UNFOLDING THE MEASUREMENT OF THE CREATIVE PERSONALITY

Abstract

Gough’s Creative Personality Scale (CPS) is a self-report personality inventory for creativity assessment. We investigated the undimensionality and the response process on the CPS from an ideal point (unfolding) perspective. The Graded Unfolding Model (GUM) was used to model binary responses and participants were 228 engineering students who completed a Greek version of the CPS. Results support the undimensionality of the CPS construct and suggest that unfolding measurement models may provide new insights to the assessment of creativity.

Keywords: Creativity measurement, ideal point models, GUM, Greece
INTRODUCTION

Creativity assessments seek to identify the creative individual and this is one of the most exciting adventures in creativity research (Runco, 2007). Psychometric tests (i.e. tests of divergent thinking, attitude and interest inventories, personality inventories, and biographical inventories) are one general approach used in the study of creativity (Hocevar and Bachelor, 1989). Researchers have devoted substantial efforts to demonstrating the psychometric quality of creativity tests; and the importance of understanding the psychometric properties of creativity measures has been recognized (Plucker and Renzulli, 1999). Yet, the techniques used in these efforts have not kept pace with the advances in psychometric theory and methods. In the field of creativity, the application of modern test theory – notably Item Response Theory (IRT) and its variants, such as unfolding models is rare. This is regrettable given that modern test theory may bring more psychometric rigor to the problem of measuring creativity, over and above traditional psychometric techniques.

Unfolding (or ideal point) models is a relative new category of psychometric methods, with a lengthy history of the conceptual underpinnings (e.g., Coombs, 1964; Davison, 1977). Unfolding models belong to the general family the IRT models; however as explained later in the text, they are more general models than the common used dominance IRT models (Stark, Chernyshenko, Drasgow, & Williams, 2006). Recently after a surge of publications aimed at developing the models (i.e., The Hyperbolic Cosine Model – Andrich and Luo, 1993; the Graded Unfolding Model (GUM) – Roberts and Laughlin, 1996; the Generalized Graded Unfolding Model (GGUM) - Roberts, Fang, Cui, and Wang, 2006 - the General Hyperbolic Cosine Model - Andrich, 1996) and estimation
procedures (e.g., De la Torre, Stark, and Chernyshenko, 2006; Luo, 2000; Roberts, Donoghue, and Laughlin, 2002) applications of these models have begun to appear (Drasgow, Chernyshenko, and Stark, 2009); Chernyshenko, Stark, Drasgow, & Roberts, 2007; Scherbaum, Finlinson, Barden, and Tamanini, 2006; Stark, et al., 2006; Weekers, & Meijer, 2008). However, more research is needed to enhance confidence in the usefulness of unfolding models in describing the response process on self-report inventories in general and EI measures in particular.

Our study heeds this call by applying the unfolding theory to a Greek version of the Gough’s Creative Personality Scale (CPS; Gough, 1979) for the Adjective Check List (Gough and Heilbrun, 1965) with a sample of engineering students. As discussed later in the paper, it is plausible that CPS data might be representable along a unidimensional unfolding scale ranging from low to high creativity. At a broader level, the contribution of this study to creativity research is twofold. First, it demonstrates that the individual adjectives constituting the CPS are indeed measuring one general construct. Second, in an informative way it demonstrates the potential usefulness of unfolding models in creativity measurement.

**CREATIVITY AS A PERSONALITY TRAIT**

According to Sternberg, (2006), there are five commonalities in the research of creativity around the world. First, creativity involves thinking that aims at producing ideas or products that are relatively novel and in some respect, compelling. Second, creativity has some domain-specific and domain-general elements; i.e. it needs some specific knowledge, but there are certain elements of creativity that cut across different domains. Third, creativity is measurable, at least to some extent. Fourth, it can be
developed and promoted. And fifth, creativity is not highly rewarded in practice, as it is supposed to be in theory.

It is well recognized that the most difficult task in evaluating creativity is to define what is to be measured; that is, to define creativity. Creativity can be understood as a product, person, press or process (Amabile, 1996; Runco, 2007) or as the interaction among “aptitude, process, and environment” (Plucker, Beghetto, and Dow, 2004). In the present study, creativity is conceived as an undimensional latent trait (Eysenck, 1995). This line of measuring creativity relates to the study of individual differences and personality attributes and includes non-cognitive aspects and set of combined traits. Studies in this line have tried to find characteristics of creative people and they could be divided into psychometric, biographical and historiometric approaches. Biographical and historiometric approaches are mainly related to the study of creative individuals and their context (see Kaufman, Plucker, and Baer, 2008); in psychometrics approaches, an attempt is made to measure facets of creativity associated with creative people (Plucker and Renzulli 1999).

