue bars represents the step parameter between ‘strongly disagree’ and ‘disagree’ and the dark blue represents the step parameter between ‘agree’ and ‘strongly agree’. Pink (disagree/slightly disagree), yellow (slightly disagree/slightly agree) and green (slightly agree/agree) represent the middle three parameters.



##### Figure 10. Item person map with Likert parameter thresholds

Figure 10 therefore adds weight to the conclusion that this questionnaire is an excellent measure of engagement in this sample. The questionnaire has items that can differentiate even at the extremes. Because of the spread of the item difficulty it means there are questions in NHSEEI that can be answered by everyone and some questions only by the most engaged. Basically this means the NHSEEI is doing what it is supposed to be doing: differentiating between less and more engaged staff. Nobody scored at the maximum or minimum and because the items all tap into slightly different aspects of staff engagement (table 2, figure 11 & 12) it suggests that it is a functional discriminatory tool between people.

Whilst this mainly adds probabilistic data its more general function is that it adds credibility to deductions we have already made by demonstrating that using different statistical techniques we arrive at the same conclusion. We have now analysed the sample in SPSS, Winsteps and jMetrik using various techniques and produced mutually supporting results.

However, as well as factor analysis and item difficulty we also need to examine item *fit*. This is done to ascertain the overall behaviour of the individual items in relation to the measurement of the putative underlying trait. In order to see how well the items fit with the underlying trait of staff engagement we can construct a bubble chart. This illustrates not just the item difficulty (table 2) on the vertical axis but also the item ‘fit’ with the overall model on the horizontal axis. See figure 11.

Infit mean square

##### Figure 11. Bubbleplot of all 28 items.

We can interpret figure 11 as follows. Infit mean square of one indicates the mean value of ‘staff engagement’. More or less than 1 indicates more variation than would be expected by the Rasch model. For example 1.3 is 30% more variation. 0.78 is 22% less variation. Response strings nearly always show some variation and this is no problem. In fact it is this activity that allows us to accept the factor structure with more confidence, given that this shows factor groupings are more likely to be a function of diverse response patterns as opposed to an artefact of the same individuals answering different items in the same way (just ticking the same response to every question on a particular page). People respond to easier items as if they were more difficult and vice versa as these judgements of difficulty come from general estimates and individuals vary greatly.

Nevertheless it can be seen that some of these items may be problematic. Items 16, 4 and 22 clearly don’t fit the model at all. They ‘underfit’ the model and therefore need to be reviewed. Items 6 and 27 on the other hand may fit the model too well and therefore artificially inflate reliability. The technical implication of ‘overfit’ is that we may conclude that our questionnaire is better than it actually is. Item 6 is almost too good to be true 15. We therefore need to examine these issues in more detail.

### Detailed analysis of each question.

Each question was individually analysed by reference to the following:

* The frequency of responses in each Likert category (from figure 2 and figure 10)
* The factor the item most closely correlates to (from table 1 and figure 8)
* The difficulty of the item and the item fit (from figure 9 and figure 11)
* A synthesis of these analyses

The purpose of gathering this data together was to facilitate deeper analysis of each question and its place within the questionnaire as a whole. The frequencies, fit statistics and factor associations have were considered together for a focused analysis. Each question was therefore analysed in this level of detail, and a full breakdown of this detailed analysis by question is available on request to the evaluation team. The body of this report focuses on the synthesis and key conclusions from this work. Here are the main findings:

* The reversed items 4, 16 and 22 do not fit the model and the reversal should be removed
* Item 14 also underfits the model. Items 6 and 27 overfit the model. However, because of their places within the factor structure these items should be retained
* The rest of the items fit very well and all contribute uniquely to the NHSEEI.

### Summary synthesis of item analysis.

Figure 11 shows the vast majority of items lie between plus or minus 0.3 of the mean. This means they all fit with the model of staff engagement and are all measuring something of the same concept. Likewise there is enough variation in these items to warrant confidence that the items all contribute something unique to the whole. The problematic items are in the minority.

