On the fractal dimension of a social measurement: I

David Andrich
Murdoch University, Western Australia

Mailing address
David Andrich
Murdoch University
Murdoch 6150
Western Australia

Acknowledgements

The research reported in this paper was supported in part by an Australian Research Council Linkage grant with the Australian National Ministerial Council on Employment, Education, Training and Youth Affairs (MCEETYA) Performance Measurement and Reporting Task Force; UNESCO's International Institute for Educational Planning (IIEP), and the Australian Council for Educational Research (ACER) as Industry Partners. Permission to use the data from the Australian Scholastic Aptitude Test (ASAT), which is a test constructed by ACER, was given by the Curriculum Council of Western Australia. This paper has benefited from the many discussions with Irene Styles on the nature of social measurement and the impacts of level of scale. The work was also helped by a much earlier paper with Barry Sheridan reporting the drop in reliability for the same ASAT data set depending on the method of its calculation and continuing discussion with Guanzhong Luo regarding the structure of the Rasch model. The programming of Barry Sheridan and Guanzhong Luo was essential to being able to estimate the relevant parameters in the analysis using the Rasch model.

*Report No. 3 ARC Linkage Grant LP0454080: Maintaining Invariant Scales in State, National and International Level Assessments. D Andrich and G Luo Chief Investigators, Murdoch University

On the fractal dimension of a social measurement: I

Abstract

Many scales in psychology, education and social measurement in general, constructed to measure a single variable, are nevertheless designed to measure different aspects of the variable. The scales are often composed of subscales which measure these various aspects. These different aspects are conceptually related to the variable and are employed because they make the scale more valid than if only one of the aspects were measured. The aspects are expected to be reasonably highly correlated, but most unlikely to be perfectly correlated. To the degree that the aspects are not perfectly correlated, to that degree the scale could be considered not to measure a unidimensional variable. This paper suggests that the concepts of fractal geometry may be useful in characterizing this non-unidimensionality. Fractal geometry of physical entities is concerned with characterizing *roughness*, and the approach taken in this paper is that the presence of dimensions other than the dominant dimension adds roughness to the measurement. However, because of some special features of social measurement when compared to physical measurement, the term *thickness* of a measurement is also introduced to complement the concept of roughness in social measurement. The distinctive feature of the approach, compared with those that attempt to characterize the different dimensions explicitly, is that the focus remains on the main variable to be measured. The calculation and interpretation of the fractal dimension is made possible by the application of the unidimensional Rasch model. The paper illustrates the calculation of the fractal dimension of a social measurement by analyzing an Australian Scholastic Aptitude Test and considers some implications of such a characterization for social measurement.

Key words: psychological measurement, social measurement, dimensionality, fractal dimension

On the fractal dimension of a social measurement: I

1. Introduction

Scales constructed in education, psychology, and social measurement in general, are intended to measure more or less of some construct. The term *variable* will be used generally in the paper for the term *construct* to simultaneously capture the intention to measure the construct, and the term *construct* is reserved for its occasional use to emphasize the qualitative distinctiveness of the variable to be measured. The term *scale* is used in two senses where the context makes clear the sense. In one sense it refers to a set of items used as an operationalization of the variable. It follows that the term subscale refers to a subset of items of such an operationalization. In the second sense, it implies a metric, and it is often used in conjunction with the term *unit*, giving the expression *unit of scale.*

Such scales are also generally summarized as measuring a *unidimensional* variable. They have been analyzed and modified using both traditional and modern test theory. In both theories there are methods for identifying various violations of unidimensionality with prescriptions as to how these might be handled depending on the circumstances. In general, the perspective is that multidimensionality somehow contaminates the scale. However, many scales are developed from an analysis of *aspects* or *components* of a construct, sometimes at more than one level. In order to capture these broader aspects of a construct, the aspects are understood to be imperfectly, rather than perfectly, related. When subscales which measure these aspects are included in a scale, the variable measured is more complex than if only one subscale were used.

Measurements of the complex variable using subscales can be seen to increase its validity beyond that which would be achieved if only one aspect were measured using only one subscale. Although seemingly not picked up in the literature on psychological, educational, and social measurement, Cronbach (1951) considered a similar issue as a trade-off between *fidelity*, the precision of measurement of a single variable, and

bandwidth, the incorporation of variables other than the main variable in the measurement.

The illustrative example used in this paper is a version of the Australian Scholastic Aptitude Test (ASAT) which is composed of 100 dichotomously scored multiple choice items (Australian Council for Educational Research, 1989). The test is of general reasoning in the generic areas of learning at secondary schools and the version analyzed in this paper was used in part for selecting students for studies at tertiary institutions and in part for equating tests in the same general areas of learning. To meet this general purpose, the items can be classified into four successive and hierarchical levels: (i) as test of educational reasoning in two general areas of learning, Humanities (H) and Social Science (SS) in one and Mathematics (M) and Science (Sc) in the other; (ii) as a test of educational reasoning in each of the four disciplines of Humanities, Social Science, Science and Mathematics; (iii) as a test of reasoning in 17 different specific reading and problem stimuli, with respectively 4, 5, 3 and 5 stimuli in each of the four discipline areas, and (iv) as a test of reasoning on each of 100 dichotomously scored multiple choice items ranging from 3 to 8 in each reading.

To elaborate on levels (iii) and (iv), each of the specific stimuli has more than one multiple choice item arising from it, and in this paper will be referred to as a *reading stem* for the items. Figure 1 shows the component structure of the ASAT. Such structures are common in education where components are focused upon in teaching and learning and then integrated into higher-order levels of achievement (Andrich, 2002). The same structure is summarized in Briggs and Wilson (2003).

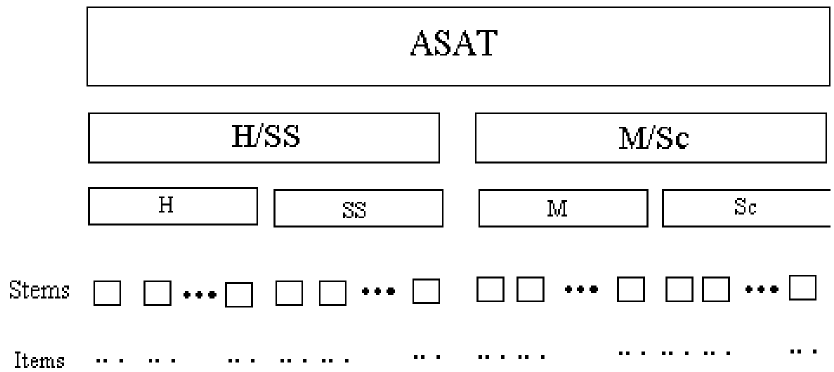

Figure 1. Component structure of the ASAT

Despite this relatively broad coverage of areas of learning in the ASAT, it is considered relevant to summarize a person's score on the test by a single number, and generally by the total number of items answered correctly. Where a scale is constructed by a composition of subscales, as the case of ASAT, it is inevitable that there is some relative multidimensionality among the items. However, eliminating the multidimensionality would contribute to reducing the complexity and therefore the validity of the ASAT for certain purposes for which it is used. Although ASAT has been used to make the point concrete, the same principles apply to other measurements of constructs carried out in terms of subscales. In these circumstances some multidimensionality at some levels of analysis may be seen as a positive property of the scale rather than evidence of some fault. *Therefore, it might be useful to be able to summarize efficiently the magnitude of this multidimensionality as a feature of the main single scale or dimension rather than as features of different dimensions*.

The purpose of this paper is to suggest that the two key concepts of fractal geometry, those of *roughness* and *self-similarity* at different units of scale, may be useful in summarizing this multidimensionality in social measurement, both numerically and conceptually. The idea is that measurement of social constructs composed of different aspects can be conceptualized as more or less *rough*, and that this roughness can be

characterized as a fractal dimension from measurements at different units of scale. However, in physical measurement, the units of measurement are defined independently of any object of measurement, and roughness characterizes only the object of measurement. In social measurement, as introduced above, the instrument of measurement itself is composed of more than one subscale, and the object of measurement is measured with such an instrument and therefore the fractal dimension is in part a property of the instrument and in part a property of the object of measurement.. The term *thickness* is coined to summarize the number of subscales in a measurement, where it is envisaged that the different subscales are interwoven, more or less as strands of a string are interwoven at successive levels to form a rope which is much thicker than each subscale. For example, in the case of ASAT, each individual item on its own characterizes a relatively *thin* variable. The items within a reading stem of say the SS subscale characterize a somewhat thicker variable, but a variable thinner than a combination of all items of the reading stems within the SS subscale. In turn, the SS subscale characterizes a thinner variable than the combination of items in reading stems of SS subscale and the H subscale. Finally, the combination of all items of all reading stems in all of the subscales of M, Sc, H and SS characterizes the thickest variable of the data set.

These metaphors, thick and thin and rough and smooth, are used in the paper and then further expanded on in the *Interpretation* section. In a companion paper (Andrich, 2006), an index of multidimensionality is derived and using it, a possible motif of mutually orthogonal dimensions is suggested to characterize the idea of roughness. The approach draws on the study of roughness of surfaces in physical measurement in terms of mutually orthogonal planes (Mandelbrot, 2002, p. 16). It is stressed that these papers are only a beginning of demonstrating the possible applications of such concepts in social measurement and that there is much further work that would need to be done to take advantage of these concepts. Some suggestions for further work, together with challenges, are covered in the *Interpretations* section at the end of the paper.

The paper begins by reviewing briefly the classic introductions to fractal geometry and by calculating the fractal dimensions of a hypothetical coastline and of a simple geometric motif. In the process it defines a fractal dimension and self-similarity and derives relevant formulae. Then by using analogous logic, an analysis of the responses to the 1989 ASAT test using the unidimensional Rasch model at different units of scale is shown. From this analysis, the fractal dimension of the measurement obtained of the ASAT test is calculated.

The paper considers further the particular feature of the analogy which makes the calculation of a fractal dimension in social measurement more challenging than its calculation in physical measurement. It is more challenging because of the availability of a well defined arbitrary unit of measurement in the former but not in the latter. The distinguishing features of the Rasch model which appears to overcome this lack of a well defined independent unit are summarized and exploited.