Self-report inventories (such as adjectives check-list, interest and attitude measures) are tools commonly used to assess creativity as a personality trait (e.g. Domino, 1994; Hocevar, 1981; Gough, 1979; Myers, 1962; Zampetakis and Moustakis, 2006; Zampetakis, 2008). The premise underlying the construction of such inventories is twofold: First, optimization of internal consistency reliability is required by including items with high item-test correlations. Second, self-report inventories assume that a dominance process underlies item responding; i.e. the probability of item endorsement increases as the trait level increases (Chernyshenko, et al., 2007). Yet, the
appropriateness of dominance assumptions for the measurement of creativity as a personality trait has not been examined explicitly.

One of the most widely used paper-and-pencil check list of the creative personality (Hocevar, 1981; Oldham and Cummings, 1996) is Gough’s Creative Personality Scale (CPS; Gough, 1979) for the Adjective Check List (Gough and Heilbrun, 1965). The scale consists of 30 adjectives: 18 are indicative of creative individuals and 12 are contraindicative of creative individuals. To assign a CPS score to respondent, researchers follow Gough’s scoring protocol: they first calculate indicative and contraindicative creativity scores separately by adding the numbers of indicative and contraindicative adjectives the respondent seems applicable; then subtracts the contraindicative score from the indicative score. The resulting value, which falls between -12 and +18, is hypothesized to indicate the respondent’s position along a latent creativity dimension. Despite the encouraging evidence about the utility and validity of the CPS scale (e.g. Amabile, Hill, Hennessey, and Tighe, 1994; Carson, Peterson, and Higgins, 2005; Batey and Furnham, 2006; Oldham and Cummings, 1996) no work has explicitly examined whether all 30 adjectives measure in fact the same latent trait.

Gough selected the 30 CPS adjectives using data obtained by 1,701 individuals; the sample covered a wide range of ages and kinds of work (research scientists, mathematicians, architects, psychology graduate students, engineering students, law students, air force officers, Gough, 1979, p. 1400); criteria of creativity were also varied, including ratings by expert judges, faculty members, personality assessment staff observers, and life history interviewers. However Gough’s scoring approach essentially treats all adjectives as equivalent, thereby ignoring the possibility that some items are
more diagnostic of being high (or low) on creativity than are other adjectives. This is in line with studies, suggesting the domain-specificity of personality variables with regards to creativity (e.g. Baer, 1998). Feist (1998) for instance, argued that, although personality dispositions do regularly and predictably relate to creative achievement in art and science, there appears to be temporal stability of these distinguishing personality dimensions of creative people; creative artists and creative scientists do not completely share the same unique personality profiles.

Modern test theory, notably IRT and its variants like unfolding models, overcome the limitation of equivalent items, by explicitly modeling test takers’ answers as a probabilistic process involving these respondents’ own trait levels, and the trait levels implied by the questions.

**UNFOLDING MODELS**

Unfolding (or ideal point) models (Coombs, 1964; Davison, 1977) is a relative new category of psychometric methods and they belong to the general family the IRT models; i.e. they are model based approaches to understanding the non-linear relationships between individual characteristics (e.g., ability, traits), item characteristics (e.g., location), and individuals' response patterns. Contrary however, to well-known IRT models (e.g., Rasch models, two and three parameter logistic models) which are based on the notion of a dominance response processes (monotonic), unfolding models are based on non-monotonic ideal point response processes. Specifically, an unfolding model represents a proximity relation (Coombs, 1964) in which the individual endorses an item to the extent that the individual and the item are located relatively close to each other on a
latent attitude continuum. Furthermore, unfolding models are more general models than the common used dominance IRT models, in the sense that, a dominance model may be considered a special case of an ideal point model, where the ideal point is allowed to be located at infinity (Stark, et al., 2006, p. 35; Weekers, & Meijer, 2008, 67). This implies that under certain circumstances (when for example, the items used are located at the higher and lower ends of the trait continuum) unfolding and dominance models may produce similar Item Response Functions (IRFs-the curves that graphically indicate the probability of item endorsement or agreement against trait level).

