Applied Mathematical Sciences, Vol. 1, 2007, no. 58, 2851 - 2862

An Empirical Study on the Transformation of Likert-scale Data to Numerical Scores

Chien-Ho Wu

Graduate School of Applied Statistics, Fu-Jen Catholic University
510 Chung-Cheng Rd., Hsin-Chuang Taipei 24205, Taiwan
stat2016@mails.fju.edu.tw

Abstract

The application of statistical methods to data analysis requires that the data set concerned should follow some particular assumptions. For example, AVOVA assumes that the response variable is normally distributed within groups, and the variances in the different groups are identical. However such assumptions are generally not observed by data collected through Likert Scales. This paper presents a computation procedure for transforming Likert-scale data into numerical scores that better follow the assumption of normality, based on the scaling procedure proposed by E. J. Snell. We have also conducted an empirical study to investigate the effects of the proposed transformation on data analysis. Finally this paper addresses the decision on whether or not that Likert-scale data should be transformed to scores that are more compliant to statistical assumptions.

Mathematics Subject Classification: 62P25

Keywords: Likert-scale, E. J. Snell, Subjective Scoring

1 Introduction

In the arena of social sciences, Likert scale (Likert, 1932) is a popular instrument to measure constructs such as attitudes, images and opinions. To facilitate data analysis, each response category on the scale is generally assigned successively an integer value. However assigning successive integer values to scale categories has also been criticised for not being realistic. For example, for an injury scale of five categories represented by none, minor, moderate, severe, and fatal, the degree of injury seriousness between severe and fatal is more significant than that between none and minor. Assigning successive integers to the scale categories would not reflect the realistic differences in injury seriousness between or among scale categories.

Although Likert-scale data can be analysed by nonparametric procedures (Agresti, 2002; Fleiss, 1981), applying parametric procedures to Likert-scale data analysis is still conveniently adopted by researchers in social sciences. Having said that, the application of parametric procedures to data analysis requires that the data set concerned should conform to some statistical assumptions. For instance, (1) both AVOVA and regression analysis assume, among others, that the observed response variables are normally distributed. (Montgomery and Runger, 2006; Neter et. al. 1992) (2) Parametric tests for population mean rely on the assumption that the sample data has an approximate normal distribution. Assumptions such as these are generally not observed by data collected through Likert Scales.

To address the aforementioned problems, we introduce a computation procedure based on E. J. Snell's scheme that can efficiently transform Likert scale data into more realistic and normally distributed data for analysis. This paper also presents an empirical study on two surveys to evaluate the effects of the transformation on data analysis. Finally we give some suggestions as to whether or not it is a good practice to proceed with the transformation.

2 REVIEW of E. J. SNELL's SCALING PROCEDURE

2.1 The Paradigm

An optimal scoring procedure was first suggested by Fisher (1938). However it takes no account of scale order and distribution assumptions, and thereby could incorrectly reject the null hypothesis in significance testing. In light of the problems arise from analysing subjective measurements in many fields, Snell (1964) presented a method of determining numerical scores for the categories of subjective scales such as Likert scale. The scores so determined, as proclaimed by E. J. Snell, are suitable for use in methods of analysis requiring assumption of normality.

The Snell method assumes that there is an underlying continuous scale of measurement along which the scale categories represent intervals. For a scale of k categories, it is defined that category s_j corresponds to the interval x_{j-1} to x_j as shown by Figure 1.

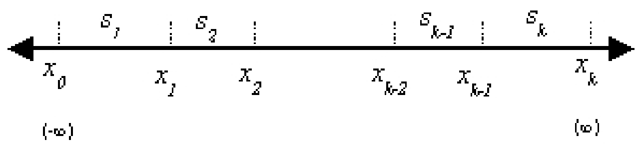

Figure 1: Intervals for scale categories.

Assume that there are m groups of observations, then the probability of an observation of group i in category s_j is defined as:

$$P_i(x_j) - P_i(x_{j-1}), \ \ where\ i = 1,\dots,m;\ j = 1,2,\dots,k \tag{1}$$

And the underlying continuous distribution function $P_i(x_j)$ takes the form:

$$P_i(x_j) = \frac{1}{1 + e^{-(a_i + x_j)}} \tag{2}$$

which is a logistic function with mean $-a_i$ and variance $\pi^2/3$. Since the logistic distribution extends in both directions to infinity, x_0 and x_k take the values $-\infty$ and ∞ respectively, and correspondingly $P_{i0} = 0$ and $P_{ik} = 1$. Also since the choice of origin is arbitrary, x_1 can be assigned 0.