Finally, some broader implications of taking a fractal geometry perspective on measurement scales composed of imperfectly correlated subscales are broached and previous attempts to deal with the same issue are acknowledged.

2. The idea of roughness and fractal dimensions in physical measurement

There are now many books of varying intellectual challenge and varying interest on fractal geometry and the related concept of chaos. However, fractal geometry, and the measurement of roughness it provides, is very much a development of the last quarter or so of the 20th century. Mandelbrot (1983), who inspired the work on fractal geometry (Lauwerier, 1987), coined the term *fractal,* deriving it from the Latin *fractus* meaning *irregular* and combined it with *dimension,* to give the idea of a *fractal dimension,* a dimension that may have a value that is not an integer (Mandelbrot, 1983, p.4). Mandelbrot states that "Fractals are a suitable language for the study of roughness wherever it is encountered." (Mandelbrot, 2002, p16). Social measurement seems to be eminently suitable for such purposes. In particular, the use of the term *fractal* in

considering a dimension that is irregular seems an eminently appropriate metaphor for social science measurements.

The demonstrations and examples here are adapted from Mandelbrot (1983) and Lauwerier, (1987).

2.1 The roughness of a coastline

The common introduction of a fractal dimension in physical measurement is in the measurement of a segment of a coastline. The central observation is that the smaller the unit in which the coastline is measured, the longer the measured length of the coastline when all measurements are expressed in the same units. Thus suppose that the measurement of a segment of a coastline is made using a solid stick 1 meter in length by placing the stick successively end on end to follow the coastline as best as possible between the two end points of the segment, and that the measurement is the count of the number of times the meter stick is placed end on end. Suppose next that another measurement of the same coastline is made in the same way using a solid stick of length 10 meters. Then the measurement, expressed in meters say, using the 1 meter stick will be greater than that obtained using the 10 meter stick. The reason for this difference in measurements is that with the smaller 1 meter stick, the details of the coastline are followed more closely than they are with the 10 meter stick, the use of which literally cuts more corners. The measurement using a solid stick of 100 meters would cut even more corners and give an even smaller measurement, as would the measurement in units of 1000 meters.

Mandelbrot (1983) cites the work of Richardson in first making this observation. The measurements were carried out on maps of different scales, but because the detail of the coastline depends on the relative scale of the map, the same effect of the size of the unit on the measurement is present as if the length of the coastline were measured as described in the previous paragraph. Of course, to make the comparison, it is necessary to convert all measurements to the same arbitrary unit, say meters, and the lengths of the

units of measurement to the same unit. Although seemingly obvious, keeping a track of the units of analysis, and agreeing on an arbitrary unit on which to express them, needs to be fore-grounded in social measurement, and is an important part of the paper.

To assess the effect of the change of unit in a systematic way, the units are reduced or amplified successively by a constant factor, which in the above example is an amplification by 10 meters. The successive units can be expressed exponentially as $10^0 m; 10^1 m; 10^2 m; 10^3 m$, giving logarithms to base 10 of 0, 1, 2 and 3 respectively. In this example, base 10 is the constant amplification or scaling factor from one unit of scale to the next. To formalize the relationship between the observed length and the unit of measurement, both are expressed in the same unit, say meters. Then to take account of 10 as the scaling factor, the logarithm to base 10 of the measurement is plotted against the logarithm to base 10 of the unit of the measurement. Mandelbrot reproduces data from Richardson showing graphs of 5 coastlines in which the log of the length is plotted against log of the measuring unit. In each case a straight line shows excellent fit, with negative slope reflecting that as the (log)unit of measurement becomes larger so the (log) measurement becomes smaller. The slope of the line characterizes the roughness of a coastline. An excellent fit to a straight line indicates that a constant amplification factor of the unit of measurement has a constant effect on the size of the measurement.

Table 1 and Figure 2 demonstrate the argument with artificial data for purposes of exposition. Notice that in Table 1, the measurement (31622) in the unit of 1 meter is to the nearest 1 meter, that the measurement (19950) in the unit of 10 meters is to the nearest 10 meters when expressed in meters, that the measurement (12500) in the unit of 100 meters is to the nearest 100 meters when expressed in meters, and that the measurement (8000) in the unit of 1000 meters is to the nearest 1000 meters when expressed in meters. This is the nature of measuring with an instrument which measures in the particular unit inherent or natural to it. Let $N^{(a)}$ be the measurement in units of *a*, that is, it is the count of the number of times the stick of length *a* is used to measure the coastline. For example, with the stick of 10 meters in Table 1, the number of times the

stick was placed end to end is given by $N^{(10)} = 1995$, which is 19950*m* when expressed in arbitrary units of 1*m*.

The regression line of the log measurement on the log unit has a slope of -0.20. However, if the only effect of measuring with different units was the precision of measurement, then the effect across different units would not be systematic, and the regression line would have a slope of 0. That the slope is not 0 indicates that the size of the measurement depends systematically on the size of the unit.

However, if the only effect of measuring with different units was the precision of measurement, then the effect across different units would not be systematic, and the line regression line would have a slope of 0.

The distinction between the natural unit of measurement of an instrument, and the re-expression of this measurement in a designated arbitrary unit which is independent of the instrument, central to Table 1, is also the basic distinction made in (i) the development of the measurement Poisson (Andrich, 2003), (ii) the articulation as to why the collapsing of adjacent categories in the unidimensional Rasch model for ordered categories destroys the model (Andrich, 1995), (iii) the comparison of empirical units of scale across different frames of reference for common items or common persons (Humphry, 2005), and (iv) the demonstration of different units of scale across different formats in setting educational benchmark standards (Heldsinger, 2005). Although relatively clear in physical measurement, this distinction is more subtle in social measurement.

Table 1 Artificial data showing the length of a coastline as a function of the size of the unit of measurement

Natural units of measurement a expressed in meters $a^{(1)}$	Measurement $N^{(a)}$ expressed in units a	Measurement expressed in meters $s_a^{(1)}$	$\log a^{(1)}$	$\log s_a^{(1)}$
1 (10^0)	31622	31622	0	4.50
10 (10^1)	1995	19950	1	4.30
100 (10^2)	125	12500	2	4.10
1000 (10^3)	8	8000	3	3.90

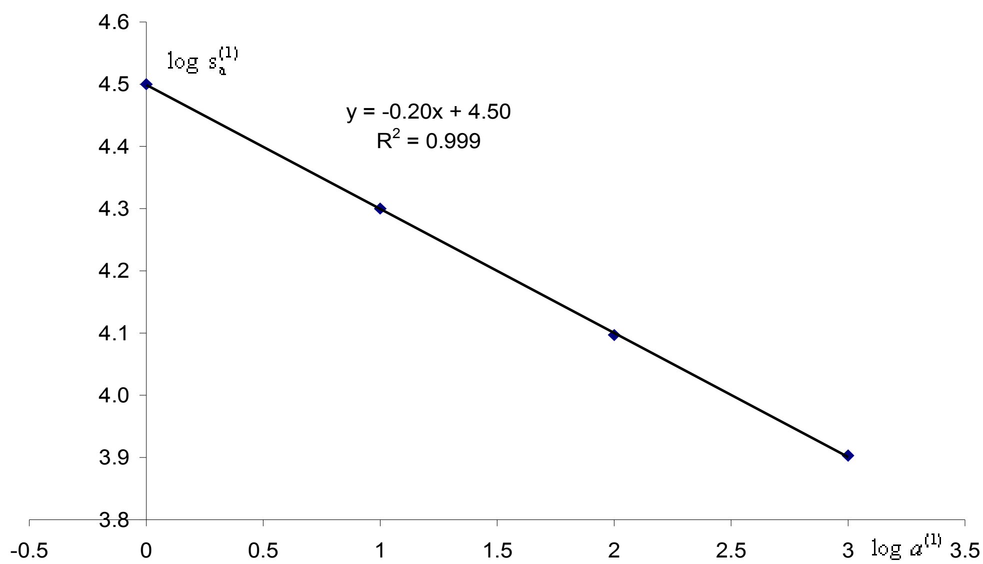

Figure 2. Hypothetical example from Table 1 of the change in log length as a function of the log of the unit of measurement.

Because of this subtlety, the hypothetical example of the fractal dimension of a coastline is presented in some detail in preparation for calculating the fractal dimension of a social measurement. The reason for the subtlety in social measurement between the unit which is an integral part of the instrument and an arbitrary unit in which the measurement can be expressed, is that unlike physical measurement, there are no well defined arbitrary units independent of any instrument, for example, such as the meter. This presents, therefore, a special challenge in deriving and calculating the fractal dimension of a social measurement. The closest one can come to at present distinguishing between an arbitrary

unit and one inherent to the instrument in social measurement, it seems, is to have a specified frame of reference of classes of items and persons which conform to the Rasch model (Rasch, 1961). In that case, the measurement of persons is independent of which particular subset of items is chosen. The special properties of the Rasch model, in which the person and item parameters can be separated in their estimation, are used in this paper to calculate a fractal dimension of a social measurement.

Returning to the case of coastlines in order to demonstrate the definition of a fractal dimension, let the linear relationship in Figure 2 be given by

$$\log s_a^{(1)} = \log s_1^{(1)} + b \log a^{(1)} \qquad (1)$$

where $s_a^{(1)}$ is the measurement expressed in units of 1 meter, but taken with the stick of a meters, that is, in units a meters long, s_1 is the measurement taken with a stick of 1 meter (31622 in Table 1), and b is the slope (-0.20 in Figure 2). The greater the slope of the regression line, the greater the change in measurement as a function of the unit in which measurements are taken, and the rougher the coastline.

From Eq. (1) , and for the coastline of Figure 1,

$$s_a^{(1)} = s_1^{(1)}(a^{(1)})^b = 31622.78(a^{(1)})^{-0.20}. \qquad (2)$$

If the measurements of the length of the coastline were constant irrespective of the unit of measurement, then it would follow that $s_a^{(1)} = s_1^{(1)}, a = 2,3,4$, in which case $b = 0$.