Three basic assumptions underlie unfolding theory (Coombs, 1964). First, it is assumed that the trait under investigation is unidimensional and that persons and stimuli (e.g., adjectives in our case) can be located along the same single dimension. Second, it is assumed that there is universal agreement regarding the order of the stimuli along this dimension. Note that this assumption holds even though individuals are expected to differ in their locations along the continuum. Finally, it is assumed that a person’s location is determined by an ideal-point process. That is, a person will be located closest to his or her most preferred stimulus. An implication of this response process is that a person has one of two reasons for not endorsing a given item and that reason depends upon the person’s location. If the person is located higher on the continuum than the item, then the person is said to disagree from above. If the person is located at a lower level, then the person is said to disagree from below (Andrich and Styles, 1998; Chernyshenko, et al., 2007). In other words under the unfolding perspective it is assumed that the probability of agreeing (or endorsing) with a statement is greatest when there is little distance between
an individual's level of latent trait and the level of the trait reflected in the item (Drasgow, et al., 2009; Stark et al., 2006).

Belonging to the IRT family, unfolding models share the same advantages of IRT over classical methods: (1) item parameters are not subpopulation dependent; (2) the person parameter is not specific to the set of items forming the test; and (3) measurement precision is not assumed to be constant; instead IRT methods allow researchers to calculate conditional standard errors of measurement (Embretson and Reise, 2000).

THE GRADED UNFOLDING MODEL (GUM)

Roberts and Laughlin, (1996) developed the GUM to describe ideal point response processes; it can be used with both graded response and binary data. The GUM estimates a single person parameter and two item parameters. The person parameter estimated by the GUM is theta ($\theta$) that is the individuals' level of the latent trait. Theta is expressed as a standardized score; an individual with $\theta=1.0$ has a value on the latent trait that is one standard deviation above the mean. The first parameter estimated by GUM is the item location parameter (delta-$\delta$). This parameter identifies the location of the item on the continuum of the latent trait. The location parameter is on the same metric as theta (i.e., z-scores). Thus, it is used to determine if an individual's level of theta is above or below an item and the size of the difference between the location of the item and the person. The second parameter is the subjective response thresholds ($\tau$). These parameters represent the location of the subjective boundaries between the response options relative to the item location parameter; they are also on a z-score metric. For each item, the number of subjective response thresholds equals the number of objective response options. The value of the subjective response threshold that is associated with the most
positive objective response is set to 0.0. For the GUM the values of $\tau$ are constant across items. Finally, the item discrimination parameter ($\alpha$), for the GUM is set equal to 1 across items.

Mathematically, the GUM is expressed as,

$$P[Z_i=z|\theta_j] = \frac{\exp\left(z(\theta_j-\delta_i) - \sum_{k=0}^{z} \tau_k\right) + \exp\left((M-z)(\theta_j-\delta_i) - \sum_{k=0}^{z} \tau_k\right)}{\sum_{w=0}^{C} \exp\left[w(\theta_j-\delta_i) - \sum_{k=0}^{w} \tau_k\right] + \exp\left[(M-w)(\theta_j-\delta_i) - \sum_{k=0}^{w} \tau_k\right]}$$  \hspace{1cm} (1)

where, $Z_i$ represents the objective response to item $i$, $z$ represents the level of agreement with the item ($z=0$ is the strongest level of disagreement and $z = C$ is the strongest level of agreement), $C$ represents the number of observable response categories minus 1, $M$ equals $2C+1$, $\theta_j$ represents the location of the individual $j$ on the latent trait continuum, $\delta_i$ represents the location of item $i$ on the latent trait continuum, and $\tau_k$ represents the constant across items, subjective response threshold.

**METHOD**

**Participants and procedure**

The sample consisted of 228 engineering students from a small Technical University ($N=1600$) that were taking part in a creativity development program. In addition to several measures concerning the program (the program aimed at the introduction of mind mapping technique in the engineering curriculum, please see Zampetakis, Tsironis and Moustakis, 2007; the rest of the measures included time
management behaviors and attitudes towards mind mapping), participants completed a Greek version of the Gough’s Creative Personality Scale. Students were informed that the program is about developing their creativity and time management behaviors. Their participation was on a voluntary basis; there was no monetary incentive to complete the study. The pencil and paper tests were completed individually in groups of 20. The sample included 102 males and 126 females aged 20-29 years ($M = 23.76$ years, $SD = 2.54$).

