Report on the Validity and Reliability of the Dutch Residency Educational Climate Test (D-RECT)

As deployed by the Medical Council in the National Trainee Experience Survey 2014

Dr. Deirdre Bennett MB MA MPH FRCPI, Senior Lecturer in Medical Education[1]

Dr. Siun O'Flynn MB BMedSc FRCPI, Head of Medical Education[1]

Dr. Margaret O'Rourke BA MA MSc PhD CPsychol CSci, Senior Lecturer[1]

Dr. Sean Hammond BA MA PhD, Senior Lecturer[2]

Medical Education Unit, School of Medicine[1] and Department of Psychology[2],

University College Cork

Background:

The Dutch Residency Educational Climate Test (D-RECT) is a 50 item scale for the evaluation of postgraduate clinical learning environments. The Medical Council used this scale as part of its Your Training Counts National Trainee Experience Survey 2014. The scale has been developed and validated for use in the Netherlands by Boor et al(1). Boor et al developed their scale on the basis of previous qualitative work on postgraduate clinical learning environments in the Netherlands (2,3) and a subsequent modified Delphi method with Dutch experts was used to refine items included. Boor et al then went on to test the construct validity of the instrument, using both Exploratory and Confirmatory Factor Analysis(1). D-RECT was shown to measure clinical learning environment along 11 distinct dimensions, or subscales, and to do so in a manner that is consistently reproducible. Generalisability testing was also undertaken to examine the number of respondents at various levels, department, hospital site etc. required to generate a reliable measure.

Transferring the D-RECT instrument to a new setting, that of Irish Postgraduate Medical Education and Training, requires re-assessment of the construct validity and reliability data referred to above. D-RECT has not, as yet, been validated for use outside the Netherlands. Whilst the features of good clinical learning environments are likely to be very similar internationally, it is possible that some of the individual items within D-RECT will be interpreted differently in the Irish context, due to differences in how training is conducted and in how healthcare is delivered. The terminology used in some of the items in D-RECT has been slightly adjusted in recognition of this by the Medical Council. Demonstration of validity and reliability of the instrument in the Irish context is essential for quality assurance of the learning environment data generated by the Your Training Counts survey.

Objective:

The project team were commissioned by the Medical Council to describe the reliability and validity evidence for the use of D-RECT in the Irish context, based on responses to the Your Training Counts National Trainee Experience Survey 2014.

Team Expertise:

The project team has a track record of excellence in Medical Education research. Dr. Bennett, project lead for this proposal from UCC, has published work using the DRECT to survey trainees of all levels under the Royal College of Physicians of Ireland(4). As such she has familiarity with the instrument and its factor structure. The team has extensive experience of learning environment tools in the undergraduate setting (5,6) and has published validity data on one such tool, the Dundee Ready Educational Environment Measure (7). The psychometric skills of the team are also evidenced in the development of a tool to measure undergraduate medical student stress, the Medical Student Stress Profile (8).

What is reliability?

D-RECT requires respondents to interpret 50 statements (or items) and indicate a level of agreement with each statement. The reliability of an instrument, such as D-RECT, refers to its ability to produce consistent scores under consistent conditions. An alpha coefficient (α), which can range from 0 to 1,

is used to indicate the degree to which individual items are consistent across the test. An alpha coefficient of 1 means total consistency: in other words, you would expect that if the clinical learning environment is good, that the individual items in a test to measure the clinical learning environment would be consistently be highly rated. We would not expect them to be totally consistent however, as within the clinical learning environment there are separate elements, or factors, which may vary. In considering acceptable reliability, the purpose of data collection and the need for precision for that purpose are taken into account. Reliability is a relative rather than absolute concept; however, an alpha coefficient >0.8 is generally considered to be a more acceptable measure of reliability.

What is validity?

The validity of a questionnaire refers to evidence that it is measuring what it claims to measure. Validity evidence can be derived from multiple sources(9):

- Evidence based on the content of the questionnaire;
- Evidence based on responses to the questionnaire items;
- Evidence based on the structure of the questionnaire; its reliability and analysis of the relationships between the responses to the individual items in the instrument.

Reliability and Validity of D-RECT in the Your Training Counts Survey

1572 responses to D-RECT were provided to the research team by the Medical Council. 596 or the responses were incomplete, due either to missing items (n=83), items deemed by trainees to be non-applicable to their current post (n= 495), or a combination of both (n=18).

General Applicability of the Items

When the D-RECT is used across different training contexts not all of the items are equally applicable. This means that in some case items will need to be skipped or unanswered. This was the case with the majority of 'missing items' in the YTC dataset; they were deemed 'not applicable'. A statistical response to this situation is to treat such items as missing and then apply imputation for missing data. The problem with this approach is that the items are not, in fact, missing but are not applicable. The 'missing' nature of such items is not random but systematic. In this case we follow the strategy of replacing 'missing' values with the mean of the 'non- missing' values. This means generating the scale score as the mean of the valid responses. In this way each respondent is directly comparable with others even though they may have answered fewer items. This procedure is predicated on the assumption that the items have all been sampled from the same domain of items and that they can be demonstrated to be highly consistent with each other. We will return to this issue later.

Reliability

The full set of D-RECT responses (50 questions) are used to examine the overall clinical learning environment. The responses to D-RECT collected in the Your Training Counts survey were shown to be a very reliable single measure, or total rating of satisfaction, in relation to overall clinical learning environment (α=0.96).

D-RECT has been designed to measure 11 distinct dimensions of the clinical learning environment using subscales, each comprising 2-8 questions. These are shown in Table 1 below along with an alpha co-efficient for each. The reliability of the D-RECT subscales is lower than the total scale (α coefficient range 0.67-0.92). The majority of these subscales have acceptable reliability ($\alpha > 0.8$) but some subscales are less reliable. Specifically these are *Work adapted to competence* (α=0.78), *Supervision* (α=0.75), and *Peer collaboration* (α=0.67). The reliability of these subscales is toward the lower end of acceptable reliability, although they are broadly consistent with the Dutch findings.

This subscale structure serves largely to demonstrate the construct validity of the general satisfaction domain and is less helpful in diagnostic terms. Nevertheless the psychometric properties appear quite persuasive. The total score appears psychometrically robust with an alpha coefficient of 0.97. The main finding here supports the notion that the items are drawn from a highly consistent item domain. Our tactic of using the mean of valid responses for scoring the scales is supported.