Much of the problem of fit appeared to be related to the reversed items. These items factored together (see section on factor analysis) yet none fit with the overall model. When they were removed and the analyses rerun without them the majority of the overfit problems disappeared (figure 12). Item 14 stands out as a consistently difficult and original question, and items 12-19 are on the edge of acceptable fit. This is not to say they are irrelevant to staff engagement. What this tells us is that it they are measuring a unique aspect of it that the Rasch model can’t account for.

##### Figure 12. Bubbleplot with reversed items removed

In order to consider whether we should keep items 12, 13, 14, 15, 18 and 19 (circled in figure 22) based on the Rasch analysis we need to go back to the factor analysis. Remember we had a four-factor structure to explain the principal component structure:

1. My experience as an individual
2. My experience with my direct line manager
3. My experience with my team
4. My experience with my organisation

Items 13 – 19 entail factors 3 & 4 (figure 8). If we remove them because of issues with item fit we would not just be removing 5 misfitting items but more significantly we would be removing 2 of the four factors we found in the principal component analysis. We would be claiming that items about my direct line manager are irrelevant to staff engagement. This would be unjustifiable, and shows why summary statistics always need to be interpreted with caution. Item 12 is a little more problematic to justify from a purely psychometric perspective. It is only moderately associated (.432) with factor 1 and therefore doesn’t add a great deal of extra information to ‘my experience as an individual’. It is a poor fit with the Rasch model. However, it is a very interesting item in itself. As well as being mainly associated with factor 1 is also associated with factor 2 (.341) and factor 4 (.403) suggesting it contributes widely but not selectively to most factors. It is not easily pigeonholed, but ‘my work gives me a sense of achievement’ is also clearly important information to understand.

In summary, from detailed individual item analysis based on a variety of measures and analyses we would recommend that all 28 items are kept for the final NHSEEI, but that the reversed items revert to positive statements in line with the rest of the questionnaire. Illogical responses to these reversed items offer the most salient explanation of the erroneous behaviour of these items within both the factor and Rasch analyses. No other items demonstrate this behaviour.

### Is engagement related to role?

We saw from figure 3 that there appeared to be differences between the levels of engagement according to role. Here we analysed these differences statistically. Analysis of variance (one way ANOVA, table 3) was run for all groups and was highly significant (p<0.000):

|  |
| --- |
| **ANOVA** |
| Total NHSEEI |
|  | Sum of Squares | df | Mean Square | F | Sig. |
| Between Groups | 22185.427 | 7 | 3169.347 | 7.804 | .000 |
| Within Groups | 472344.494 | 1163 | 406.143 |  |  |
| Total | 494529.921 | 1170 |  |  |  |

#### Table 3. ANOVA in profession

This result is mainly accounted for by the managers’ high scoring and the support staffs’ low scoring. In this sample the managers are the most engaged and the support staff the least, and these findings are statistically significant.

### How reliable is this finding?

Recall on page 16 we showed the distribution of responses to item 29 alongside the distribution for total score NHSEEI. Here we statistically analysed these variables and found a clear correlation (table 4) between item 29 (Overall working within my organisation is 0-10) and the NHSEEI

|  |
| --- |
| **Correlations** |
|  | Overall, working within my organisation.... | Total NHSEEI |
| Overall, working within my organisation.... | Pearson Correlation | 1 | .781\*\* |
| Sig. (2-tailed) |  | .000 |
| N | 1193 | 1193 |
| Total NHSEEI | Pearson Correlation | .781\*\* | 1 |
| Sig. (2-tailed) | .000 |  |
| N | 1193 | 1193 |
| \*\*. Correlation is significant at the 0.01 level (2-tailed). |

#### Table 4. Correlation between NHSEEI and item 29

Pearson’s correlation is significant (0.781, p<0.000) inferring a very strong association between the two measures. This means the self-rated score for item 29 is a strong indicator of degree of staff engagement and could therefore provide a useful summary metric. We therefore repeated the ANOVA using item 29 as the independent variable and got a similarly significant result (p<0.000) in relation to finding statistical differences between professions in terms of engagement.