In Eq. (1), in Table 1 and in Figure 2, all measurements, taken with the different units, are already expressed in the common unit of 1 meter – this is signaled by the superscript of (1) in $s_a^{(1)}$ and $a^{(1)}$. The general relationship in Eq.(1) is also conventionally expressed

in terms of the measurement $N^{(a)}$ in the natural unit of the measuring stick as shown below.

From the definition of $N^{(a)}$,

$$N^{(a)}a^{(1)} = s_a^{(1)} \quad . \tag{3}$$

Substituting from Eq. (3) into Eq. (1) for $s_a^{(1)}$ gives

$$\log (N^{(a)}a^{(1)}) = \log s_1^{(1)} + b \log a^{(1)} , \tag{4}$$

that is, $\log N^{(a)} + \log a^{(1)} = \log s_1^{(1)} + b \log a^{(1)}$,

$$\log N^{(a)} - \log s_1^{(1)} = b \log a^{(1)} - \log a^{(1)},$$

and $\quad \log N^{(a)} - \log s_1^{(1)} = (b - 1) \log a^{(1)}$,

from which

$$\log (N^{(a)} / \ s_1^{(1)}) = (b - 1) \log a^{(1)}, \tag{5}$$

and $\quad \log (N^{(a)} / \ s_1^{(1)}) = (1 - \mathrm{b})(- \log a^{(1)})$,

$$\log (N^{(a)} / \ s_1^{(1)}) = (1 - \mathrm{b})(\log(a^{(1)})^{(-1)}),$$

$$\log(N^{(a)} / s_1^{(1)}) = (1 - \mathrm{b})(\log (1/a^{(1)})) ,$$

giving the expression

$$D \equiv (1 - b) = \frac{\log(N^{(a)} / s_1^{(1)})}{\log(1/a^{(1)})} . \tag{6}$$

Eq. (6) gives the definition of the fractal dimension which has the value of D.

In the case of the hypothetical coastline in Figure 2, $D = 1.20$. As anticipated above, in the case where there is no systematic change in the measurement as a function of the size of the unit, the slope $b = 0$ - then the fractal dimension is simply $D = 1$ and the coastline segment is a smooth straight line. In the case of a smooth curve, there are changes in the value of $s_a^{(1)}$ as a function of a, but in the limit as $a \to 0$, the changes become smaller and eventually reach an asymptote beyond which there is no further change. In the examples considered in this paper there is no need to deal with this case. Instead, the case where the data fit a line with finite slope as in Figure 2 is considered sufficient.

This linearity is important in highlighting the other main feature of fractal geometry, that of replication at different levels of scale, or as it is also known, that of self-similarity. The reason the plotted data fit a straight line is that when the scale is magnified, similar relationships among the measurements hold at this new unit of scale. This is demonstrated in greater detail with the example of the Koch motif summarized below in which the self-similarity is perfect and in which case the fit to the straight line is also perfect.

2.2 The Koch motif and self replication

To consolidate the principle of self-similarity, the Koch motif shown in Figure 3 is constructed from first principles. It involves taking a line segment, and then dividing it into some number of segments; in the example of Figure 3, this is 3. Then each new length is 1/3 of the original length. At the second level, the middle segment of 1/3 is removed and replaced by two joined segments of exactly the same length. This increases the actual number of segments, termed the length of the motif, and also increases the roughness of the outline of the motif. This is evident from Figure 3 where the lines of

replication of the successive steps are shown. Although the calculation of the fractal dimension is simpler in this case than that of the coastline, in order to consolidate the approach in anticipation of the case of a social measurement, this simplicity will be demonstrated by first completing a regression of the log lengths against the log units as in Figure 2 of the coastline example.

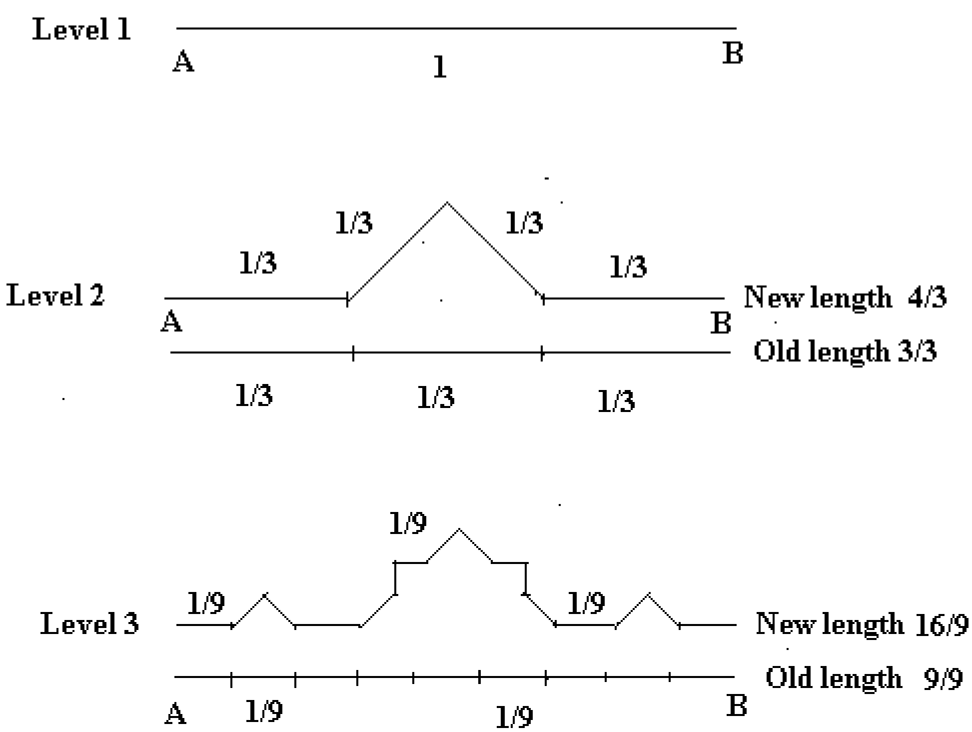

Figure 3. Three levels of self replication at 1/3 the scale for the Koch motif

Table 2 shows the measurements of the Koch motif of Figure 3 in different units, and the corresponding logarithms of the units and the log lengths. Figure 4 shows this regression.

Table 2 Data from the Koch motif of Figure 3

Units of measurement a expressed in terms of a unit length $a^{(1)}$	Length $N^{(a)}$ expressed in units of measurement a	Length $s_a^{(1)}$ expressed in terms of the original unit ($a^{(1)}$)	$\log a^{(1)}$	$\log s_a^{(1)}$
1 (1/ 3^0)	1	(1/1) $= 4^0/3^0$	0log(3)	0
1/3 (1/ 3^1)	4	(4/3) $= 4^1/3^1$	-1log(3)	0.125
1/9 (1/ 3^3)	16	(16/9) $= 4^2/3^2$	-2log(3)	0.250

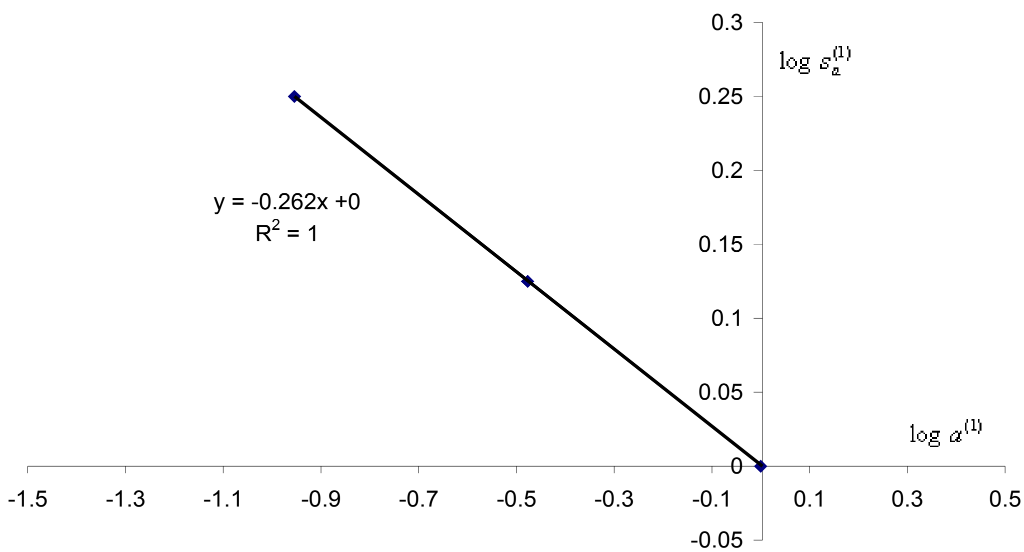

Figure 4. Regression line of the log measurements and log units in the Koch motif

Once again, the slope of the log measurement plotted against the log units is negative: b = -0.262. In this case the correlation is perfect and because $s_1^{(1)} = 1, \quad \log s_1^{(1)} = 0$, the regression line passes through the origin. Accordingly, Eq. (6) can be used to calculate D.

From the first part of Eq. (6), the fractal dimension of the Koch motif is

$$D = (1 - b) = 1 + 0.262 = 1.262.$$

Using the second part of Eq.(6), and for $a \neq 1$,

$$D = \frac{\log N^{(a)}}{\log(1/a^{(1)})} = \frac{\log 4^1}{\log 3^1} = \frac{\log 4^2}{\log 3^2} = \ldots., \tag{7}$$

that is, $D = \frac{\log N^{(a)}}{\log(1/a^{(1)})} = \frac{1\log 4}{1\log 3} = \frac{2\log 4}{2\log 3} = \ldots. = \frac{\log 4}{\log 3} = 1.262.$ (8)

Table 3 shows the values from Eq. (7) for different values of a.