Because participants were Greek-speaking, all 30 adjectives were first translated into Greek by two translators, who compared their versions until agreeing on the most correct translation, and then back-translated into English by a bilingual, native English speaking translator, following the procedure recommended by Brislin (1980). The few discrepancies between the original English version and the back-translated version resulted in adjustment in the Greek translation based on direct discussion between the translators.

**Measures**

Gough’s Creative Personality Scale, is a measure that has predicted high levels of creativity across multiple studies and diverse samples. Of the 30 trait adjectives of the scale, 18 are indicative of creative individuals: capable (ikanos), clever (eksipnos), confident (ehemithos), egotistical (egokentrikos), humorous (xioumoristas), informal (aplos), individualistic (atomiostis), insightful (oksiderkis), intelligent (efiis), interests wide (pola endiaferonta), inventive (efebretikos), original (afthentikos), reflective (afthormitos), resourceful (polimixanos), self-confident (me aytopepithisi), sexy (esthisiakos), snobbish (iperoptis), and unconventional (asimbibastos); the remaining 12
are contraindicative of creative individuals: cautious (prosektikos), commonplace (koinos), conservative (sintiritikos), conventional (sigatabatikos), dissatisfied (disarestimenos), honest (entimos), interests narrow (periorismena endiaferonta), well-mannered (kosmios), sincere (elikrinis), submissive (endotikos), suspicious (kaxipoptos), and artificial (anelikrinis). According to Gough’s scoring protocol, 1 point is given each time one of the 18 positive items is checked, and 1 point is subtracted each time one of the 12 negative items is checked. The theoretical range of scores is therefore from -12 to +18.

Reported reliability coefficients for the CPS are often about 0.80 (Cropley, 2000) although Gough and Heilbrun (1965) themselves reported an internal consistency coefficient of 0.63, and test-retest reliabilities of about 0.70, depending on gender. Coefficient alpha (for the 30 items) for the present sample was 0.70.

**Analytic Strategy**

Prior analyses we recoded the obtained data such as each of the checked adjectives was given a value of one; that is the respondent considers the adjective applicable. Non checked items received a value of zero (i.e. not applicable). Item parameters for the GUM were estimated with the GGUM2004 computer program (version 1.1) using a marginal maximum likelihood (MML) approach (Roberts, Donoghue, and Laughlin, 2002; Roberts et al., 2006). GGUM2004 was also used to compute the item and test information functions. Free copies of the program are readily available to readers at: [http://www.psychology.gatech.edu/unfolding/FreeSoftware.html](http://www.psychology.gatech.edu/unfolding/FreeSoftware.html).

The assessment of model-data fit for IRT models, in general and unfolding models in particular is not a trivial task (Drasgow, et al., 1995; Chernyshenko, et al., 2001). Two
alternative approaches to assess model fit have explicitly been advocated to be applied for fitting IRT models: (1) chi-square square goodness of fit tests for single items (singlets), pairs of items (doublets), and three items (triples) (Drasgow, et al., 1995), and (2) graphical inspection of observed vs. expected item response curves (Fit plots; Chernyshenko et al., 2001; Drasgow et al., 1995).

Although GGUM2004 contains item and model fit statistics, such as infit and outfit statistics, according to the technical manual (p.34) these statistics are generalized from cumulative IRT applications and are not mathematically deduced for GUM. Little is known about their distribution, their power, and their Type I error rate. Therefore the MODFIT (version 1.1) computer program (Stark, 2001) was used to compute chi-square statistics and fit plots for the GUM (MODFIT is freely available at: http://work.psych.uiuc.edu/irt).

Following Drasgow, et al., (1995) and Chernyshenko, et al., (2001) we examined model-data fit with fit plots and chi-square square goodness of fit tests for single items, pairs, and triples. Chernyshenko et al., (2001) demonstrated the usefulness of fit plots by showing that some kinds of misfit can be detected by fit plots but not by summary statistics. The idea behind fit plots is to plot both the expected category response function and the observed category response function with approximate 95% confidence intervals (CI; estimated from the observed frequency of category choices, see Drasgow et al., 1995). Deviations of expected and observed category response functions, i.e., if the approximate 95% CIs do not include the expected response functions, indicate model misfit.
For the chi-square square goodness of fit test, indices are formed of the difference between the expected frequency of responses for the options and the observed frequency of responses for the options. Drasgow et al., (1995) found that best fitting models had small, (below 3.0), chi square to degrees of freedom ratios for item singles as well as small ratios for pairs and triples. To summarize the results of chi-square analyses for the model, frequency tables were constructed for the chi-square statistics having values in seven intervals. Means and standard deviations of $\chi^2/df$ ratios were also computed for each subscale.