####

#####

##### Figure 13. Mean score for item 29 by profession

|  |
| --- |
| **ANOVA** |
| Overall, working within my organisation.... |
|  | Sum of Squares | df | Mean Square | F | Sig. |
| Between Groups | 242.742 | 7 | 34.677 | 8.382 | .000 |
| Within Groups | 4811.429 | 1163 | 4.137 |  |  |
| Total | 5054.171 | 1170 |  |  |  |

#### Table 5. ANOVA by profession using item 29

Again this result can be mainly accounted for by the positive mean score of the managerial staff (7.35) and the least positive score attributable to the support services (5.36). The fact that these results were comparable adds credibility to the inference that these are meaningful differences. Whilst the sample size of managers could have been cited as a weakness of this inference it is nevertheless a statistically significant finding from two different (but related) measures. Managers in this sample are significantly more engaged than their colleagues. Support services significantly less so. However, recall that even their scores are positive.

## Discussion

This section interprets these various analyses. It contextualises this interpretation by reflecting on the underpinning theory of the questionnaire and its relationship and importance to the inferences we have made. It finishes by making recommendations for further study. First a quick recap of the main findings within this study.

### What this study found

* The response rate was excellent at just over 56%. 1193 responses were included for analysis.
* According to both NHSEEI and item 29 the most engaged profession were the managers. The least engaged were support services.
* The four main factors established through principal component analysis mirrored the structure of the questionnaire:
1. My experience as an individual
2. My experience with my direct line manager
3. My experience with my team
4. My experience with my organisation
* Factor analysis is best explained by the item placement within the questionnaire. Because of the variation in responses and the strength of the underpinning theory this is a sign of reliability.
* Rasch analysis showed the majority of items fit the model of staff engagement very well with the following exceptions:
* The reversed items 4, 16 and 22 do not fit the model. Items 12, 13, 14, 15 & 18 underfit the model. Item 6 overfits the model
* Factor analysis accounts for the anomalous fit statistics with the exception of the reverse items. These should be amended.
* The rest of the items show enough variation to contribute uniquely to the overall construct of staff engagement.

Before we can make any recommendations from these findings it is important to consider the strengths and weaknesses of the analytic approach we have adopted here.

### Context and theory: why these findings are valid

The credibility of inferences that can be drawn from any questionnaire is a function of the robustness of the initial conceptual model 16. A valid scale is one that allows us to make accurate inferences about someone. What this means in relation to recommendations for this scale is that the central question always remains:

 ‘Do the results of this study allow us to draw the inferences about the people that we wish to make’ 17

The importance of this apparently straightforward statement cannot be overstated. It impacts upon scale development by focusing attention on the *purpose* of the scale 18. Validation is a unitary construct that tests the theory and the measure at the same time 17. This in turn highlights the importance of grounding the NHSEEI development in a sophisticated and testable conceptual model of what staff engagement is and how it is expressed, understood and evaluated. Because this scale is theoretically grounded, developed by staff and modified through a process of consultation over a series of robust cycles the results of this study *do* allow us to draw the inferences we wish to make. Therefore, when we scrutinise the factor structure or item fit we can be confident that these issues are conceptually rich.

### A justification for factor analysis

The goal of factor analysis is to look at all the pairwise relationships (correlations) between variables (questionnaire items) and identify those variables that correlate most closely with each other. The purpose of this is to group the items into ‘factors’. Once we have done this we need to interpret what these factors mean. Because the questionnaire was built from the principles in figure 1 we would hope that factor analysis would show us that our groups of items do in fact map on to these principles. If they do then the factor analysis has shown that the underpinning structure is consistent with our theoretical position. The questionnaire measures what we hope it is measuring.

However, in a sense this is circular logic. We are bound to interpret the factors as expressions of the principles because those principles drove the development of the questionnaire in the first place. This is why we tried to interpret the factor structure without reference to underpinning theory in the interim report. It was subsequently possible to show that this interpretation was also consistent with the original drivers. This principle holds in this final report, although the factors are best explained by their categorical placement within the questionnaire. Nevertheless this shows a limitation of factor analysis and is why we went further and used Rasch analysis to deepen our understanding of the performance of the items within questionnaire.