Table 3 Measurements and the units of measurement for the Koch motif

$1/a^{(1)}$	$N^{(a)}$	$D = \frac{\log(N^{(a)})}{\log(1/a^{(1)})}$	D
$1/1=1/(1/3^0=3^0)$	$1=4^0$	Undefined	Undefined
$1/3=1/(1/3^1)= 3^1$	$4=4^1$	$\frac{1\log 4}{1\log 3}$	$\frac{\log 4}{\log 3}=1.262$
$1/9=1/(1/3^2)= 3^2$	$16=4^2$	$\frac{2\log 4}{2\log 3}$	$\frac{\log 4}{\log 3} = 1.262$

3. The Rasch unidimensional model

The paper now turns to a brief summary of points relevant to analyzing data according to the unidimensional Rasch model before applying the above principles, with some adaptations, in calculating the fractal dimension of the measurement provided by a version of the ASAT.

The Rasch model has now been studied and advanced considerably (Fischer, & Molenaar, 1995; Van der Linden, & Hambleton, 1997; Boomsma,. Duijn, & Sniders, 2001) since Rasch's key publications (Rasch, 1960; Rasch, 1961) which are distinguished by the property of sufficiency of the person and item parameters. Following a sequence of derivations in Rasch (1961), Andersen, (1977), Andrich (1978) and Wright and Masters (1982), the unidimensional model takes the form

$$\Pr\{X_{ni} = x\} = \frac{1}{\gamma_{ni}} \exp(\kappa_{xi} + x(\beta_n - \delta_i)) \tag{9}$$

where (i) $X_{ni} = x$ is an integer random variable characterizing $m_i + 1$ successive categories, (ii) β_n and δ_i are locations on the same latent continuum of person *n* and item *I* respectively, (iii) $\kappa_{xi} = -\sum_{k=0}^{x} \tau_{ki}, \tau_{ki}, k = 1,2,3...m_i$ are m_i thresholds which divide the continuum into to $m_i + 1$ ordered categories and which, without loss of generality, have the constraints $\tau_{0i} \equiv 0$, $\sum_{k=0}^{m} \tau_{ki} = 0$, (iv) $\gamma_{ni} = \sum_{x=0}^{m_i} \exp(-\sum_{k=0}^{x} \tau_{ki} + x(\beta_n - \delta_i))$ is a normalizing factor that ensures that the probabilities in Eq. (9) sum to 1. The parameters $\kappa_{xi}, x = 0,1,2,...m_i, \kappa_{0i} = \kappa_{mi} \equiv 0$, are known as category coefficients - they are used for algebraic convenience and are useful in the exposition of this paper. The thresholds, $\tau_{ki}, k = 1,2,3...m_i$, which are points at which the probabilities of responses in one of the

two adjacent categories are equal, are the parameters that can be interpreted as locations on the continuum which define the successive categories.

Replacing the category coefficients according to

$$\kappa_{xi} = -\sum_{k=0}^{x} \tau_{ki}; \quad \tau_{0i} \equiv 0 \tag{10}$$

in Eq. (9) gives

$$\Pr\{X_{ni} = x\} = \frac{1}{\gamma_{ni}} \exp(-\sum_{k=1}^{x} \tau_{ki} + x(\beta_n - \delta_i)) \tag{11}$$

in terms of thresholds.

Figure 5 shows the probabilities of responses in each category, known as category characteristic curves (CCCs) for an item with 5 thresholds and 6 categories, together with the location of the thresholds on the latent trait.

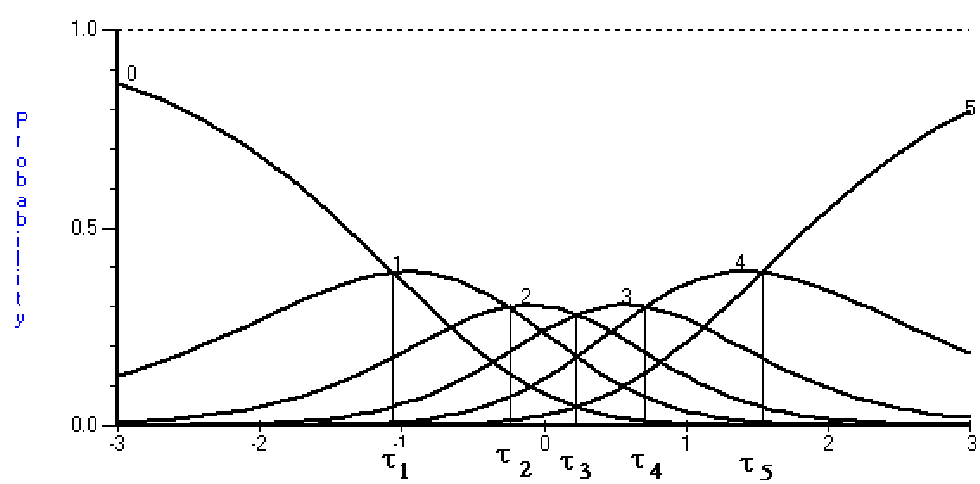

Figure 5. Category characteristic curves showing the probabilities of

responses in each of 6 ordered categories

As indicated earlier, the distinguishing characteristic of the Rasch class of models is that they have sufficient statistics for the parameters. This implies that the person and item parameters are separable so that in particular, the estimates of the item parameters can be obtained independently of the values of the person parameters, (Rasch, 1960; 1961; Andersen, 1977; Andersen and Olsen, 2001).

To illustrate this efficiently for the case of the person parameter β_n , let $\delta_{ki} = \delta_i + \tau_{ki}$, $\delta_{0i} \equiv 0$. Then Eq.(11) simplifies to

$$\Pr\{X_{ni} = x\} = \frac{1}{\gamma_{ni}} \exp \sum_{k=1}^{x} (\beta_n - \delta_{ki}) = \frac{1}{\gamma_{ni}} \exp(x\beta_n - \sum_{k=1}^{x} \delta_{ki}). \qquad (12)$$

In this form, the thresholds δ_{ki} are immediately comparable across items; in the form of Eq. (10), the τ_{ki} are referenced to the location δ_i of item *i*, which is the mean of the thresholds δ_{ki} for each item, that is, $\delta_i = \overline{\delta_{ki}}$. In the case of dichotomous responses, the model specializes to have only the one threshold, $\delta_{i1} = \delta_i$, for each item.

The sufficient statistic for the person parameter β_n is simply the total score $r_n = \sum_{i=1}^{I} x_{ni}$. For example, for a pair of items *i* and *j*, and using directly the responses x_{ni} , x_{nj} , the conditional equation for the pair of responses given the total score r_n , is given by

$$\Pr\{x_{ni}, x_{ni} \mid r_n = x_{ni} + x_{ni}\} = \frac{1}{\Psi_{nij}} \exp \sum_{k=1}^{x_{ni}} (-\delta_{ki}) \exp \sum_{k=1}^{x_{nj}} (-\delta_{kj}), \qquad (13)$$

where $\Psi_{nij} = \sum_{(x_{ni}, x_{nj} | r_n)} \exp \sum_{k=1}^{x_{ni}} (-\delta_{ki}) \exp \sum_{k=1}^{x_{nj}} (-\delta_{kj})$ is the summation over all possible pairs of

responses given a total score of r_n. Eq. (13) is clearly independent of the person parameters β_n, $n = 1,2,...N$ reflecting that the total score r_n is the sufficient statistic for the person parameter β_n. It can be used to estimate the item threshold parameters independently of the person parameters. However, it is not very efficient in the case that there are many categories and particularly if some categories of some items have zero frequency. Different adaptations of Eq. (13), which take advantage of the sufficiency property of the model and which overcome the problems encountered with small or zero frequencies, are available. The software used in the analysis of ASAT in this paper, RUMM2020 (Andrich, Sheridan and Luo, 2005) uses a pairwise conditional algorithm described in detail in Andrich and Luo (2003) and can handle the zero frequencies in categories which are present in some of the analyses. In order not to digress from the substance of the paper, the details of the estimation of the item parameters procedure will not be considered in here. Some issues with estimation are, however, considered briefly in the *Interpretations* section.

Because there are generally many less items than persons, the procedure for estimating the person parameters by conditioning out the item parameters is not feasible. Therefore, the person parameters are estimated following the estimation of the item parameters using the latter as known. Direct maximum likelihood estimates of the person parameters are biased and RUMM2020 also provides a weighted likelihood procedure (Warm, 1989) that virtually eliminates bias for more than 20 or so dichotomous items (Luo and Andrich, 2004).

The sufficiency of the total score r_n for the person parameter β_n has a second critical element for the purposes of this paper. Specifically, for every total score on a given set of items, there is only one corresponding location estimate. It can be calculated from an analysis of a data set even if no person achieves that total score in that particular data set.

3.1 Different arrangements of the data for analysis using the Rasch model

The Rasch model for ordered categories is motivated (Andrich, 1978) by considering a single item which has a response format which involves ordered categories, for example, a rating of Fail, Pass, Credit, and Distinction, of some performance. However, once derived, the model can be applied to data that has the structure of Eq. (12) in that the response variable is a count with successive integers beginning with zero, even if the data do not arise directly from a response in one of $m_i + 1$ ordered categories as in the original motivation. In that case, the parameters of the model are interpreted consistently with the construction of the data in the $m_i + 1$ response categories, and not necessarily as ordered categories on a continuum (Andrich, 1985). This is exploited in the example of the successive analyses of ASAT.

In the analyses of the ASAT data in this paper, the dichotomous items are summed in different combinations before being analyzed by the model of Eq. (12). In particular, in the first such combination, each person's score is summed across items within the 17 reading stems giving each person 17 scores,

$$y_{nj} = \sum_{i=1}^{I_j} x_{nij}\,,\; j = 1,2,3...17\,, \tag{14}$$

where x_{nij} is the response of person n to item i in reading stem j and is therefore an integer ranging from 0 to I_j.

In the analysis of data according to the model of Eq. (12), it is postulated that the person parameter can be characterized by a single real number β_n. In common terminology, the model is *unidimensional* - it characterizes a single linear continuum. The item parameters are located on the same continuum.