Subscale	N of Items	Raw Mean	Standard Deviation	Item Mean	Alpha
Supervision	3	10.92	2.69	3.64	0.75
Coaching and Assessment	8	22.66	7.31	3.22	0.92
Feedback	3	7.91	3.18	2.66	0.83
Teamwork	4	15.33	3.13	3.83	0.82
Peer Collaboration	3	11.60	2.17	3.88	0.67
Professional Relations	3	9.70	2.62	3.25	0.81
Work Adapted to Competence	4	13.61	3.39	3.41	0.78
Consultants Role	8	27.94	5.34	3.52	0.83
Formal Education	4	15.38	3.21	3.84	0.83
Role of Speciality Tutor	6	20.16	5.11	3.38	0.86
Patient Sign Out	4	13.19	3.45	3.30	0.81
Total	50	170.83	31.72	3.44	0.96

Table 1.

Spearman-Brown Prophecy Formula

Because we advocate a scoring method in which the mean of valid item responses are taken as the scale scores, a further analysis is indicated to show how many non applicable items are acceptable for scoring each scale. We used the Spearman-Brown Prophecy formula to estimate the number of items required to produce a score reliability of 0.80. It is clear that for the smaller subscales non applicable items will render the scale scores unreliable. In fact some subscales do not reach this criterion and should instead have items added. The ratio column in Table 2 (below) describes the proportion by which the item numbers may be modified to ascertain an alpha of 0.80. Ratios over 1 indicate scales in which more items are required to meet this criterion. The column Opt shows the rounded-up number of items required to meet the alpha of 0.80 while the column Tol indicates the tolerance for non-applicable items. A minus number in this column indicates that more items not less are required. This clearly shows the inadvisability of relying upon the 11 subscale scores as many of the scales are inadequately furnished with items. Nevertheless, the total score is very tolerant of non-applicable items.

Subscale	N of Items	S-B Ratio	Opt	Tol
Supervision	3	1.333	4	-1
Coaching and Assessment	8	0.348	3	5
Feedback	3	0.819	3	0
Teamwork	4	0.878	4	0
Peer Collaboration	3	1.970	6	-3
Professional Relations	3	0.938	3	0
Work Adapted to Competence	4	1.128	5	-1
Consultants Role	8	0.819	7	1
Formal Education	4	0.819	4	0
Role of Speciality Tutor	6	0.651	4	2
Patient Sign Out	4	0.597	3	1
Total	50	0.167	9	41

Table 2.

Validity

The validity evidence relating to the content of D-RECT is found firstly in the manner of its original development. The items used are derived from expert opinion, previous research and theories of how people learn in workplaces. This theoretical foundation suggests that the D-RECT has content validity beyond its original context of use. Content validity for the Irish context is also supported by the prior use of D-RECT for measurement of clinical learning environment, with a subset of the trainee group surveyed in Your Training Counts (4). Furthermore, the consultation exercise undertaken by the Medical Council in relation to selection of the D-RECT instrument and the expert bench-marking process, strongly support the validity of the content of D-RECT for use in Ireland.

D-RECT scores were shown to be generalisable from groups defined by specialty, postgraduate training body, training network/ area (site) and stage of training to the wider population within that group. The generalisability of scores in further sub-categories, defined by combinations of these variables, for example postgraduate training body and stage of training, are less robust. A bootstrapping exercise was conducted to determine that, where a minimum of thirty respondents is achieved in each group, the generalisability score is more acceptable (g index >0.8) and it is possible to generalise from these groups with confidence (see Appendix A). While it is possible to use D-RECT scores from smaller groups of respondents, a more cautious approach is required in generalising to the wider population within that group.

Factor Analysis

Our first analysis was to carry out a Multiple Group Factor Analysis (MGFA) on all data with complete data. This is a restricted or 'confirmatory' factor analysis in which the 50 items are constrained to conform to the a-priori 11 factor model. A least squares procedure was used (10). This is a non-iterative procedure for fitting observed data to an expected pattern. Fitting such models is often problematic because the coefficients most often used are either sample size dependent or heuristic estimates. We have elected to uses a very simple fitting index proposed by Fleming (11) which has

the advantage of allowing us to gauge the fit to the model for each item and each factor separately as well providing a general model fit parameter. Fleming's index is essentially a signal to noise ration and ranges between 0 and 1. It may be interpreted in much the same way as a reliability coefficient, a higher value indicating better fit. An additional advantage of this method is the fact that a very large number of competing models can be tested alongside the theoretical target model. We generated 10,000 randomly generated models and estimated the fit for each. The target fit is then compared to the randomly generated model. Experience suggests that the distribution of random fit indices is essentially normal although there is no statistical guarantee of this. As a rough guide a standardised deviation of the target fit from the mean random fit provides a method for judging whether the target model has been confirmed. In addition, it is expected that the target model should show a better fit than any of the random models before we can assert that the model is confirmed.

The factor analysis is presented in Table 3. The full factor pattern for the 11-factor model is presented. The overall fit of the model is 0.92. The mean fit for 10,000 models is 0.74 with a standard deviation of 0.0167. This renders a z score of 10.78, placing the target fit over 10 standard deviations above the mean fit, no random model exceeded the target value. Trainees in both countries understand the questionnaire similarly and respond in similar patterns. D-RECT is measuring a stable construct in both Ireland and the Netherlands. Therefore it is an instrument with good construct validity.

TABLE 3
Multiple Group 'Confirmatory' Factor Analysis of the 50 Items of the D-RECT