Furthermore, in the graphical assessment of global model fit, GGUM2004 provides plots of observed and expected responses as a function of theta-delta differences. These theta-delta differences are sorted and then classified into a specified number of homogenous groups of approximately equal size. The average observed and expected responses based on the GUM are calculated for each group. These averages are then plotted against the average theta-delta value for each group. Large discrepancies indicate portions of the latent continuum in which the model does not adequately fit the data (Roberts, et al., 2000).

Finally, if a unidimensional unfolding model like GUM, can fit patterns of responses, then a single latent trait is sufficient to account for item responding (Chernyshenko et al., 2001); that is to say, the proposed unidimensional model provides good fit, and there is little reason to suspect that the data are multidimensional.

**RESULTS**

**Model data-fit and parameter estimates**
Table 1 lists the adjectives used and the associated item parameter estimates (item locations) derived from the GUM. Fit plots for all category response functions indicated no significant misfit and the expected response functions were almost always included in the approximate 95% CIs for the observed response functions. The plot of expected/observed grouped average responses provided visual evidence about the fit of the global model to the observed data (Figure 1).

-------------
Insert Table 1 here
-------------

-------------
Insert Figure 1 here
-------------

Chi-square to degrees-of-freedom ratio fit statistics, provided by MODFIT, showed favourable results indicating the appropriateness of the GUM for the 30 item scale (Table 2).

-------------
Insert Table 2 here
-------------

Since the chi-square tests and the fit-plots did not show large differences between the observed responses and the expected response functions, the unidimensional GUM was accepted for the CPS scale; that is, a single latent trait is sufficient to account for item responding in the CPS.

Adjectives’ and scale properties
The GUM ordered the majority of the adjectives in a logical fashion corresponding to low and high creativity content. Recall that in an unfolding representation the adjectives of the CPS are interpreted as differing in the extend to which they describe low or high creativity.

Specifically the 18 adjectives indicative of creative individuals were all positive; that is they were located on the positive (high) end of the creativity continuum. The location value under the GUM is the point on the latent continuum where the IRF attains its maximum value. For example, on the adjective “Intelligent”, the IRF attains its maximum value at 4.167 (almost four standard deviations above the mean).

For the 12 adjectives that are contraindicative of creative individuals, all but three (honest, well mannered, and sincere), were negative; that is they were located on the negative (low) end of the creativity continuum. For the tree adjectives that were located in the unexpected direction, no evidence of significant misfit was observed.

Six out of the 30 adjectives (capable, clever, confident, honest, well-mannered, and sincere), exhibited nonmonotonic IRFs over the range -4 to +4 showing folding (see Figure 2), which is characteristic of an ideal point response process.

For the remaining adjectives, the empirical IRFs were monotonically increasing, s-shaped functions. For these adjectives, the location parameters were very extreme and there were no respondents located above the adjective’s location. Under the unfolding perspective when the items are extreme relative to the sample of persons, as in our case,
then the empirical IRFs will be approximately monotonic, although they might exhibit a slight degree of nonmonotonic behaviour in the extreme portions of the latent continuum (Roberts et al., 2000). In contrast, items that are located near the center of the person distribution will exhibit markedly nonmonotonic, bell-shaped IRFs. Because of this, unfolding models may be able to be conceptualized as a more general model under which the dominance approach falls.

The mean of the estimated theta distribution was 0.00 with a standard deviation of 0.57. The distribution did not differed significantly from a normal distribution according to the Shapiro and Wilk criterion ($W = 0.989$, $p = 0.073$). Figure 3 illustrates the test information function (TIF) and standard error of measurement for the 30 items in the analysis. The CPS demonstrates the highest levels of measurement precision at the lower levels of the creativity continuum. From the standard error plot it can be seen that the standard error is $\sigma_{GUM} \leq 0.35$ for trait levels of about $-1.0 \leq \theta \leq 0.0$, which is considerably lower than a standard error computed under the CTT approach. Under the classical approach, equal standard errors of measurement would be assumed for all trait levels and estimated from the reliability estimate of the test, e.g., by using Cronbach’s $\alpha = 0.70$ for the present sample, resulting in a standard error estimate of $\sqrt{1-\alpha} = 0.55$.