The model and the estimation equations further imply that no other factors, other than the person and item parameters and random variation, will affect the response: this implies stochastic independence in the sense that

$$\Pr\{X_{ni} = x_i ; X_{nk} = x_k\} = \Pr\{X_{ni} = x_i\}\Pr\{X_{nk} = x_k\}, \tag{15}$$

in the case of the original items, and

$$\Pr\{Y_{nj} = y_j ; Y_{nh'} = y_h\} = \Pr\{Y_{nj} = y_j\}\Pr\{Y_{nh'} = y_h\}, \tag{16}$$

in the case of the 17 reading stems.

These two features, independence and unidimensionality, which can both be violated are conceptually different and can be empirically different. Features of these violations are presented in greater detail in Andrich (1985) and in the companion paper (Andrich, 2006). In the latter paper the focus is on violations of multidimensionality.

Here we consider briefly one key difference and one key equivalence of the analyses, first when each person's score is the dichotomous response $x_{ni}, x_{ni} \in \{0,1\}, i = 1,2,3...100$ and second, when the person's response is the sum $y_{nj}, y_{nj} \in \{0,1,2,3,...m_j\}, j = 1,2,3,...17$ in the case of the ASAT example.

3.1.1 A difference in analyzing reading stems rather than the original items

Briefly, for the purposes of this paper, the effect of analyzing the responses in terms of reading stems can be understood by referencing the responses to the binomial distribution. Thus suppose that the items within any reading stem all have the same value for their location parameters $\delta_j = \delta_i, i = 1,2,3,...I_j$, and further suppose the responses are stochastically independent according to Eq. (15). Then given the person parameter, the responses of $y_{nj} = \sum_{i=1}^{I_j} x_{nij}$ follow the binomial distribution giving defined values for the

category coefficients according to $\kappa_x = \ln\begin{pmatrix} m_j \\ x \end{pmatrix}$. That is, the Rasch model of Eq. (9) specializes to

$$\Pr\{X_{nij} = x\} = \frac{1}{\gamma_{nij}} \exp(\ln\begin{pmatrix} m_j \\ x \end{pmatrix} + x(\beta_n - \delta_j)), \tag{17}$$

$x \in (0,1,2,...m_j)$.

This means that the thresholds τ_{ki} will have specific values detailed in Andrich (1985) for a series of m_j. In particular, from Eq. (10),

$$\tau_x = \kappa_{x-1,i} - \kappa_{xi} = \ln\begin{pmatrix} m_j \\ x-1 \end{pmatrix} - \ln\begin{pmatrix} m_j \\ x \end{pmatrix}. \tag{18}$$

Suppose now that either the condition of independence or the condition of unidimensionality is violated, that is, suppose that the responses of items within a reading stem have a secondary feature in common with each other, but that it is not common with the items in other reading stems. For example, suppose that unidimensionality is violated according to $\beta_{nj} = \beta_n + c_s\beta_{ns}$ where β_n is the same ability of person *n* as in Eq. (9), β_{ns} is an additional, specific ability component required to answer all the items in reading stem *j*, and c_s is a relative weighting of this component. In the case of no violation, $c_s = 0$. In the case $c_s > 0$ when unidimensionality is violated , the effect on the distribution of responses $Y_{nij} = y_j$ is that they will no longer be distributed binomially – the distribution will be flatter than the binomial in the sense that there will be relatively more responses towards the extremes of 0 and m_j than there are in the binomial distribution. This occurs because a person who has one item right within a reading stem will tend to get others right as well at a greater rate than given by Eq. (9), and vice versa, thus increasing the

rate of scores towards 0 and m_j . The consequence is that, when estimated according to Eq. (9) without taking account of the non unidimensionality, the threshold estimates are closer together than would be obtained according to the binomial coefficients. The effect is shown graphically and numerically in Andrich (1985).

The companion paper (Andrich, 2006) considers this violation of dimensionality of the model in greater detail and derives a summary index of multidimensionality based on the above rationale. For the purposes of the present paper it is sufficient to note that summing the items within reading stems in the ASAT, and in subscales in general, and using these sums as item responses in the model of Eq. (9), *absorbs* features of the responses within the subscales that are unique to the subscale. This, in turn, is reflected in the threshold estimates.

As will be seen in the analysis of the ASAT, however, there are two related additional manifestations in the analysis as a consequence of the effects on the threshold estimates, in particular on the unit of scale of the estimates of abilities for each raw score and on the standard deviation of the person distribution. The paper returns to this issue following the analysis of the ASAT example. It is noted that if no violations of unidimensionality or independence are present, then these manifestations in the item and person parameter estimates are not present, other than through random effects, irrespective of how the items are summed to form subscales. This is shown formally in the companion paper (Andrich, 2006).

3.1.2 An equivalence in analyzing reading stems rather than the original items

Although the item and person parameters estimated will show effects depending on the way the data are analyzed, the total score of a person remains the sufficient statistic for the person parameter, that is, each total score $r_n = \sum_{i=1}^{I} x_{ni} = \sum_{j=1}^{J} y_{nj}$ has associated with it a person parameter. The availability of this estimate, most importantly, does not depend on

any person having any particular total score. As is shown with the ASAT analysis, the feature which is exploited in the paper is that the values of these estimates are different in a very particular way depending on the structuring of the data in the analysis, for example, summing or not summing the scores within the subscales.

4. Analysis of the ASAT data

As indicated above, forming subscales and using the Rasch model to analyze the data accordingly manifest themselves in both the item and person parameters. However, for the purpose of the present paper, only the effects on the person parameters will be considered. Sheridan and Andrich (1996) describe in detail the effects on the item parameters for the first two analyses of the ASAT reported below.

In the present paper, the data are analyzed four times according to the model of Eq.(9): (i) as 100 dichotomous items; (ii) as17 polytomous items formed by summing the dichotomous responses within each reading stem; (iii) as 4 polytomous items formed by summing the responses of the reading stems within each of the disciplines of Humanities, Social Sciences, Mathematics and Natural Science; (iv) as two polytomous items formed by summing the two items of the Humanities and Social Science disciplines and by summing the two items from Mathematics and Natural Sciences disciplines.

Table 4 shows the relevant results. Because the total scores range from 0 to 100, the person location estimates for only selected total scores are shown. Although RUMM2020 extrapolates a location estimate for extreme scores of 0 or 100, the theoretical estimates of these are $\pm\infty$ respectively, and so are not considered. As it happens, no person in this sample had these extreme scores. The total number of people in the analysis was 987.

It is evident from the columns of Table 4 that the values of the person locations for each raw score are not the same across the different analyses. In particular, the extreme values shrink in the successive analyses. Two standard deviation values are also shown in Table 4 for each analysis: first, the standard deviations of the location estimates corresponding

to each total score*, irrespective of how many people have a particular total score*, termed *Std Dev Scale* and notated $S_a, a = 1,2,3,4$; second, the standard deviations of the person estimates which are a function of the number of persons with each total score, termed *Std Dev Distribution,* and notated $s_a, a = 1,2,3,4$.

Table 4 The location value in logits corresponding to each total score from the four different analyses of the ASAT

Total Score r	As 100 Items a=1 ($\hat{\beta}_r$)	As 17 items a=2 ($\hat{\beta}_r$)	As 4 items a=3 ($\hat{\beta}_r$)	As 2 Items a=4 ($\hat{\beta}_r$)
1	-4.488	-3.871	-3.366	-2.782
2	-3.959	-3.347	-2.904	-2.432
3	-3.605	-2.998	-2.601	-2.204
4	-3.335	-2.735	-2.373	-2.030
5	-3.117	-2.524	-2.189	-1.886
6	-2.932	-2.346	-2.033	-1.762
.	.	.	.	.
45	-0.217	-0.123	0.028	0.131
46	-0.171	-0.091	0.049	0.144
47	-0.126	-0.060	0.070	0.156
48	-0.081	-0.029	0.090	0.168
49	-0.036	0.002	0.109	0.180
50	0.009	0.034	0.128	0.192
51	0.054	0.065	0.147	0.202
52	0.099	0.096	0.165	0.213
53	0.144	0.127	0.183	0.223
.	.	.	.	.
94	2.907	2.353	1.723	1.237
95	3.089	2.526	1.863	1.333
96	3.305	2.733	2.031	1.447
97	3.572	2.991	2.242	1.589
98	3.924	3.334	2.526	1.779
99	4.449	3.852	2.964	2.078
Mean	0.000	0.024	-0.002	0.004
Std Dev Scale S_a	1.803	1.436	1.130	0.895
S_a expressed in terms of S_1: $S_a^{(1)} = S_1 / S_a$	1.000	1.256	1.596	2.014
Std Dev Distribution s_a	0.830	0.612	0.395	0.260
s_a expressed in terms of S_1: $s_a^{(1)} = s_a S_a^{(1)}$	0.830	0.769	0.630	0.524

4.1 Calculating the fractal dimension of the ASAT measurement

Clearly, in calculating the fractal dimension of the measurement of ASAT there is no independent unit of measurement and there is no length as such as is available in the measurements of coastlines and of the Koch motif. However, values analogous to both are available in the analysis in Table 4.

4.1.1 Transforming the unit of measurement to the same scale

First consider the unit of the scales. The estimates of values associated with the raw scores for each of the four analyses, which are not a function of the frequencies of observed raw scores, are in their respective natural units of the measurements. Their standard deviations $S_a, a = 1,2,3,4$ are taken to reflect their respective natural units of scale. The natural unit of each analysis is a property of the data, their structure, and the model of analysis. As is evident from Table 4, the values $S_a, a = 1,2,3,4$ vary across analyses. In particular, as the maximum total score increases for the smaller number of items within the successive subscales, so the standard deviation S_a of the scale decreases. Thus the value of the standard deviation $S_1 = 1.803$ for the analysis of 100 discrete dichotomous items is approximately twice the value of the standard deviation $S_4 = 0.895$ for the analysis of two polytomous items with a maximum score of 50 each. This is taken to imply that the size of unit of scale of the former is approximately half that of the latter – the unit with the smaller value is larger when expressed on the same scale.