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	Simplicity
1	0.46	0.02	-0.03	-0.00	0.01	-0.05	0.09	-0.00	0.06	-0.03	0.00	0.91
2	0.57	-0.01	-0.02	-0.06	-0.00	0.00	-0.05	0.00	-0.00	0.00	0.03	0.97
3	0.52	-0.01	0.05	0.06	-0.01	0.04	-0.04	-0.00	-0.05	0.03	-0.03	0.93
4	0.05	0.43	-0.09	0.03	0.03	-0.01	-0.03	0.02	-0.07	-0.04	-0.00	0.88
5	-0.03	0.44	-0.04	0.01	-0.00	0.01	0.02	0.00	0.01	-0.00	-0.02	0.97
6	0.01	0.38	-0.05	0.00	0.04	0.02	0.00	0.02	0.10	-0.08	-0.01	0.85
7	0.00	0.40	0.04	-0.05	-0.02	0.01	0.00	-0.01	-0.03	0.01	0.05	0.93
8	-0.01	0.43	0.00	-0.01	0.00	-0.01	-0.04	0.02	-0.01	0.07	-0.01	0.95
9	0.02	0.39	0.01	0.01	-0.01	-0.00	0.01	-0.02	0.02	0.01	0.01	0.97
10	-0.03	0.35	0.11	-0.00	-0.05	-0.00	0.03	-0.02	-0.05	0.04	-0.03	0.83
11	-0.03	0.42	0.01	0.01	0.00	-0.01	-0.00	-0.00	0.02	-0.02	0.01	0.97
12	0.01	0.21	0.33	-0.02	-0.02	0.01	0.00	0.01	-0.03	0.02	0.01	0.68
13	-0.00	-0.11	0.76	0.01	0.01	0.00	-0.00	0.00	0.01	-0.00	0.00	0.97
14	-0.00	-0.10	0.77	0.01	0.01	-0.01	0.00	-0.01	0.02	-0.02	-0.01	0.97
15	0.03	0.01	-0.01	0.60	0.01	0.04	0.02	0.04	-0.00	-0.02	-0.03	0.98
16	-0.03	-0.02	0.03	0.73	-0.06	-0.04	0.03	-0.00	-0.01	0.01	-0.01	0.98
17	-0.01	-0.00	0.02	0.68	-0.04	0.00	-0.06	0.00	0.01	0.02	-0.02	0.98
18	0.01	0.01	-0.04	0.56	0.08	0.00	0.00	-0.04	0.00	-0.01	0.06	0.95
19	-0.02	0.00	-0.00	0.06	0.74	0.02	-0.01	-0.00	-0.00	-0.00	-0.00	0.99
20	-0.03	0.03	-0.00	0.01	0.75	-0.00	-0.00	0.00	-0.03	-0.04	0.02	0.99
21	0.05	-0.04	0.01	-0.08	0.63	-0.01	0.01	0.00	0.03	0.04	-0.01	0.96
22	0.03	-0.04	0.02	-0.02	-0.00	0.70	0.00	-0.02	-0.00	0.00	0.00	0.99
23	-0.03	0.06	-0.01	0.00	0.01	0.66	-0.01	-0.01	0.00	-0.00	0.02	0.98
24	0.00	-0.01	-0.00	0.02	-0.01	0.67	0.01	0.03	-0.00	0.00	-0.02	0.99
25	0.06	0.02	-0.05	-0.01	0.04	0.00	0.48	0.00	0.03	-0.07	0.00	0.93
26	0.00	0.03	-0.04	0.07	-0.05	0.00	0.53	-0.01	-0.03	-0.00	-0.01	0.95
27	-0.03	-0.02	0.00	-0.05	-0.00	-0.02	0.49	0.01	-0.01	0.08	-0.00	0.95
28	-0.02	-0.04	0.08	0.00	0.01	0.01	0.45	0.00	0.00	-0.00	0.01	0.94
29	0.02	-0.12	0.04	0.05	-0.01	0.02	0.06	0.40	-0.23	0.12	-0.02	0.63
30	-0.07	0.02	0.04	0.01	-0.01	-0.01	0.01	0.48	-0.28	0.10	-0.04	0.69
31	-0.08	0.07	-0.01	0.02	0.00	-0.03	-0.04	0.46	-0.14	0.04	-0.02	0.83
32	0.01	0.03	0.00	0.01	0.04	0.00	-0.00	0.35	0.02	-0.08	-0.01	0.92
33	0.07	0.05	-0.05	-0.05	0.02	-0.04	-0.03	0.30	0.14	-0.04	0.01	0.69
34	0.04	0.02	-0.03	-0.04	0.03	-0.08	0.00	0.33	0.16	-0.07	0.03	0.69
35	-0.03	-0.08	0.04	-0.02	-0.04	0.19	-0.00	0.25	0.11	0.01	-0.01	0.50
36	0.03	-0.01	-0.05	0.01	-0.02	-0.04	0.00	0.28	0.22	-0.08	0.06	0.54
37	0.03	-0.01	-0.04	0.01	0.05	-0.02	0.01	0.11	0.41	-0.09	0.04	0.85
38	-0.01	0.04	-0.04	-0.01	-0.01	0.05	-0.05	0.03	0.46	-0.01	0.04	0.93
39	0.00	-0.02	0.03	0.00	0.01	-0.02	0.00	-0.08	0.63	0.04	-0.03	0.96
40	-0.02	-0.00	0.05	-0.01	-0.05	-0.00	0.03	-0.06	0.57	0.06	-0.05	0.94
41	-0.02	-0.03	-0.02	-0.00	0.02	-0.01	0.06	0.01	0.04	0.46	-0.12	0.89
42	-0.02	-0.01	-0.04	0.00	-0.00	0.01	0.01	0.00	0.00	0.54	-0.13	0.93
43	0.03	-0.03	-0.00	0.01	-0.01	0.01	0.05	-0.01	-0.03	0.50	-0.05	0.96
44	-0.03	0.07	-0.06	-0.01	0.03	-0.00	-0.03	0.01	0.08	0.40	-0.01	0.88
45	0.03	0.01	0.08	0.01	-0.03	-0.04	-0.03	-0.00	-0.06	0.32	0.18	0.66
46	0.01	-0.00	0.06	-0.01	-0.00	0.02	-0.05	-0.02	-0.03	0.36	0.13	0.82
47	0.03	-0.05	-0.01	-0.00	-0.00	-0.00	-0.05	0.01	-0.04	0.23	0.35	0.66
48	-0.00	0.01	0.04	0.01	-0.03	-0.00	0.03	-0.00	-0.03	-0.01	0.48	0.97
49	0.03	0.02	-0.05	0.00	0.02	-0.03	0.01	-0.01	0.05	-0.15	0.54	0.90
50	-0.06	0.00	0.02	-0.01	0.01	0.04	0.00	0.00	0.02	-0.06	0.48	0.95
Factor Fit	0.93	0.90	0.93	0.97	0.97	0.95	0.94	0.96	0.77	0.86	0.87	0.92

Factor Correlation Matrix

I	1.00										
II	0.74	1.00									
III	0.53	0.68	1.00								
IV	0.40	0.45	0.32	1.00							
V	0.26	0.22	0.15	0.37	1.00						
VI	0.43	0.46	0.36	0.36	0.31	1.00					
VII	0.60	0.63	0.49	0.51	0.37	0.45	1.00				
VIII	0.62	0.68	0.48	0.53	0.32	0.56	0.68	1.00			
IX	0.57	0.61	0.43	0.46	0.28	0.48	0.54	0.74	1.00		
X	0.56	0.62	0.50	0.46	0.35	0.47	0.68	0.67	0.56	1.00	
XI	0.58	0.68	0.56	0.39	0.29	0.43	0.60	0.63	0.54	0.75	1.00

Z=10.78

Summary

In summary, the content and constructs of D-RECT have good validity. The total scale and the majority of subscales are reliable. These findings mean that the total D-RECT score measures clinical learning environment in the Irish context, with a high level of precision. Most of subscale scores also measure their targets, the distinct dimensions of the clinical learning environment, with acceptable precision. Some subscales do so with less precision than we would like, and caution is required in their interpretation. The small number of items in some of the subscales is an important contributor to this lack of precision. Generalising from respondent groups of at least 30 trainees, to the wider population of trainees, is valid. More detailed analysis, at subscale level and with smaller groups of respondents should be interpreted more cautiously. Overall, this makes D-RECT a useful way to examine the clinical learning environment in Ireland, especially when using total score from larger numbers of respondents. Care should be taken when interpreting results based on small groups of trainees, especially at subscale level.