Let n_{ij} denotes the number of observations of group i in category s_j and N_j the total number of observation in category s_j, i.e.

$$N_j = \sum_{i=1}^{m} n_{ij} \tag{3}$$

Then the maximum likelihood estimates of the parameter x_j can be obtained by solving the equations (4) and (5).

$$\hat{x}_{k-1} - \hat{x}_{k-2} = \ln\left(\frac{N_{k-1}}{\sum\limits_{i=1}^{m}(n_{i,k-1} + n_{ik})\hat{P}_{i,k-1} - N_{k-1}} + 1\right) \ \ for\ j = k-1 \tag{4}$$

$$\hat{x}_j - \hat{x}_{j-1} = \ln\left(\frac{N_j}{\sum\limits_{i=1}^{m}(n_{ij} + n_{i,j+1})\hat{P}_{ij} + \frac{N_{j+1}}{e^{(\hat{x}_{j+1} - \hat{x}_j)} - 1} - N_j} + 1\right) \ \ for\ j = 2,\cdots,k-2 \tag{5}$$

2.2 The Approximate Solution

Snell had shown that an adequate approximate solution for x_j can be achieved by replacing the theoretical proportions $\hat{P}_{ij}$ in equations (4) and (5) by the *observed* (accumulative) proportions p_{ij}. Provided that there are no obvious irregularities in the data that may cause significant discrepancies between the theoretical and observed proportions, the estimates of x_j thus obtained will be close to the true values for $\hat{x}_j$. Once the class boundaries x_j are estimated, mid-points are taken as scores for categories, i.e.

$$s_j = \frac{(x_j + x_{j-1})}{2} \; for \; j = 2, 3, \cdots, k-1 \qquad (6)$$

For the extreme category s_k, we take $(x_{k-1}+1.0)$ or $(x_{k-1}+1.1)$ respectively as its score according to whether the average of the observed proportions in category x_k (i.e. $1 - p_{i,k-1}$) is less than 0.10, or lies between 0.10 and 0.20. A point with a similar distance below x_1 is taken as the score for s_1.

2.3 Some Limitations

As pointed out by Snell, the application of the scaling procedure should take into account the following facts:

1. The method takes no account of the experiment design behind the data.
2. The method is not applicable if there is only one observation in each group.
3. There should be relatively few observations in the extreme scale categories.
4. If the grouping is coarse, the estimation and analysis of the parameters a_i is preferable to the use of scores.
5. Irregularities in data should be within tolerance. In cases of doubt, iteration to a more accurate solution is recommended.

These facts constitute a check list for deciding on the appropriateness of the data set for scaling.

3 The COMPUTATION PROCEDURE

Before the transformation we need to (1) check the validity of the collected data, (2) preprocess missing values in the data set, (3) evaluate the appropriateness of the transformation, and (4) select a computation tool. It is possible to use a spreadsheet-like tool such as Excel to do the approximate transformation. However, it may be error prone and troublesome when the size of the data set is large. This paper presents an algorithm as shown in the following section that can be used as the core of a generalized tool for scoring.

3.1 The Algorithm

The following is an algorithm for scoring based on Snell's scaling procedure. The algorithm is straightforward and simple with the idea of KISS principle in mind for illustration purposes. A more detailed description of the algorithm can be found in the appendix.

Procedure SnellTransformation (#OfRespondents, #OfScaleItems, #OfCategories)
BEGIN