### A justification for Rasch analysis

Rasch analysis takes a very different approach to factor analysis in that it begins by assuming we are measuring a unitary construct. One of the purposes of Rasch analysis is to test this assumption. So instead of looking for factors we are looking to assess the degree to which all the items measure the same underpinning trait. This is called the latent trait in the Rasch literature11 and the latent trait we are referring to here is staff engagement. The Rasch model tests all the items and all the people taking the test against each other at the same time. The purpose is to see which items (and people) fit the model and which don’t. The starting point is to calculate the proportion of items answered correctly by each person, and the proportion of people successfully answering a particular item. These raw score totals allow us to estimate person ability and item difficulty. Person ability and item difficulty are sufficient to allow a calculation of the odds of success for a particular person on a particular item. Because we are using odds we can convert these odds into a log scale, and converting odds into a log scale allows for standardisation of expression for item difficulty and person ability, hence we can represent them on the same scale.

### Factor analysis and Rasch analysis: complementary or contradictory?

By using different statistical methods we are essentially viewing the data from different standpoints. This is fundamentally useful as it gives us a better picture than relying on one technique19. However, from a statistical perspective caution must be maintained when generating assumptions grounded in different measurements. On the positive side there is increasing credibility in combining Rasch analysis with confirmatory factor analysis to study dimensionality, as we have done here20. In essence no single model can detect all possible sources of misfit and so it makes sense to view any dataset from multiple angles21. Factor analysis and Rasch analysis both provide evidence to support inferences regarding invariance within a particular context, and in this study they offer complementary explanations for the findings.

On the negative side we must remain aware that these inferences are logically discrete. That is, Rasch and factor analysis are embedded in different philosophical deconstructions of what constitutes measurement. Because of these different assumptions they are strictly speaking, incompatible, mutually exclusive models.22 However, even at this level of philosophical deconstruction common sense prevails. For example Saltzberger states that

 ‘*proper scale development and analysis should never be confined to a statistical procedure (even if that procedure utilizes the Rasch Measurement), but should be guided by a theory of the construct to be measured’* 22 (p1376).

This brings us back to our discussion on context and theory. Rasch analysis and factor analysis help us understand the data by viewing it in different ways and both are very useful. A judgement does not need to be made on logical compatibility. It is not the techniques of data analysis that drive the credibility of the findings but the quality of the data and its underpinning theory. Because the data is of such high quality we can be reasonably certain that where Rasch and factor analysis illuminate similar patterns then these patterns are generalisable.

### Where are the weak points?

The main psychometric problem was created by attempting to eliminate nonsensical responses. Reversing some items to ascertain logical incompatibility is a standard method of data cleaning. We were aware that reversing some items does not eliminate this problem altogether23 and this in fact turned out to be the case. The reversed items 4, 16 and 22 ended up as some of the least well fitting items according to the Rasch analysis. Lack of fit can of course be explained by irrational responding and so there is a possibility that these items may not have provided the check against irrationality they were designed to avoid. There is also the possibility that these items may have been answered entirely rationally but that they are not the best indicators of staff engagement. However, a check of the raw data would suggest that according to the pattern of responses to other questions, people had spotted the reversal of these items. It is difficult to know what to suggest in these cases, except to acknowledge that people are inherently irrational24 and thus mitigating this may not be entirely possible through any technique other than ascertaining large samples.

### Item 29: A global metric?

A suggestion made by the evaluation team after the interim report was the possible addition of a simple scale to ascertain level of engagement. This method has a successful history in other areas of psychometrics such as the distress thermometer for example25, where certain cut-offs on the scale are reliable indicators of severity of clinical distress. In pilot 3 we therefore added the global engagement question 29: ‘*overall working within my organisation is a (0 – 10 scale) experience*’. If successful this could provide a very quick standalone measure. Because the concept of engagement is so broad this type of question would only make sense at the end of the questionnaire. That is, the questionnaire could *not be replaced* by such a measure, but there is a strong possibility that such a measure may *correlate* with the questionnaire. As such it could be an important standalone metric to convey a global measure of engagement.