-------------------------
Insert Figure 3 here
-------------------------

The mean of the CPS scores under the traditional scoring protocol was 3.22 ($SD=2.77$). This value is close to the mean value reported in Gough’s (1979, p. 1402)
original study for 66 engineering students ($M = 3.88$, $SD = 3.94$), providing initial evidence for the validity of the Greek CPS. We explored the correlation between the creativity scores obtained through GUM and the traditional scoring protocol (Cough, 1979). We found a moderately large correlation ($r = 0.82$, $p < 0.0001$). A scatter plot of the theta values about the diagonal line indicated that the theta values are differently ordered, especially at the middle range (Figure 4). Additionally, under GUM non-significant differences were observed between male and female students; however under the traditional scoring protocol, differences were statistically significant with males exhibiting higher creativity scores.

DISCUSSION AND CONCLUSIONS

This study is the first to apply modern test theory in the form of unfolding analysis to a creativity related scale. Specifically, we used the Graded Unfolding Model (GUM) (Roberts, et al., 2000; Roberts, et al., 2006) to model binary responses (endorsement/no endorsement); GUM was successfully fitted to data obtained with the Greek version of the Gough’s Creative Personality Scale (CPS) in a sample of engineering students. The results support the hypothesis that the data form a single unidimensional unfolding scale, although three of the contraindicative adjectives (honest, well mannered, and sincere) were located in the high creativity continuum.

A possible reason for the “misplacement” of the three contraindicative adjectives is that, they were placed in the wrong direction during the initial construction of the scale. In his original paper, Gough presents only 3 examples for the adjectives egotistical, original and conservative and their correlations with the criterion ratings of creativity. However, only one correlation out of four was statistically significant for the adjective
egotistical; only two for the adjective original; and three for the adjective conservative (Gough, 1979, p. 1401). This raises some concern about the characterization of the adjectives as indicative and contraindicative of the creative individual. Gough himself (1979, pp. 1404) noted that the CPS “is moderately valid measure of creative potential”.

Another possible interpretation for the misplacement of the three contraindicative adjectives relates to the domain-specificity of personality variables with regards to creativity (e.g. Dawson, D'andrea, Affinito, and Westby, 1999; Feist, 1998). For instance, creative artists and creative scientists do not completely share the same unique personality profiles. Doyle (1971) found that artists see honesty as more central than other characteristics that are frequently mentioned as typical of creative work (i.e. originality). In their research Dawson, et al., (1999, p.59) reported that college students’ serving as expert judges in the establishment of a pattern of traits that constitutes the creative individual, viewed the creative individual as having such characteristics as "sincere," "well mannered," and "responsible."

Parameter estimates, test information function, and the standard error plot demonstrated that the CPS scale can be used to measure trait creativity accurately at lower to middle scores of the trait level scores; however, at higher trait creativity scores the standard error increases (see Figure 3). Furthermore, our results indicate that all the items of the CPS had extreme item location parameters; this implies that the items included in the scale do not evenly span the creativity continuum and this in turn may adversely affect the accuracy of measurement.

CPS was constructed with classical test theory methods, which encourage the elimination of items located near the center of the person distribution in favor of items
located at more extreme locations (see Table 1). However, we found 6 adjectives showing evident folding. That finding was intriguing given that CPS was constructed with classical test theory methods, which tend to eliminate items showing nonmonotonicity over typical trait level. In order future research to construct a test for the creative personality with a relatively high but flat Total Information Function, more neutral adjectives are required. Neutral items (those with delta parameters close to zero) generally help to measure respondents who are above and below average, whereas positive and negative items provide high information in the middle and at extremes (Chernyshenko et al., 2001).

In conclusion, these analyses were conducted to improve our understanding of the CPS under an unfolding perspective. Single-peaked response functions have been rarely used in substantive research although modern computing algorithms have overcome the problems of time-consuming and laborious traditional analysis of single-peaked response data. Furthermore, unfolding models are more general models than dominance IRT models.