References:

1. Boor K, Van Der Vleuten C, Teunissen P, Scherpbier A, Scheele F. Development and analysis of D-RECT, an instrument measuring residents' learning climate. Med Teach [Internet]. 2011 Jan [cited 2013 Jun 28];33(10):820–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21355691

2. Boor K, Scheele F, Van Der Vleuten CPM, Teunissen PW, Den Breejen EME, Scherpbier AJJA. How undergraduate clinical learning climates differ: A multi-method case study. Med Educ. 2008;42:1029–36.

3. Boor K, Teunissen PW, Scherpbier AJJA, van der Vleuten CPM, van de Lande J, Scheele F. Residents' perceptions of the ideal clinical teacher-A qualitative study. Eur J Obstet Gynecol Reprod Biol. 2008;140:152–7.

4. Bennett D, Dornan T, Bergin C, Horgan M. Quality of Postgraduate Training in Ireland: Expectations and Experience. Ir J Med Sci [Internet]. 2014; Available from: http://link.springer.com/article/10.1007%2Fs11845-013-1060-5#page-1

5. Kelly M, Bennett D, O'Flynn S. General practice - the DREEM attachment? Comparing the educational environment of hospital and general practice placements. Educ Prim Care. 2012;23(1).

6. Bennett D, Kelly M, O'Flynn S. Are the bigger hospitals better: DREEM on? Ir J Med Sci. Springer London; 2010;179(4):515–9.

7. Hammond SM, O'Rourke M, Kelly M, Bennett D, O'Flynn S. A psychometric appraisal of the DREEM. BMC Medical Education. 2012. p. 2.

8. Rourke M, Hammond S, Flynn S, Boylan G. The medical student stress profile: A tool for stress audit in medical training. Med Educ. 2010;44:1027–37.

9. Wetzel AP. Factor analysis methods and validity evidence: a review of instrument development across the medical education continuum. Acad Med [Internet]. 2012 Aug [cited 2014 Nov 5];87(8):1060–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22722361

10. Harman H. Modern Factor Analysis. 3rd ed. Chicago: University of Chicago Press; 1976.

11. Fleming JS. An Index of Fit for Factor Scales. Educational and Psychological Measurement. 1985. p. 725–8.

Appendix A: G study results

Generalisability testing was conducted across the various groups within the study population. The G coefficients are generally supportive. A standard analysis of variance was also conducted, to demonstrate the differences between groups for each grouping variable. Sometimes there will be a statistically significant difference between groups and yet the G coefficient is still quite high. This indicates that the qualitative meaning of the measure is equivalent across groups although the quantity of satisfaction with training may vary. The following tables describe these results.

Descriptive Statistics for Site

Site	Mean	Std. Deviation	N
Dublin Midlands	3.3084	.70189	138
Dublin North East	3.3526	.65866	265
GP Practice	3.7361	.53815	204
Mental Health Services	3.8227	.68003	107
Midwest	3.4826	.60193	73
Paediatrics	3.3678	.55229	79
South/South West	3.3450	.62647	238
West/North West	3.3153	.64579	194
Dublin East	3.3360	.61889	247
Other	3.8252	.58298	27
Total	3.4370	.65141	1572

Oneway ANOVA

$F = 20.802$ $p<0.001$ $\eta^2 = 0.096$

The F ratio is highly significant indicating group differences. You can see that the Mental Health Services, Other and GP Practice groups had the highest satisfaction with training. The lowest were Dublin Midlands and West/Northwest. However the effect size of 0.096 is quite small and as the mean scores reflect the 1-5 ordinal scale you can see that they are all above 3. So everyone is happy just some are happier than others.

G-Analysis (Items by Site)

	MS	DF	Variance Component
Site	110.61	9	1.83
Item	95.18	49	9.44
interaction	2.09	441	0.14
Residual(m)	19.28		
Residual(i)	0.73		

Absolute G = 0.76

This G-Analysis shows the computational steps in estimating G. Here the G is 0.76. This is pretty good although generalisability perfectionists tend to suggest 0.80 as the target to aim for.

Descriptive Statistics for Stage of Training

Stage of Training	Mean	Std. Deviation	N
Basic Specialist Training Programme	3.3697	.64794	406
G.P. Training Programme	3.5044	.63872	357
Higher Specialist Training Programme	3.6256	.59403	430
Intern Training Programme	3.0281	.60912	238
Registrar Training Programme	3.6391	.60843	84
Run-through (basic and higher) Specialist TP	3.4573	.58801	54
Other	3.9165	.39139	3
Total	3.4370	.65141	1572

Oneway ANOVA

F = 27.073 p<0.001 η^2 = 0.094

G-Analysis (Items by Stage of Training)

	MS	DF	Variance Component

Stage of Training	267.79	6		4.98
Item	7.07	49		0.91
interaction	3.33	294		0.37
Residual(m)	18.59			
Residual(i)	0.72			

Absolute G = 0.96

Descriptive Statistics for GP Training

GP Training	Mean	Std. Deviation	N
Not in GP Training	3.3957	.65317	1360
In GP Training	3.7018	.57458	212
Total	3.4370	.65141	1572

Oneway ANOVA

F = 41.528 p<0.001 η^2 = 0.026

G-Analysis (Items by GP Training)

	MS	DF	Variance Component
Main	92.38	1	1.45
Item	41.82	49	20.54
Interaction	5.69	49	2.48
Residual(m)	20.01		
Residual(i)	0.73		