1. **declare** integer array scaleData [#OfRespondents+1][#OfScaleItems+1],
   integer frequencies [#OfScaleItems+1][#OfCategories+1],
   integer array totalForNj [#OfCategories+1],
   integer array totalForGi [#OfScaleItems+1],
   integer array accumulativeFrequencies [#OfScaleItems+1][#OfCategories+1];
2. **declare** real array accumulativeProportions [#OfScaleItems+1][#OfCategories+1],
   real array interval[#OfCategories+1],
   real array X [#OfCategories+1],
   real array scores[#OfCategories+1];
3. Initialize all declared array elements to 0.
4. Populate scaleData array.
5. Compute the frequency of each category (n_{ij}) for each scale item. Let frequencies[i][j]=n_{ij}.
6. Compute the total frequency (N_j) for each response category. Let totalNj[j]=N_j.
7. Compute the total response frequency (G_i) for each scale item. Let totalForGi[i]=G_i.
8. Compute the accumulative frequencies of response categories (f_{ij}) for each scale item. Let accumulativeFrequencies[i][j]=f_{ij}.
9. Compute the accumulative proportions for response categories (p_{ij}) for each scale item. Let accumulativProportions[i][j]=p_{ij}.
10. Compute the intervals ($x_j - x_{j-1}$) corresponding to the non-extreme response categories.
11. Compute *interval boundaries* (x_j) for non-extreme response categories.
12. Compute scores (s_j) for *non-extreme* response categories. Let score[j]=s_j.
13. **If** the avg of the observed proportions in the first category is less than 0.10 **then**
    score[1] = X[1] -1.0
    else
    score[1] = X[1] -1.1
14. **If** the avg of the observed proportions in the last category k is less than 0.10 **then**
    score[k] = X[k-1] + 1.0
    else
    score[k] = X[k-1] + 1.1

END

3.2 Complexity of the Algorithm

The memory space required for the execution of the algorithm is declared by statements 1 and 2. For input data set of size n (defined as: $\#OfRespondents \times \#OfScaleItems$) the memory space for data needed for computation can be estimated by the following expression, assuming that each array element requires 4 bytes of memory space:

$$4_{bytes} * [(\#OfRespondents * \#OfScaleItems) + 3 * (\#OfScaleItems * \#OfCategories) + 4 * \#OfCategories + \#OfScaleItems]$$

It can be easily seen from the above expression that the total space requirement is a linear combination of terms dominated by $\#OfRespondents$ and $\#OfScaleItems$. Under the uniform cost criterion (Aho et.al., 1976), the space complexity is simply *O(n)*.

The time complexity of the algorithm is dominated by the loops spread across the rest of the algorithm. Since many of the instructions within loops perform simple computations only, it is reasonable to assume the uniform cost criterion for time complexity, and thus the total execution time is a linear combination of the execution time for statements 3 through 14. Without lost of generality, the time complexity of the algorithm is *O(n)*.

4 The CASE STUDY

4.1 The Melancholic Survey

To investigate the effect of the transformation on Likert-scale data analysis, we have conducted a field survey on the melancholic status of junior high school students by using a well-recognized Melancholy Self-Testing Scale provided by John Tung Foundation[1]. The scale is a 4-category scale consisting of 18 items. Alongside the aforementioned scale, the questionnaire for the survey also contains items related to the demographic characteristics of respondents.

After validation and verification of the collected data, we generated two sets of data for analysis from the original survey data, each consists of 398 valid observations. The first data set is derived from assigning successive integers to scale categories. The second data set is generated by applying the Snell scaling procedure to the categories of the scale. The scores for the scale categories, shown below, in the second data set are computed by resorting to the procedure as described in section 2.

$$S1{=}{-}1.0 \quad S2{=}0.704 \quad S3{=}1.974 \quad S4{=}3.541$$

[1]The web site for John Tung Foundation can be reached at http://www.jtf.org.tw/

Since the scale is a summation scale, the computed scores have been adjusted to 1, 2.704, 3.974 and 5.541 respectively for data analysis. The salient distribution features of the two data sets are shown in Table 1.

Item#	Data set 1 (Integer scores)			Data set 2 (Snell scores)		
	skewness	kurtosis	normality	skewness	kurtosis	normality
1	1.715	2.713	X	1.494	1.787	X
2	0.624	-0.329	X	0.447	-0.423	X
3	0.860	-0.194	X	0.713	-0.401	X
4	0.818	-0.537	X	0.711	-0.651	X
5	2.042	4.542	X	1.780	3.085	X
6	1.263	0.711	X	1.114	0.330	X
7	0.737	-0.493	X	0.609	-0.614	X
8	0.932	-0.164	X	0.801	-0.363	X
9	0.426	-0.907	X	0.316	-0.880	X
10	0.751	-0.447	X	0.613	-0.567	X
11	0.617	-0.535	X	0.472	-0.598	X
12	1.176	0.670	X	0.997	0.229	X
13	1.232	0.452	X	1.112	0.177	X
14	0.935	-0.356	X	0.839	-0.492	X
18	2.490	5.295	X	2.382	4.767	X
16	1.840	2.576	X	1.690	2.012	X
17	1.304	0.544	X	1.197	0.292	X
18	1.477	1.011	X	1.375	0.731	X
Total Score	1.032	0.801	X	0.951	0.621	X

Note: (1) X represents the rejection of normality. (2) α=0.05 (3) Statistics computed by SPSS.