Theoretically, some questions arise from the analyses presented herein. For example, it is interesting to understand, how could an individual disagree from above with the adjective “clever”. The unfolding model suggests that individuals, who are extremely creative, are more reserved in how they use the response options and are therefore less likely to endorse the adjective “clever”. Maybe extremely creative people are more aware of what the term means and how they perceive themselves, making them less likely to endorse the adjective “clever” compared to less creative individuals.
The analyses we have offered have some limitation. First, the small sample size of the present study is restrictive and may be considered small for the GUM. According to De la Torre et al., (2006), it is plausible that MML estimates of GUM item parameters may degrade when binary responses are used with small samples. Furthermore, estimates of standard errors in MML are asymptotic and may be inaccurate when the sample size is small. Our results however presented in Table 1 suggest that the standard errors of the MML estimates were acceptable, even for adjectives with extreme location parameters. However, although, Roberts and Laughlin's (1996) simulations have shown that when responses fit the GUM, then accurate estimates could be obtained with as few as 100 respondents and 15 to 20 items, future research should use larger samples in order to gain confidence about model parameters. The small sample of the study restricted our analyses to the use of the GUM (where response option thresholds are held constant across items and the discrimination parameter is equal to 1 for all items). Future research should apply more complicated unfolding models with larger data sets. In fact the GGUM2004 software contains a set of eight related unfolding IRT models that vary in terms of which item parameters are constrained versus free to vary. Second, the sample used is unrepresentative of the larger population. However, IRT item parameters are subpopulation invariant (Embretson and Reise, 2000). Finally, the study is limited to one particular culture, rather than being cross-cultural. Consequently, future studies might want to consider the implications of our work for different populations. These limitations represent, in any case, opportunities to advance in our efforts, to better understand the nature and measurement of creativity.
ACKNOWLEDGEMENTS

This research has been partially supported by a postdoctoral research scholarship granted to the author by the Greek State Scholarship Foundation (IKY-801/2009).
REFERENCES


Triandis & J. W. Berry (Eds.), Handbook of cross cultural psychology (pp. 398-

structure of the creative achievement questionnaire. Creativity Research Journal,
17, 37-50.

Fitting item response theory models to two personality inventories: Issues and

personality scales under the assumptions of an ideal-point response process:
Toward increasing the flexibility of personality measures. Psychological
Assessment, 19, 88-106.


Davison, M. L. (1977). On a metric, unidimensional unfolding model for attitudinal and
developmental data. Psychometrika, 42, 523-548.

estimation of item parameters for the Generalized Graded Unfolding Model.


Figure 1. Average observed item responses (squares) and average expected item responses (solid line) as a function of mean estimated (theta-delta) value.
Figure 2. Item response functions for the adjectives (a) Capable, (b) Clever, (c) Confident, (d) Honest, (e) Well mannered and (f) Sincere. Each curve represents the expected item response given theta and the item parameter estimates for that item.
Figure 3. The test information function (TIF) for all of the items in the analysis along with the standard error (SE) plot.
Figure 4. Scatter-plot comparisons of creativity estimates from GUM and traditional scoring system (CPS). Every circle represents a person’s trait estimates under the two perspectives.

\[ r^2 = 0.6566 \]
Table 1. Summary statistics for the data and GUM parameter (in descending order) and standard error estimates