Absolute G = 0.63

Descriptive Statistics for Speciality

Speciality (Block)	Mean	Std. Deviation	N
Anaesthesia	3.5317	.54634	121
Emergency Medicine	3.2480	.69785	65
General Practice	3.7018	.57458	212
Surgery	3.2491	.67580	250
Medicine	3.3103	.63551	455
Obstetrics & Gynaecology	3.2374	.62153	81
Occupational Medicine	4.0127	.30001	4
Ophthalmology	3.4326	.41604	15
Paediatrics	3.3853	.58278	115
Pathology	3.4578	.71001	54
Psychiatry	3.8615	.59212	124
Public Health Medicine	4.0055	.48327	4
Radiology	3.7516	.57386	50
Other	3.3225	.52274	22
Total	3.4370	.65141	1572

One Way ANOVA

F = 13.650 p<0.001 ω^2 = 0.102

G-Analysis (Item by Speciality)

MS DF Variance Component

Main	113.67	12	1.89
Item	29.50	49	2.27
interaction	2.75	588	0.21
Residual(m)	18.95		
Residual(i)	0.71		

Absolute G = 0.82

Descriptive Statistics for Training Body

OVERSEE	Mean	Std. Deviation	N
Intern Training Network	3.0863	.56841	165
Royal College of Physicians in Ireland	3.4226	.63077	332
Royal College of Surgeons in Ireland	3.5090	.62726	145
The College of Anaesthetists of Ireland	3.5970	.53558	87
The College of Psychiatry of Ireland	3.8822	.64920	54
The Faculty of Radiologists	3.8457	.53351	7
The Irish College of General Practitioners	3.3884	.59846	170
The Irish College of Ophthalmologists	3.4089	.63743	9
Other	3.4457	.73916	7
Total	3.4167	.63434	976

One Way ANOVA

F = 11.921 p<0.001 η^2 = 0.090

G-Analysis (Item by Training Body)

	MS	DF	Variance Component
Training Body	220.12	8	4.03

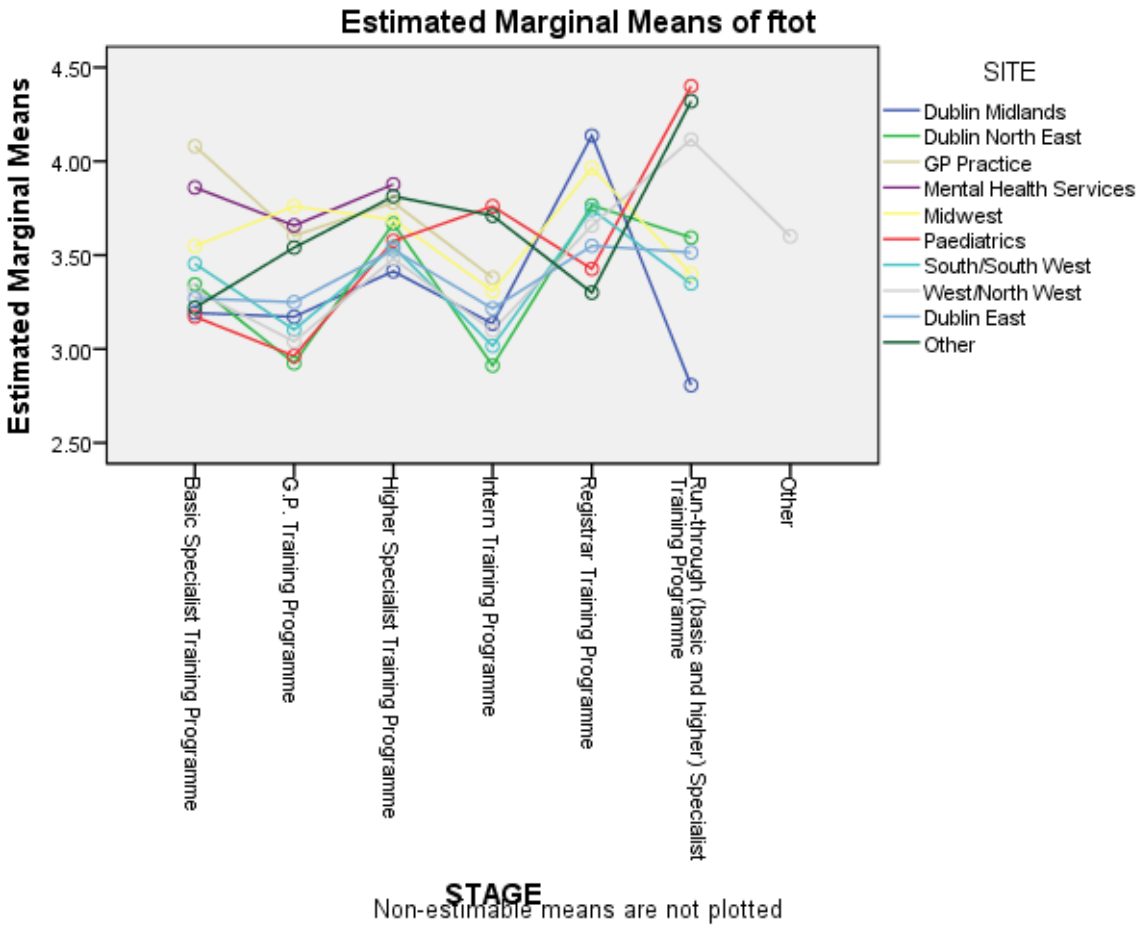
Item	31.41	49	3.41
Interaction	3.54	392	0.31
Residual(m)	18.465		
Residual(i)	0.715		