This is in fact what happened in this sample. There was a very strong correlation between the NHSEEI and item 29 (table 4). This is an important finding and suggests this question should be retained. Because item 29 follows 28 valid items questioning staff engagement it is highly likely this is an accurate metric of any particular individual’s global judgement of their engagement. Because it correlates so well with the NHSEEI item 29 may have a function for reporting and data validation/cross checking purposes. However, we would strongly caution against this metric ever *replacing* the NHSEEI. The NHSEEI can be just as effectively reduced to a single score for ease of reporting and this single score entails information from 28 validated items generated by NHS Scotland staff themselves, not just a single global judgement.

### What did NHSScotland staff think of the NHSEEI?

Following this final pilot three focus groups were held with participating NHS boards. The following is a very brief summary of themes.

* The issues of anonymity persisted in small teams. If people do not feel they can be genuinely anonymous then there remains a potential for disengagement with process. However, it was interesting to note that even these cautious individuals recognized and supported the necessity of the exercise
* Participants sometimes missed the managers’ name, even though it was specified on the questionnaire. Some participants wanted to talk about a different supervisor instead. Some managers were unhappy about being named but again all saw the necessity and benefit of the process.
* Focus group participants were all very positive about the style and layout of questionnaire. They were particularly impressed with the short time it took, mainly in relation to comparable time-consuming questionnaires they had previously completed. Despite one or two comments about reducing the scale the six point scale was broadly popular.
* One or two IT issues persisted with access. However there may be a perceived advantage in retaining a paper version as it mat assuage anxieties regarding anonymity
* A few semantic issues were raised despite the provided definitions. For example some participants remained unclear as to what constituted ‘organisation’ or ‘sufficiently visible’.
* Reverse items were spotted and some valued them for their ability to stop people and make them think. By contrast other participants suggested they were unnecessary as they introduced confusion.
* One participant suggested putting a timescale on the questionnaire (eg ‘in the last 12 months…) and this fits with PROMs literature on specifying timescale26
* There was also consensus that the scale in item 29 was the wrong way round and should be worded positively. At the moment it goes from 10 – 0. It should be the other way round
* Finally there was a very strong message that these results need to be published and fed back to participants for action. Otherwise the NHSEEI will come to be seen as a ‘paper exercise’ and disengagement will ensue.

In conclusion then, the NHSEEI is a robust measure of staff engagement, meaningful and important to staff. Its coproduction with the staff has facilitated a theoretical credibility which in turn has generated psychometric validity and positive practical consequences. The outstanding response rate suggests that there is a genuine desire for staff voices to be heard and that they have endorsed this measure as a means to that end. In psychometric terms the Rasch analysis showed that most of the 28 items fit with the model of staff engagement generated by the staff and project team. With the exception of the reversed items the factor analysis then accounted for those items that did not appear to fit so well. This means that the NHSEEI is a valid and reliable measure of staff engagement consistent with its scope and purpose. We propose the following recommendations.

## Recommendations

Remove the reversal of items 4, 16 and 22, otherwise leave the 28 items as they are. Add a timescale to specify the response period.

Keep item 29 but reverse the scale to go from 0 – 10 instead of 10 – 0. Responses to question 29 should be utilised as a snapshot for crosschecking levels of staff engagement and ascertaining response logic. They should not be used as a standalone metric.

Keep sample sizes comparably high in subsequent iterations to maintain the level of credibility achieved here.

Feedback results to staff. Their engagement in the process of scale development has enhanced credibility and generalisability. In order to maintain this validity staff must remain engaged in the process of dissemination, feedback and action.

Test NHSEEI nationally. Rollout the NHSEEI simultaneously with any proposed continuation of previous staff survey to:

* establish construct validity between the two measures and to
* ascertain which measure staff prefer.

If you need to discuss any part of this report we would be delighted to hear from you

|  |  |
| --- | --- |
| Dr Austyn SnowdenReader in Mental Health | Dr Ewan MacArthurSenior Lecturer in Mathematics |
| austyn signature.bmp | Description: signature |

## References

1. Snowden, A. & MacArthur, E. *Validation of the NHS Scotland Employee Engagement Index: Interim Report*. 40 (2013).

2. Francis, R. *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry Executive summary*. (London, 2013).

3. NHSScotland *Staff Governance Standard. A Framework for NHSScotland Organisations and Employees, 4th edition*. 22 (Edinburgh., 2012).