<table>
<thead>
<tr>
<th>Adjective</th>
<th>Mean</th>
<th>SD</th>
<th>GUM item parameter (δ)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snobbish</td>
<td>0.05</td>
<td>0.22</td>
<td>7.052 (0.31)</td>
</tr>
<tr>
<td>Individualistic</td>
<td>0.10</td>
<td>0.30</td>
<td>6.323 (0.22)</td>
</tr>
<tr>
<td>Unconventional</td>
<td>0.17</td>
<td>0.38</td>
<td>5.678 (0.17)</td>
</tr>
<tr>
<td>Egotistical</td>
<td>0.17</td>
<td>0.38</td>
<td>5.677 (0.18)</td>
</tr>
<tr>
<td>Insightful</td>
<td>0.21</td>
<td>0.41</td>
<td>5.430 (0.17)</td>
</tr>
<tr>
<td>Self-confident</td>
<td>0.27</td>
<td>0.45</td>
<td>5.035 (0.16)</td>
</tr>
<tr>
<td>Resourceful</td>
<td>0.30</td>
<td>0.46</td>
<td>4.918 (0.16)</td>
</tr>
<tr>
<td>Sexy</td>
<td>0.32</td>
<td>0.47</td>
<td>4.806 (0.15)</td>
</tr>
<tr>
<td>Inventive</td>
<td>0.38</td>
<td>0.49</td>
<td>4.488 (0.14)</td>
</tr>
<tr>
<td>Intelligent</td>
<td>0.45</td>
<td>0.50</td>
<td>4.167 (0.15)</td>
</tr>
<tr>
<td>Wide interests</td>
<td>0.53</td>
<td>0.50</td>
<td>3.820 (0.15)</td>
</tr>
<tr>
<td>Original</td>
<td>0.55</td>
<td>0.50</td>
<td>3.745 (0.14)</td>
</tr>
<tr>
<td>Humorous</td>
<td>0.61</td>
<td>0.49</td>
<td>3.489 (0.15)</td>
</tr>
<tr>
<td>Reflective</td>
<td>0.62</td>
<td>0.49</td>
<td>3.434 (0.13)</td>
</tr>
<tr>
<td>Confident</td>
<td>0.69</td>
<td>0.46</td>
<td>3.119 (0.15)</td>
</tr>
<tr>
<td>Clever</td>
<td>0.72</td>
<td>0.45</td>
<td>2.935 (0.16)</td>
</tr>
<tr>
<td>Well-mannered</td>
<td>0.74</td>
<td>0.44</td>
<td>2.816 (0.17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>-----</td>
<td>-----</td>
<td>------------</td>
</tr>
<tr>
<td>Sincere</td>
<td>0.77</td>
<td>0.42</td>
<td>2.670 (0.17)</td>
</tr>
<tr>
<td>Honest</td>
<td>0.84</td>
<td>0.37</td>
<td>2.239 (0.18)</td>
</tr>
<tr>
<td>Capable</td>
<td>0.85</td>
<td>0.36</td>
<td>2.133 (0.19)</td>
</tr>
<tr>
<td>Cautious</td>
<td>0.60</td>
<td>0.49</td>
<td>-3.532 (0.15)</td>
</tr>
<tr>
<td>Conventional</td>
<td>0.44</td>
<td>0.50</td>
<td>-4.258 (0.16)</td>
</tr>
<tr>
<td>Suspicious</td>
<td>0.36</td>
<td>0.48</td>
<td>-4.598 (0.15)</td>
</tr>
<tr>
<td>Conservative</td>
<td>0.25</td>
<td>0.43</td>
<td>-5.164 (0.17)</td>
</tr>
<tr>
<td>Commonplace</td>
<td>0.15</td>
<td>0.35</td>
<td>-5.888 (0.20)</td>
</tr>
<tr>
<td>Submissive</td>
<td>0.14</td>
<td>0.35</td>
<td>-5.927 (0.21)</td>
</tr>
<tr>
<td>Narrow interests</td>
<td>0.08</td>
<td>0.27</td>
<td>-6.599 (0.26)</td>
</tr>
<tr>
<td>Dissatisfied</td>
<td>0.07</td>
<td>0.25</td>
<td>-6.802 (0.28)</td>
</tr>
<tr>
<td>Artificial</td>
<td>0.02</td>
<td>0.13</td>
<td>-8.200 (0.29)</td>
</tr>
</tbody>
</table>

*Note: For the GUM, item discrimination parameter \(\alpha\) is set equal to 1 across items; the values the subjective response thresholds \(\tau\) are constant across items and equal to -3.947. Standard error estimates appear in parentheses.*
Table 2. Frequencies of the values of the chi-square statistic to degrees of freedom from the model fit analysis for the GUM model for all 30 adjectives of the CPS

<table>
<thead>
<tr>
<th>Model</th>
<th>&lt;1</th>
<th>1 - &lt;2</th>
<th>2 - &lt;3</th>
<th>3 - &lt;4</th>
<th>4 - &lt;5</th>
<th>5 - &lt;7</th>
<th>&gt;7</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUM</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.27</td>
<td>0.18</td>
</tr>
<tr>
<td>Doubles</td>
<td>20</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1.21</td>
<td>1.68</td>
</tr>
<tr>
<td>Triples</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.66</td>
<td>1.29</td>
</tr>
</tbody>
</table>