Absolute G = 0.90

Analysis of Site by Stage

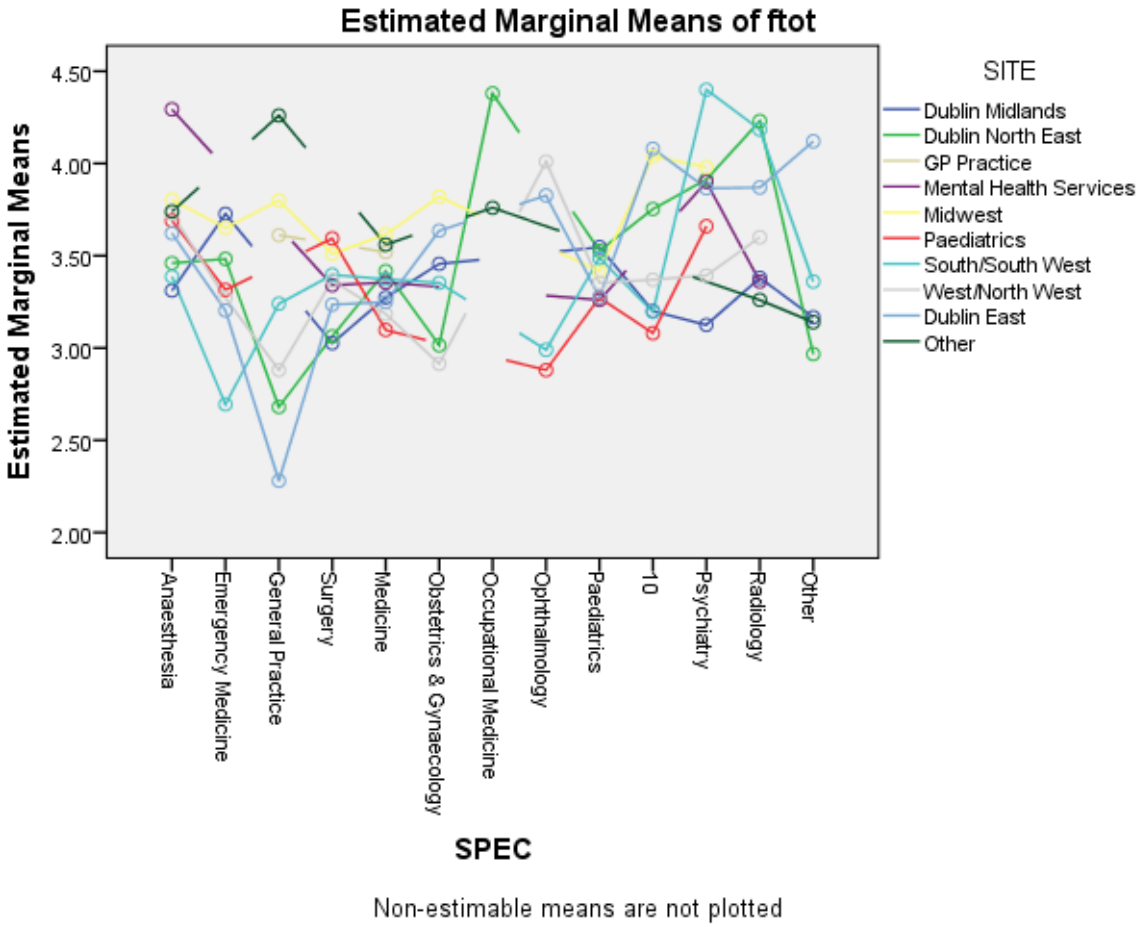
Site	1.56	9	4.10**	0.30	0.25
Stage	1.52	6	4.02**	0.08	0.16
Site by Stage	0.44.	40	1.16	0.05	0.06
Residual	0.38				

G=0.54

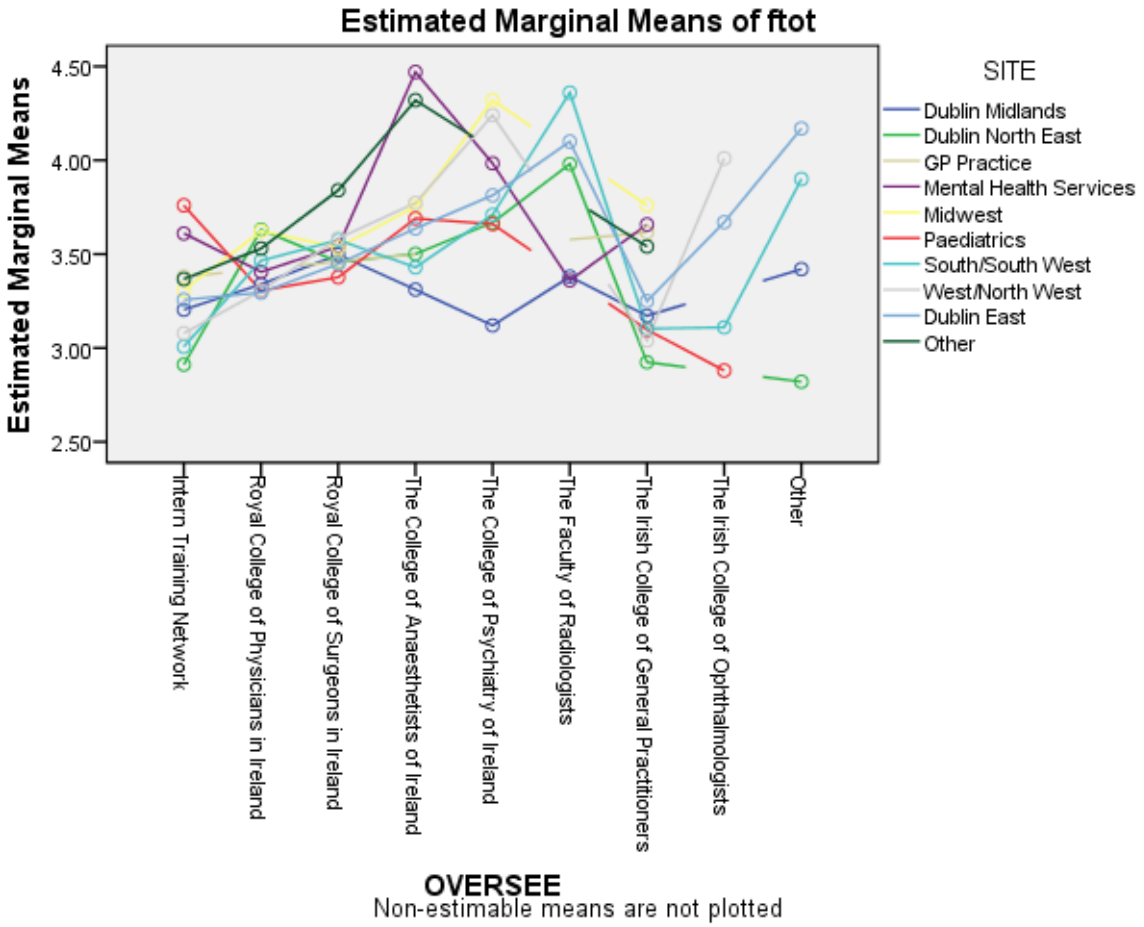

Estimated Marginal Means of ftot
Estimated Marginal Means
4.50
4.00
3.50
3.00
2.50
SITE
Dublin Midlands
Dublin North East
GP Practice
Mental Health Services
Midwest
Paediatrics
South/South West
West/North West
Dublin East
Other
Basic Specialist Training Programme
G.P. Training Programme
Higher Specialist Training Programme
Intern Training Programme
Registrar Training Programme
Run-through (basic and higher) Specialist Training Programme
Other
STAGE
Non-estimable means are not plotted

Site	0.42	9	1.06	0.03
Spec	0.79	12	2.01	0.07
Site x Spec	0.47	66	1.19	0.08
Residual	0.39			

G= 0.89



Oversee	1.21	8	3.21	0.09		0.113	
Site	0.75	9	1.99	0.05		0.074	g = 0.81
Oversee by Site 0.46	51	1.22	0.07		0.080		
Residual	0.38						



Bootstrapping Method: Numbers needed for a reliable response

There is a general view that the sample in each group should be as large as possible to make for reliable results. The question of what the minimum acceptable number has been addressed using a bootstrapping method, in which the data is sampled with replacement to generate varying subsample sizes. Since this is dependent on the individuals selected a program was written to randomly sample from the existing data generating 100 different results for each sample size. The average G-coefficient was then taken as representative of that particular sample size. Sample sizes varied between 3 and 100.The smaller the sample the smaller the G as expected. As the sample size increases so does the generalisability. These analyses were carried out for all the independent variables and the consistent finding was that around 30 cases in each group was required to rely upon a good generalisability (set at 0.8).