In summary, the standard deviations of the location values, $S_a, a = 1,2,3,4$ reflect in a well-defined sense that the measurements in the different analyses are in different natural units (Andrich, 1995, 2003). In addition, however, by the construction of the data for the respective analyses, there are also some substantive differences in these measurements at different levels of scale. The ramifications of these differences are considered further in the Interpretations section.

The unit of scale in each of the analyses can be brought to the same arbitrary value by an appropriate transformation of each. In this analysis, the scale of each is brought to the same unit as the first analysis of 100 distinct items and is notated, by analogy to the notation in the coastline example, as $S_a^{(1)}$. This transformation is given by $S_a^{(1)} = \frac{S_1}{S_a}, a = 1,2,3,4$. For example, $S_2^{(1)} = \frac{S_1}{S_2} = 1.256$ in the case of the second analysis with 17 reading stems.

Again, the transformation is analogous to converting the measurement in units of 10 or 100 meters to expressions in terms of 1 meter in the case of the measurements of the coastline.

4.1.2 Transforming the estimates of the standard deviations of persons to the same scale

The standard deviations $s_a, a = 1,2,3,4$ of the person distributions are an indication of the degree to which the ASAT has spread the persons. This standard deviation is taken to correspond to the length in the coastline and Koch motif examples. For the successive analyses, the standard deviations, as for the scales, also decrease, ranging from 0.830 to 0.260. However, as indicated above each is in its own unit of scale and therefore cannot be compared directly. The successive *values* are smaller in part because they are expressed in terms of larger units. To be compared, the standard deviations $s_a, a = 1,2,3,4$ need to be transformed to the same unit of scale.

To transform the $s_a, a = 1,2,3,4$ to the same unit of scale, the same transformation $S_a^{(1)}$ as that for transforming the scales to the same unit is applied: $s_a^{(1)} = s_a S_a^{(1)}$. That is, the standard deviation of each distribution is transformed to the scale of the first analysis with the 100 distinct items. For example for the second analysis as 17 items,

$s_2^{(1)} = s_2 S_2^{(1)} = (0.612)(1.256) = 0.769$. These values are shown in the last row of Table 4. It is clear that when transformed to the same scale, the values of $s_a^{(1)}$ increase – they increase because they are expressed in terms of the smaller unit of the first analysis compared to the units of the subsequent analyses.

4.2 The fractal dimension of the measurement of ASAT

The two transformations, first the standard deviation of the estimates for each raw score, and second the transformation of the standard deviations of the persons into the same units, are possible, it seems, because the estimates of the locations of the total scores can be calculated without involvement of the distribution of the persons. This issue is considered again in the Interpretations section.

The final result to notice in Table 4 is that even after the transformation of the standard deviations of the person distributions to the same unit of scale, the values $(s_a^{(1)})$ are different. In particular, they are smaller with each successive analysis in which it has been considered that the units are successively larger.

By analogy to the calculations of the fractal dimension of a hypothetical coastline and the Koch motif, the remaining step to take, therefore, is to calculate the fractal dimension from the standard deviations in Table 4. This will be done by regression using an equation analogous to Eq. (1), and directly by applying a formula to be derived which is analogous to Eq. (6).

4.2.1 The fractal dimension of ASAT using regression

First, consider the regression approach. The relevant values, $S_a^{(1)}$ taken to reflect the common unit of the first analysis, and $s_a^{(1)}$ which is the measurement taken in the different units but expressed in terms of $S_a^{(1)}$, are reproduced from Table 4 in Table 5 for convenience. These values are augmented by their natural logarithms.

The relevant relationship, again by analogy to the case of the coastline and the Koch motif, is

$$\ln s_a^{(1)} = \ln s_1^{(1)} + b \ln S_a^{(1)} . \qquad (19)$$

Table 5 Values of the standard deviation S_a and s_a expressed in the unit of scale set to $a = 1$ for the 100 item set and their respective logarithms

Analysis		$S_a^{(1)}$	$s_a^{(1)}$	$\ln S_a^{(1)}$	$\ln s_a^{(1)}$
100 items	1	1.000	0.830	0.000	-0.186
17 items	2	1.256	0.769	0.228	-0.263
4 items	3	1.596	0.630	0.467	-0.462
2 items	4	2.014	0.524	0.700	-0.647

Eq. (19) formalizes the non-linear relationship between transformations of the units and the measurements actually taken in the different unit. More explicitly, the measurement changes exponentially relative to the change in unit of scale:

$$s_a^{(1)} = s_1^{(1)} (S_a^{(1)})^b , \qquad (20)$$

again by analogy to Eq. (2).

Figure 6 shows the regression line of Eq. (19) and it is evident that a straight line with slope -0.676 fits the data extremely well, giving a fractal dimension according to the first part of Eq.(6) of

$$\hat{D} = 1\text{-}b = 1.676 .$$

4 2.2 The fractal dimension of ASAT using the ratios of Eq. (7)

Using minimum least squares regression provides one way of estimating D. However, it is instructive to calculate it using the approach analogous to that of Eq. (7). It involves the measurements s_a taken in the natural unit of the analysis.

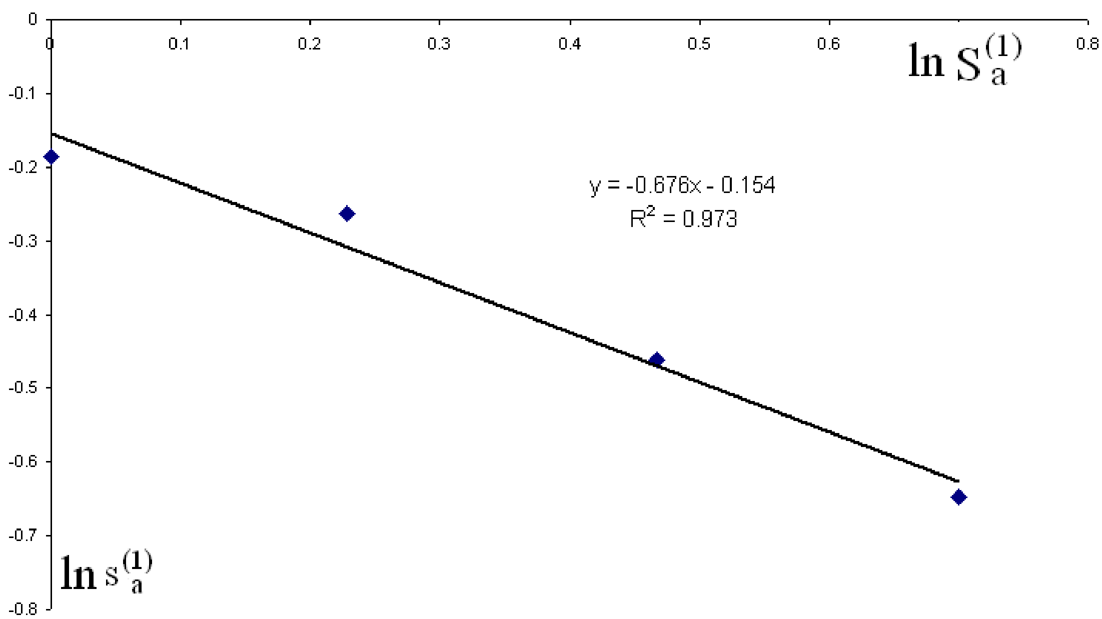

Figure 6. Regression of $\ln s_a^{(1)}$ against $\ln S_a^{(1)}$ for the ASAT data

From Table 4, $s_a^{(1)} = s_a S_a^{(1)}$ and $S_a^{(1)} = S_1 / S_a$. Substituting for $s_a^{(1)}$ and $S_a^{(1)}$ in Eq. (19) gives

$$\ln\left(s_a S_a^{(1)}\right) = \ln s_1^{(1)} + b\ln\left(S_1 / S_a\right),$$

that is,

$$\ln s_a + \ln S_a^{(1)} = \ln s_1^{(1)} + b\ln(S_1 / S_a). \qquad (21)$$

From Table 4 again, $s_1^{(1)} = s_1 S_1^{(1)}$ and $S_1^{(1)} = 1$. Therefore, $s_1^{(1)} = s_1$. Further, $S_a^{(1)} = S_1 / S_a$. Substituting S_1 / S_a for $S_a^{(1)}$ and s_1 for $s_1^{(1)}$ in Eq. (21) gives

$$\ln \mathrm{s_a} + \ln(S_1 / S_a) = \ln s_1 + b \ln(S_1 / S_a). \tag{22}$$

That is,

$$\ln(\mathrm{s_a} / s_1) = b \ln(S_1 / S_a) - \ln(S_1 / S_a), \tag{23}$$

$$\ln(\mathrm{s_a} / s_1) = (b - 1)\ln (\mathrm{S_1} / \mathrm{S_a}), \tag{24}$$

and $$\ln(\mathrm{s_a} / s_1) = (1 - \mathrm{b})\ln (\mathrm{S_a} / \mathrm{S_1}), \tag{25}$$

that is, $\mathrm{s_a} / s_1 = (\mathrm{S_a} / S_1)^{1-b} = (\mathrm{S_a} / S_1)^D$

Finally, for $S_1 \neq S_a$,

$$D = 1\text{-}b = \frac{\ln\ (s_a/s_1)}{\ln\ (S_a/S_1)} \quad . \tag{26}$$

Eq. (26) is analogous to Eq. (7) of the Koch motif. However, unlike the Koch motif, the ratios for the ASAT test are not constant. This is because the regression line of Figure 6 does not reflect a perfect relationship: rather than $R^2 = 1$, $R^2 = 0.973$. If the deviation from a perfect correlation is due only to random variation, then the self-similarity at different levels of scale is preserved. However, if there is systematic variation, as it is likely in real data, then the replication at different levels of scale is much more complex. This point in relation to the ASAT test data is reconsidered in the Interpretations section.

An alternative estimate of D, which is not a least squares estimate, can be obtained from Eq. (26).