4. The Scottish Government Rights , Relationships and Recovery : Refreshed Action plan 2010 - 2011. *Nursing* (2010).at <http://www.scotland.gov.uk/Resource/Doc/924/0097678.pdf>

5. The Scottish Government *Safe and Well at Work: Occupational Health and Safety Framework for NHSScotland*. (2011).

6. The Scottish Government Patient Rights (Scotland) Act 2011. (2011).at <http://www.legislation.gov.uk/asp/2011/5/schedule>

7. The Scottish Government *The Healthcare Quality Strategy for NHSScotland*. *Network* (Edinburgh, 2010).at <http://www.scotland.gov.uk/Resource/Doc/311667/0098354.pdf>

8. The Scottish Government *NHSScotland Chief Executive’s Annual Report*. (2012).

9. Hellerstein, J.M. *Principles and methods of data cleaning*. 1-42 (2008).

10. Parasuraman, A., Berry, L.L. & Zeithaml, V.A. Refinement and reassessment of the SERVQUAL scale. *Journal of Retailing* **67**, 420-450 (1991).

11. Bond, T.G. & Fox, C.M. *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*. 339 (Routeledge: New York, NY, 2007).

12. Bacon, N. & Hoque, K. NHS Scotland Staff Survey 2010 Highlights report. (2011).

13. Information Services Division NHS Scotland Workforce. **6**, (2012).

14. Information Services Division NHS Scotland Workforce. **6**, (2012).

15. Bond, T.G. & Fox, C.M. *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*. 340 (Routeledge: Oxon, 2007).

16. Pilkonis, P.A., Choi, S., Reise, S.P. & Stover, A.M. Item banks for measuring emotional distress from the Patient Reported Outcomes Measurement Information System (PROMIS): Depression, Anxiety and Anger. *Assessment* **18**, 263-283 (2011).

17. Streiner, D. & Norman, GR. *Health Measurement Scales: A Practical Guide to their Development and Use.* (Oxford University Press: New York, NY, 2008).

18. Cornwell, J. Why we need a national framework for patient experience. (2012).at <http://www.kingsfund.org.uk/blog/patient\_exp.html>

19. Risjord, M., Moloney, M. & Dunbar, S. Methodological Triangulation in Nursing Research. *Philosophy of the Social Sciences* **31**, 40-59 (2001).

20. Engelhard, G. Transactions of the Rasch Measurement SIG. **26**, (2012).

21. Randall, J. & Engelhard, G. Using Confirmatory Factor Analysis and Rasch Measurement Theory to Assess Measurement Invariance in a High Stakes Reading Assessment. *Applied Measurement in Education,* **23**, 286-306. (2010).

22. Saltzberger, T. The Rasch model and factor analysis: Complementary or mutually exclusive? *Rasch Measurement Transactions* **26**, 1376-8 (2012).

23. Swain, S.D., Weathers, D. & Niedrich, R.W. Assessing Three Sources of Misresponse to Reversed Likert Items. *Journal of Marketing Research* **45**, 116-131 (2008).

24. Kahneman, D. *Thinking Fast and Slow*. (Penguin: London., 2011).

25. Mitchell, A.J. Short screening tools for cancer-related distress: a review and diagnostic validity meta-analysis. *J Natl Compr Canc Netw* **8**, 487-494 (2010).

26. Fries, J., Rose, M. & Krishnan, E. The PROMIS of better outcome assessment: responsiveness, floor and ceiling effects, and Internet administration. *The Journal of rheumatology* **38**, 1759-1764 (2011).

Appendix 1. NHSEEI

##

1. http://www.scotland.gov.uk/Publications/2012/06/9560/5 [↑](#footnote-ref-1)
2. \* (42,061/154424) [↑](#footnote-ref-2)
3. # Kolmogorov-Smirnov both significant p<0.000 [↑](#footnote-ref-3)