The bootstrap output is shown here:

3	0.477143579274882
4	0.451393907097954
5	0.449443635305183
6	0.48009245089408
7	0.501850252025502
8	0.509705920331037
9	0.508059212501517
10	0.538873300674487
11	0.566428343502949
12	0.595405156392312
13	0.600246426029656
14	0.594203417126117
15	0.62063788179498
16	0.670484998887938
17	0.673872203541281
18	0.704409840477895
19	0.713775837457267
20	0.704327905908406
21	0.727671921718245
22	0.736813497148803
23	0.74562025282408
24	0.76392590397937
25	0.768176934422792
26	0.755093963832333
27	0.777232576056231
28	0.780679623380783
29	0.792821408125703
30	0.786554470927521
31	0.801743724377137
32	0.795811225503417
33	0.813841147241551
34	0.811737990511755
35	0.820859666998235
36	0.830579878796614
37	0.838711432506402
38	0.838154512695195
39	0.825664864267199
40	0.848188191874693
41	0.837597594024188
42	0.849229595564261
43	0.849764348360232
44	0.860150432122486
45	0.860586317019843
46	0.859850062629013
47	0.857263101342257
48	0.869618689595987
49	0.869776380782416
50	0.873584351107635
51	0.875859014477095
52	0.879475076011732
53	0.877053554190818
54	0.878505385137734
55	0.885452850543588
56	0.89088036584469
57	0.886192624978841
58	0.884034001165725
59	0.895907916830121
60	0.890825254025574
61	0.896952273104257
62	0.894200933298018
63	0.901988800663089
64	0.904539968582975
65	0.902558730798732
66	0.903802678553113
67	0.905406362483408
68	0.906402112530308
69	0.90718459327102
70	0.90876839185738
71	0.908519951260131
72	0.908990623338058
73	0.90795399674005
74	0.913590679666964
75	0.914428380493964
76	0.914950559626167
77	0.917446177374705
78	0.917293916168616
79	0.917832234406549
80	0.919849630283958
81	0.921695859659142
82	0.920647933746945
83	0.922295119750958
84	0.92089535068693
85	0.924945470700249
86	0.9252970616929
87	0.927918304898143
88	0.925583646666029
89	0.92788119770048
90	0.927156095285643
91	0.928981701600359
92	0.926185625590314
93	0.931774980511517
94	0.92934083184441
95	0.932376769054623
96	0.933332041375124
97	0.934915552749925
98	0.93258820023887
99	0.93090672929336
100	0.936009855767013
3	0.456509545140115
4	0.466104621191062
5	0.447785684106585

6	0.499410496895318
7	0.522789079789981
8	0.53611042940431
9	0.547851096520742
10	0.579522571089779
11	0.591174590532119
12	0.636381174709654
13	0.647031239938375
14	0.646384370265218
15	0.663462777667734
16	0.68365782229689
17	0.686344947259199
18	0.719753335443724
19	0.729286143911743
20	0.748410172069782
21	0.760193555315095
22	0.752161611145549
23	0.777471792836581
24	0.784263174497664
25	0.793605141455819
26	0.805326517822043
27	0.796313230400335
28	0.817034720885904
29	0.806696708788677
30	0.818477629218421
31	0.834408629892001
32	0.835965584129811
33	0.835470124194342
34	0.845362242559355
35	0.845303439857218
36	0.850921331206763
37	0.851901049604684
38	0.850873270462334
39	0.859007770314312
40	0.858952161542044
41	0.865705811385845
42	0.863495768041341
43	0.870390571145213
44	0.87003582135997
45	0.873846176144972
46	0.882892585400974
47	0.876936099544784
48	0.880745271228065
49	0.886277986572866
50	0.883403605315843
51	0.885351022942324
52	0.892158378512648
53	0.891427613319576
54	0.891646334283256
55	0.898985283434277
56	0.896747158994303
57	0.902337303459437
58	0.901593749198362
59	0.902030114009783
60	0.904378663960302
61	0.909939627745615
62	0.907532804876944
63	0.905188161632251
64	0.909500494801744
65	0.912541893036011
66	0.909850420756846
67	0.915090436283529
68	0.915366678133373
69	0.9176372560444
70	0.915774427268251
71	0.916417834441692
72	0.919430572046225
73	0.919816598041226
74	0.919466944257013
75	0.920578235462884
76	0.923223452558027
77	0.926489912514572
78	0.928106575858018
79	0.925640788145503
80	0.925035780077073
81	0.92846961726761
82	0.930108858455515
83	0.932074358016175
84	0.93199439005057
85	0.931788039403913
86	0.931811320426303
87	0.931873018270729
88	0.932110192195828
89	0.933152240774882
90	0.934034897221624
91	0.934383024391647
92	0.935427587504023
93	0.934820575698661
94	0.937090876884862
95	0.939613421690337
96	0.936204292580419
97	0.941422127193469
98	0.940158406315466
99	0.940756532679716
100	0.942098069219823
3	0.392458416801336
4	0.365625260951923
5	0.371295509291002
6	0.383652244724763
7	0.387076974276521
8	0.398344033116536
9	0.442865683132237