Specifically, each ratio in Eq. (26) is an estimate $\hat{D}$ of D. That is,

$$\hat{D}_2 = \frac{\ln(s_2/s_1)}{\ln(S_2/S_1)};\ \hat{D}_3 = \frac{\ln(s_3/s_1)}{\ln(S_3/S_1)};\ \hat{D}_4 = \frac{\ln(s_4/s_1)}{\ln(S_4/S_1)} \tag{27}$$

A single estimate that is more stable than any of the ones in Eq. (27) is found in a similar way to pooling variances in standard analysis of variance calculations:

$$\hat{D} = \frac{\ln(s_2/s_1) + \ln(s_3/s_1) + \ln(s_4/s_1)}{\ln(S_2/S_1) + \ln(S_3/S_1) + \ln(S_4/S_1)}, \tag{28}$$

that is

$$\hat{D} = \frac{\sum_{a=2}^{4} \ln(s_a/s_1)}{\sum_{a=2}^{4} \ln(S_a/S_1)}. \tag{29}$$

Table 6 shows the relevant values from Table 4 for the estimate of D according to Eq. (29).

Table 6 Values of $S_a, \ln S_a; s_a, \ln s_a$

Analysis a (n items)	S_1	S_a	s_1	s_a	$\ln(S_a/S_1)$	$\ln(s_a/s_1)$	$\hat{D}$
a=1;(100)	1.803	1.803	0.830	0.830	0.000	0.000	
a=2;(17)	1.803	1.436	0.830	0.612	-0.228	-0.305	
a=3;(4)	1.803	1.130	0.830	0.395	-0.467	-0.743	
a=4;(2)	1.803	0.895	0.830	0.260	-0.700	-1.161	
				Average	-1.395	-2.208	1.582

The value of $\hat{D}$=1.582 is slightly different from that obtained from the value of $\hat{D}$= 1.676 obtained from least squares regression. As indicated earlier, this difference results from the data not fitting the linear regression perfectly and the latter not being based on the least squares criterion.

From both calculations, it can be concluded that the fractal dimension is different from 1. If the value were 1, then the linear transformation of the unit, and the same transformation of the standard deviations of the distributions, would give the same value for each analysis, and the measurement would be strictly unidimensional. This can be seen explicitly from Eq.(25) which gives

$$\mathrm{s_a} / s_1 = (\mathrm{S_a} / S_1)^{1-b} = (\mathrm{S_a} / S_1)^{D} . \tag{30}$$

Eq. (30) shows once again that when the unit of measurement is transformed linearly in order to express it in a common metric, the actual standard deviations of the measurements expressed in the same units are transformed non linearly.

Setting $b = 0$ and $D = 1$, gives

$$\mathrm{s_a} / s_1 = \left(\mathrm{S_a} / S_1 \right) \tag{31}$$

which then shows that the change in the standard deviations of the measurements would be in the same ratio as the standard deviations of the units from the corresponding analyses, as required.

It is reiterated that if there is no violation of independence or unidimensionality within any of the groupings, and the data fit the dichotomous Rasch model, then there is no change of unit when analyses are carried out with different groupings. Thus these changes are a function of violations of unidimensionality or independence. The algebra of this assertion in the case of the violation of unidimensionality is presented in the companion paper (Andrich, 2006).

It is clear that it would be interpreted from the related analyses that the variable that the ASAT test measures is not smooth. In the sense considered above, the last analysis of the data as two items with maximum scores of 50 each - an analysis which has the largest

unit and which combines discipline areas in pairs - measures the thickest and smoothest variable. It is the thickest in the sense that it characterizes the measurement of a variable that is common to the four discipline areas of Mathematics, Natural Science, Social Science and Humanities. It is the smoothest in the sense that in cutting out the distinguishing aspects of the different discipline in the areas by using large units - it deals only with the variable that is common across all discipline areas. This interpretation is further demonstrated in the companion paper.

The measurement from the first analysis which retains 100 distinct items, is as thick as the last analysis in the sense above, but because it captures the features of all items using smaller units, it is also the roughest. A measurement which would be relatively thinner than either of the first or last analyses, and smoother than the first analysis but rougher than the last, would result, for example, from an analysis of the reading stems within one discipline area. Such an analysis is not reported here where the focus is on the data as a whole.

5. Interpretations and further issues

The conceptualization of a test in terms of fractal geometry outlined above puts new slants on issues that have been canvassed in different ways in other literature, from the traditional factor analysis of raw scores to more modern approaches (e.g. Andrich,1985; Wang, Bradlow &Wainer, 2002; Briggs &Wilson, 2003; Reckase, 1997; Reckase & McKinley, 1991; Wang, Bradlow, & Wainer, 2002). The fractal approach is considered an additional element to the repertoire of methods for dealing with this complex issue.

Some of the new slants may have particular advantages. There are of course many new issues and challenges that it raises. Some of these are also canvassed in this section.

5.1 Possible advantages of the fractal approach

The first general advantage of the fractal approach is that it makes very clear that dimensionality is conceptually and empirically a relative, and not an absolute, concept. It suggests that the typical question that might be asked - *Is a scale unidimensional?* could be changed productively to *How unidimensional is a scale?* For some purposes it may be better to have a thicker and rougher scale in the same domain than for other purposes. The ASAT is an example of a relatively thick and rough scale, measuring a complex variable, when the analysis is that of 100 distinct items.

Second, and following the first point, the fractal further modifies the question of unidimensionality to be explicitly relative to a specified frame of reference as a whole, and not simply relative to one set of items of a scale (Rasch, 1977). The fractal dimension is in part a product of the object of measurement and in part a product of the instrument of measurement within a frame of reference which articulates the material conditions of the engagement between the instrument and the object. In regard to constructs that are considered parts of the entities of measurement, many social measurements are composed of different aspects at different levels of scale (Andrich, 2002). In education, for example, the traditional subjects of mathematics, natural science, language (e.g. English) and so on, are composed of aspects that themselves can be considered in terms of a continuum of achievement. For example, *English* may be defined in terms of the aspects of *listening, speaking; viewing; reading; and writing*; *mathematics* can be considered to be composed of the aspects of *number, measurement, chance and data, space and algebra* (Curriculum Council, 1998). Each of these aspects can, in turn, be divided into thinner constructs, for example, algebra may be further divided into the study of polynomials to the order of two, and so on. Increasing levels of achievement in English or mathematics will involve increasing levels of achievement on the different aspects, and these different aspects will be interwoven in teaching, learning and assessing to provide the thicker and rougher variable. The level at which

measurement is carried out depends on the purpose, and the purpose may be require thinner or thicker and smoother or rougher variables.

Third, the approach used in the ASAT example illustratively in this paper, involved items that were clearly placed into hierarchical levels. It would be possible to consider identifying clusters *post hoc*, and then checking the degree of roughness and thickness of the measurement relative to such classifications. This raises all the same kinds of issues that contrasts confirmatory and exploratory analysis of data and is not entered into here. The strongest argument for the approach demonstrated in this paper is when the construction of the test for a particular purpose has *a-priori* different subscales at different levels, as does the ASAT. The different analyses, which together show the roughness of the measurement, capture the overall changes from one level of analysis to the next.

Fourth, the way of transforming measurements to place them on the same unit of scale used in this paper is different from an alternative common method which involves transforming two measurements of the same people to have the same distribution, and in particular the same mean and standard deviation. It is, instead, analogous to the equating that is carried out in modern test theory when item parameters are anchored from a previous analysis to fix the unit of scale. However, most particularly in applications of the Rasch model, it is typical to carry out the anchoring with a view to fixing the origin, say the mean of common items, and generally not to fixing the unit of scale as well. The unit of scale is effectively fixed when values of some items are anchored rather than just their mean, but this distinction and its many ramifications are generally not considered. Thus the fractal approach brings the unit of scale into profile.

Fifth, the fractal perspective with the ASAT example raises the question as to which analysis is the correct one. It follows that the answer to this question, already implied, is that it depends on the purpose, but that an important part of the fractal analysis is to understand the composition of the construct and the data that reflects it. It supports the position that the task of data analysis according to explicit models is not merely to model

the data but also to understand them (Andrich, 2004). When the data are analyzed as two items with maximum scores of 50 rather than as more items with a lesser maximum score, variations of dimensionality at the lower levels of scale in the hierarchy are absorbed in the responses and the variable is, in this sense, less rough than that obtained with the other analyses - it reflects essentially the variable which has the common features of Humanities/Social Sciences and Mathematics/Science. The variable is also relatively thick. This analysis may be the ideal one for some purposes.

In a complementary manner, the variable analyzed with the 100 distinct items is the thickest and the analysis gives the roughest measurement, because the features of the items, not being absorbed in the response structure, appear in the person estimates and in particular increase the value of the standard deviation, even when expressed in the same unit. Depending on the circumstances, this analysis might be the most appropriate, for example, when predicting a complex variable such as performance at tertiary level. In some cases, considering only the items within each of the major disciplines to form an independent scale may be the most appropriate. However, in making comparisons among any of these measurements, it is seems necessary to consider the unit of scale of each analysis.

Sixth, the fractal perspective draws attention to the property of sufficiency of the total score for the person parameter in the Rasch model. In the estimation of the item parameters, the person parameters were eliminated in each of the analyses in this paper. However, the parameter that is eliminated depends on the structure of the data and the form of the analysis. Sufficiency with the elimination of the person parameter provides many advantages, but it does not imply that the parameter eliminated is exactly the same parameter when the same data are structured differently in the different analyses. Different analyses seem to capture the distinct aspects of the subscales differently, and in this paper, these differences are expressed in terms of roughness.

Seventh, however, the advantage of sufficiency in the Rasch model seems essential in being able to extract two features of the analysis, only one of which is required in the

examples of physical measurement. Thus in the coastline example, the unit of scale is well defined, as is the way of converting one to the other, and the only issue is the actual measurement in the different units. In the analyses of the ASAT above it was necessary to consider the unit of measurement as well as the measurements themselves. Clearly, only the relative values of the unit from different analyses can be obtained, with the unit of the first analysis of the ASAT with 100 distinct items being taken as the arbitrary unit of reference. However, this last feature is no different from the measurement of the coastline example.