10	0.452352588646643
11	0.465657052431922
12	0.462394945083739
13	0.489176819942276
14	0.517905394555099
15	0.544992193263649
16	0.516377773611652
17	0.545005360236093
18	0.53689759763378
19	0.579536828697026
20	0.587272778380801
21	0.577554893870379
22	0.592111447116348
23	0.624184972197171
24	0.586988171387712
25	0.609999593080651
26	0.6607416093436
27	0.667717302729391
28	0.648777202873545
29	0.68383300954244
30	0.690298669266088
31	0.67877636845694
32	0.695294098948797
33	0.709755236512735
34	0.701084961685728
35	0.70599135241318
36	0.732521190840758
37	0.71247827595138
38	0.730359272093377
39	0.747591631211757
40	0.718136692967474
41	0.751315716088115
42	0.746069175427065
43	0.75725934290038
44	0.770954190687444
45	0.758102794122314
46	0.776411565523116
47	0.767012809799948
48	0.769107468440606
49	0.792761534298153
50	0.773124241290498
51	0.787588144406482
52	0.798118843320067
53	0.783894279218446
54	0.801956490390007
55	0.81366288324979
56	0.800460569250459
57	0.802785568551843
58	0.810586433226983
59	0.817368211963056
60	0.813588222021965
61	0.826544112335728
62	0.826487257690683
63	0.827762947212688
64	0.831043137114912
65	0.832565751942452
66	0.827636443298668
67	0.84062237075806
68	0.838003644593138
69	0.838737842122649
70	0.826619314107185
71	0.837926628303755
72	0.844248362856745
73	0.849572131343311
74	0.856544373025711
75	0.859337632369663
76	0.850538444999252
77	0.85695015354798
78	0.852155996566496
79	0.859212616850944
80	0.857509116206436
81	0.863384143739708
82	0.863803055180411
83	0.862870759056437
84	0.871396092640249
85	0.865001996957187
86	0.866471302674295
87	0.871033342957397
88	0.871563311412243
89	0.873173594317713
90	0.869444370917112
91	0.874517036386136
92	0.875419749312841
93	0.875752890270421
94	0.880865394870306
95	0.885572638036323
96	0.880000136851938
97	0.879089354335669
98	0.883503202780417
99	0.885498704259648
100	0.883872795525719
3	0.898759504222007
4	0.927230698564652
5	0.939575288976049
6	0.947873179116216
7	0.955973808892173
8	0.961633645456115
9	0.965965328405246
10	0.968911877792829
11	0.971410087222537
12	0.973902205481255
13	0.97542643717133

14	0.977019163399238
15	0.978495787957828
16	0.980424250417689
17	0.981281196854616
18	0.982331361805182
19	0.982912691755299
20	0.984341553890015
21	0.984980837064769
22	0.985253667646882
23	0.9859105421965
24	0.986784318691806
25	0.987380000026571
26	0.987838753243007
27	0.988358736295248
28	0.98864645703052
29	0.989076470880988
30	0.989505729391915
31	0.989809482321453
32	0.990108935024512
33	0.99035654368749
34	0.990611464583658
35	0.99097678727151
36	0.991206819272485
37	0.991471644507469
38	0.991498356689434
39	0.991804855552452
40	0.991944265727747
41	0.992239473639405
42	0.992315103856156
43	0.992575770091666
44	0.992788767452526
45	0.992911833389537
46	0.993037806729837
47	0.993167710542771
48	0.993318104254199
49	0.993445624576227
50	0.99359028318875
51	0.993812777244662
52	0.993830371517748
53	0.993986857282797
54	0.994098047661636
55	0.994205892580843
56	0.994223928851249
57	0.994420833934141
58	0.994479196150836
59	0.99458864348691
60	0.994652464926655
61	0.994738323319103
62	0.994802457907962
63	0.994913968789255
64	0.995019873206785
65	0.995070580654629
66	0.995168957795898
67	0.995210243235158
68	0.995306202684298
69	0.995340916614943
70	0.995411883737412
71	0.995460055365781
72	0.99556382553761
73	0.995597409030119
74	0.995696649177874
75	0.995743885769101
76	0.995795934726667
77	0.995841838873573
78	0.995913096904716
79	0.995925901122247
80	0.99594742712557
81	0.996029398366813
82	0.996087233928813
83	0.996120824640679
84	0.9961942981712
85	0.996209286950674
86	0.996283016292002
87	0.996300294901711
88	0.99635876658433
89	0.996401817522471
90	0.996430923321965
91	0.996501825862168
92	0.996528545637126
93	0.996538009577117
94	0.996565950694119
95	0.99664183088089
96	0.996668947024571
97	0.996696591414315
98	0.996700659679501
99	0.996753668722683
100	0.996793524450135
3	0.55061816076829
4	0.53992002684202
5	0.506328971753335
6	0.537961598316799
7	0.531795174336167
8	0.562526868651046
9	0.545631868673336
10	0.568923615879521
11	0.505921631447555
12	0.549179971064513
13	0.555665799618746
14	0.55953553171146
15	0.546571995825194
16	0.562277398382158
17	0.532455886512802

18	0.511818952669332
19	0.553955026332997
20	0.573845567224341
21	0.575513346011046
22	0.543817474758287
23	0.57257184728616
24	0.588415917450775
25	0.572301807104017
26	0.561746781444551
27	0.566059945933252
28	0.595325367175089
29	0.606876630885311
30	0.598548730760006
31	0.571518215132941
32	0.551444732380464
33	0.560479587022662
34	0.594030386634995
35	0.594418322530465
36	0.609858683040467
37	0.630563559868414
38	0.602560622125331
39	0.617218358724781
40	0.595434765756391
41	0.635161103738285
42	0.619124607101097
43	0.639249744248225
44	0.621158822745887
45	0.627146094045901
46	0.679835564338431
47	0.633261949920558
48	0.604862163275264
49	0.647161488617645
50	0.605749442847207
51	0.629431510103513
52	0.636035033429764
53	0.6287121147766
54	0.653598594175976
55	0.621510350218027
56	0.659777670299409
57	0.634181611112851
58	0.680131574646545
59	0.667086253625654
60	0.615864088080782
61	0.682944989972342
62	0.66733527043783
63	0.677861823707444
64	0.704176327907704
65	0.664314676377711
66	0.647044346917039
67	0.64870515632751
68	0.681040009321103
69	0.657065297028321
70	0.647083231170976
71	0.67588514597925
72	0.677100980383013
73	0.716642366302035
74	0.681737908788695
75	0.694037351433423
76	0.696945897987467
77	0.711662458331176
78	0.736272613723521
79	0.724443387169598
80	0.696825443799516
81	0.711590044219939
82	0.708029858469222
83	0.669269906928886
84	0.729872368283263
85	0.692776564236981
86	0.690344999300612
87	0.689667162793201
88	0.682119224770041
89	0.665652563863328
90	0.696491110217271
91	0.709983446429669
92	0.698606088968194
93	0.721603193196391
94	0.745280754359178
95	0.694255075660239
96	0.697566054309552
97	0.739607303493875
98	0.707777655406998
99	0.740023297848621
100	0.712126530834279