Eighth, the analysis inevitably raises the issue of model fit. The quality of fit in general will appear better when the unit of scale is larger, that is, when the vagaries of multidimensionality and dependence are absorbed in the response structure and the measurement is in larger units, for example as two items with a maximum score of 50 rather than 100 items each with a maximum score of 1. The structure of the ASAT, for example, ensures that many items in the latter analysis of 100 distinct items will not fit perfectly a purely unidimensional model. However, it might still be desired to retain the thickness and roughness of the measurement of a single variable rather than characterizing the data multidimensionally. In that case it needs to be recognized that the tests of fit with unidimensional models are in fact tests of deviations of the data from perfect fit to the model, and that items may not show perfect fit, but may still contribute to improving the precision of measurement of a relatively thick and rough variable. This is a perspective well understood informally, but the concept of a fractal measurement formalizes, and in some sense legitimizes, this perspective. Clearly, items need to be assessed to check that they do indeed add to the quality of measurement, that they do not have any technical problems such as more than one right answer in a multiple choice item, that they are reasonably targeted, and so on. However, relatively perfect fit to a thin and smooth unidimensional model may not only be unnecessary for many purposes, but may be counterproductive for some purposes where a thick and rough variable is required to capture a complex construct,. At present, this perspective seems absent – concern tends to be to get as good a fit as possible, including eliminating items that do not fit. It is relevant to note that the two parameter model of Birnbaum (1968) does meet the

requirements because the discrimination of each item is postulated to be a property of the item, whereas in the conceptualization of this paper, any change in unit is brought about by constructing subscales composed from the items.

5.2 Some further issues to be considered from the approach

Having considered some possible advantages of the fractal perspective, it is necessary to indicate some limitations or challenges. First, and continuing from the last point above, ways of dealing with tests of fit at different levels of scale will need to be studied including issues of power at different levels of scale. Perhaps the biggest shift would be somehow to make explicit the criteria external to the analysis for deciding on the relevant level of scale rather than on considering fit as an issue only internal to the data. This will also imply recognizing that decisions on the quality of measurements involve professional judgment rather than only statistical criteria, and therefore are more akin to similar decisions in physical sciences (Hedges, 1987).

Second, and the most obvious one, is the consequence of not having a definition of an independent arbitrary unit of the kind available in physical science. One disadvantage is that it seems not possible to construct a motif of the kind shown with the Koch motif and the idea of roughness needs to be taken metaphorically. The companion paper does show a possible way of constructing a diagram to capture the roughness of a social measurement (Andrich, 2006).

Third, the idea that the amplification factor at successive levels of scale will be the same, that is, the exact self-similarity at different levels of scale will prevail, is unlikely. This of course is also true in physical measurement. Termed affine similarity (Mandelbrot, 2002, p.14), it complicates the calculation of a fractal dimension considerably. For example, in the ASAT data, it does seem that the effect of level of scale is a little different between the 100 item analysis and the 17 item analysis compared to the effects of the other two successive analyses. This can be seen from the graph in Figure 6, where, if the first analysis is removed, the fit to the straight line is virtually perfect –

$\hat{R}^2 = 0.9998$, $\hat{b} = -0.8126$, and $\hat{D} = 1.8126$. In these last three levels of analysis, the self-similarity is therefore also virtually perfect! The change in the log measurements ($\ln(s_a^{(1)})$) relative to the change in the scale ($\ln(S_a^{(1)})$) from the analysis of 100 items to the analysis of 17 reading stems is greater than the change from the analysis of 17 reading stems to the analysis of four discipline areas and from the analysis of four discipline areas to the analysis of two combined discipline areas, with the last two changes being virtually identical.

Fourth, forming the analyses by combining the data at the different levels of scale, generates a substantial number of categories with zero frequencies. For example, in Analysis 3 with four items, category 1 in each item had a zero frequency, and two other categories in each of the items also had zero frequencies. The method of analysis of the present paper can handle this feature. However, it only estimates four of the principal components associated with the thresholds, making the rank of the model of estimation of substantially smaller than the maximum for the successive analyses. Estimation algorithms of full rank which can handle missing data without changing the model need to be developed.

Fifth, the ASAT data set was complete in the sense that every person had completed every item. If there had been missing data, the calculation of the unit of scale would be difficult, as every combination of items provides a sufficient statistic based on different items. Thus although the estimation of parameters is performed in the presence of missing data, the study of the unit of scale would need to be done differently. In the case of random missing data, it would not be a problem to eliminate those people with missing responses. However, this is not as straightforward with structurally missing data. It seems that the study of the scale would need to be carried out through the item parameters. This, too, is not straightforward because every possible combination of items has its own set of total scores for which a location estimate can be obtained. If the fractal approach is pursued, these and other issues will arise and need to be studied.

Finally, an important challenge is to show formal relationships between the approach suggested in this paper and the other approaches that have addressed the same issue of dimensionality of social measurement within a unidimensional framework.

In summary, the paper is an introduction of the ideas of fractal geometry into social measurement and illustrates its application by applying the Rasch unidimensional measurement model to analyze the same data set at different levels of scale. The intention is that this approach be added to the range of approaches reported in the literature that deal with scales that are often, explicitly by their construction, not perfectly unidimensional. It is recognized that much work needs to be done from the fractal perspective to make it applicable. The companion paper (Andrich, 2006) takes this introduction a step further.

References

Andersen, E.B. (1977). Sufficient statistics and latent trait models. *Psychometrika, 42,* 69-81.

Andersen, E. B. & Olsen, L. W. (2001) The life of Georg Rasch as a Mathematician and as a Statistician. In Boomsma, A. , van Duijn, M. A. J. & Sniders, T. A. B. (Eds.) Essays in item response theory. New York: Springer.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43* (4), 561-574.

Andrich, D. (1985). A latent trait model for items with response dependencies: Implications for test construction and analysis. In S. Embretson (Ed.),*Test design: Contributions from psychology, education and psychometrics.* Academic Press, New York. (Chapter 9, pp. 245-273.)

Andrich, D. (1995). Models for measurement, precision and the non-dichotomization of graded responses. *Psychometrika, 60,* (1) 7-26.

Andrich, D. (2002) A framework relating Outcomes Based Education and the Taxonomy of Educational Objectives. *Journal of Studies in Educational Evaluation*. 28, 35-59.

Andrich, D. (2003) On the distribution of measurements in units that are not arbitrary *Social Science Information*, *42*, 4, 557 - 589.

Andrich, D. (2004) Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical Care,* 42, 7 – 16.

Andrich, D. (2006) On the fractal dimension of a social measurement: I.I. Unpublished paper, Social Measurement Laboratory, Murdoch University, Perth. Australia.

Andrich, D. and Luo, G. (2003) Conditional estimation in the Rasch model for ordered response categories using principal components. *Journal of Applied Measurement 4,* 205 - 221.

Andrich, D., Sheridan, B.S. and Luo, G. (2005) RUMM2020: Rasch Unidimensional Models for Measurement. RUMM Laboratory: Perth, Western Australia.

Australian Council for Educational Research. (1989) Australian Scholastic Aptitude Test, Series R. ACER: Hawthorn, Victoria, Australia.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord and M.R. Novick, *Statistical theories of mental test scores* (pp. 397 –545). Reading, Mass.: Addison-Wesley.

Boomsma, A. , van Duijn, M. A. J. & Sniders, T. A. B. (2001) (Eds.) Essays in item response theory. New York: Springer.

Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement*, *4(1),* 87-100.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

Curriculum Council (1998) Curriculum Framework. Curriculum Council. Osborne Park, Perth, Australia.

Fischer, G. H. and Molenaar, I.W. (1995) (Eds.) Rasch models: foundations, recent developments, and applications. New York: Springer.

Hedges, L. V. (1987) How hard is hard science, how soft is soft science? The empirical cumulativeness of research. *American Psychologist, 42 (5)*, 443-455.

Humphry, S. (2005) Maintaining a common arbitrary unit in social measurement. Unpublished PhD Dissertation, Murdoch University, Murdoch, Western Australia.

Heldsinger, (2005). Reporting benchmarks in an invariant metric. Unpublished EdD Dissertation, Murdoch University.

Lauwerier, H. (1987) *Fractals: Endlessly repeated geometrical figures*. Princeton: Princeton University Press.

Luo, G and Andrich, D. (2004). Weighted likelihood estimation of person locations in the Rasch model. *Second International Conference On Measurement In Health, Education, Psychology and Marketing: Developments with Rasch Models.* Murdoch University, Perth, Western Australia.

Mandelbrot, B.B. (1983). *The Fractal Geometry of Nature*. San Franciso**:** Freeman.

Mandelbrot, B.B. (2002) Some reasons for the effectiveness of fractals in mathematics education. In Frame M. L. & Mandelbrot, B.B. *Fractals, graphics and mathematics education.* Mathematical Association of America: Washington.

Rasch, G. (1960/80). *Probabilistic models for some intelligence and attainment tests.* (Copenhagen, Danish Institute for Educational Research). Expanded edition (1980) with foreword and afterword by B.D. Wright, (1980). Chicago: The University of Chicago Press.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman (Ed.). *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. IV*, (pp.321-334). Berkeley CA: University of California Press.

Rasch, G. (1977). On specific objectivity: an attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy 14,* 58-94.

Reckase, M.D., & McKinley, R.L. (1991). The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement, 15 (4)*, 361-373.

Reckase, M.D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement, 21*, 25-36.

Sheridan, B,S., and Andrich, D. (1996) Local dependence and its effect on the discrimination of items. *Paper presented at the Annual Meeting of the AERA, New York,* NY.

Van der Linden, W. & Hambleton, R. (Eds.) (1997). Handbook of Modern Item Response Theory. pp139-152. New York: Springer.

Wang, X., Bradlow, E.T., & Wainer, H. (2002). A general Bayesian model for testlets: Theory and applications. *Applied Psychological Measurement*, *26 (1)*, 109-128.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54,* 427-450.

Wright, B.D. & Masters, G.N. (1982). *Rating Scale Analysis: Rasch Measurement.* Chicago: MESA Press.