Table 1: The salient distribution features of the data sets.

Table 2 illustrates the important statistics when applying the selected statistic techniques to the data sets.

Techniques	Data set 1 (Integer scores)	Data set 2 (Snell scores)
Factor Analysis	Cronbach's α = .923 KMO = .932 Bartlett's test of sphericity: (p=.000**) Number of λ>1: 3 Total variance explained: 56.52% Grouping of items after Varmax transformation: same as data set 2	Cronbach's α = .923 KMO = .934 Bartlett's test of sphericity: (p=.000**) Number of λ>1: 3 Total variance explained: 56.44% Grouping of items after Varmax Transformation: same as dataset 1
Pearsonχ^2 Test	◇Gender(Male vs. Female): .000** ◇Sports Intensity (Strong vs. Normal) For DF Junior High School: .000** For KC Junior High School: .223 For Data Set: .475 ◇Sports Type vs. Melancholy Classification: .561 ◇School vs. Melancholy Classification: .704	◇Gender(Male vs. Female): .000** ◇Sports Intensity (Strong vs. Normal) For DF Junior High School: .004* For KC Junior High School: .218 For Data Set: .345 ◇Sports Type vs. Melancholy Classification: .391 ◇School vs. Melancholy Classification: .183
AVOVA	Sports Type vs. Melancholy Status: .521	Sports Type vs. Melancholy Status: .534

Note: (1) α=0.05 (2) Statistics computed by SPSS.

Table 2: Results from applying selected data analysis techniques.

4.2 The Occupational Stress Survey

The survey data related to the occupational stress of Taiwanese labour (Li, 2002) is used as another example for illustration. In that particular survey, the occupational stress is measured by a modified six-category scale derived from the OSI-2 scale items. The scores for the response categories of the scale are shown below:

$$S1=-1.1,\quad S2=0.458,\quad S3=1.394,\quad S4=2.456,\quad S5=3.744,\quad S6=5.448$$

Since the scale is also a summation scale, the computed scores have thus been adjusted to 1, 2.558, 3.494, 4.556, 5.844 and 7.548 respectively for data analysis. The salient features of the two data sets are shown in Table 3.

Features	**Data set 1 (Integer scores)**	**Data set 2 (Snell scores)**
Distribution features	None of the items conforms to the normality assumption.	None of the items conforms to the normality assumption. However, skewness has been improved.
Factor Analysis	Cronbach's α=.959 KMO=.952 Bartlett's test of sphericity:(p=.000**) Number of λ>1: 7 Total variance explained: 60.174% Grouping of items after Varmax transformation: slightly different from data set 2.	Cronbach's α=.958 KMO=.950 Bartlett's test of sphericity:(p=.000**) Number of λ>1: 7 Total variance explained: 59.869% Grouping of items after Varmax Transformation: slightly different from data set 1.
AVOVA	◇Education level vs. stress: p=.016* Homogeneity of variances: p=.051 ***Age group vs. stress: p=.058*** ***Homogeneity of variances: p=.532*** ◇Work shift vs. stress: p=.367 Homogeneity of variances: p=.000** ◇Sports type vs. stress: p=.296 Homogeneity of variances: p=.000** ◇Exercise frequency vs. stress: p=.338 Homogeneity of variances: p=.000** ◇Sports intensity vs. stress: p=.054 Homogeneity of variances: p=.233 ◇Job title(manager or otherwise) vs. stress: p=.919; Homogeneity of variances: p=.669 ◇Years of service vs. stress: p=.026* Homogeneity of variances: p= .002*	◇Education level vs. stress: p=.022* Homogeneity of variances: p=.036* ***Age group vs. stress: p=.047**** ***Homogeneity of variances: p=.621*** ◇Work shift vs. stress: p=.286 Homogeneity of variances: p=.000** ◇Sports type vs. stress: p=.367 Homogeneity of variances: p=.001** ◇Exercise frequency vs. stress: p=.334 Homogeneity of variances: p=.000** ◇Sports intensity vs. stress: p=.063 Homogeneity of variances: p=.236 ◇Job title(manager or otherwise) vs. stress: p=.856; Homogeneity of variances: p=.576 ◇Years of service vs. stress: p=.025* Homogeneity of variances: p=.002*

Note: (1) α=0.05 (2) Statistics computed by SPSS.

Table 3: The salient features of the occupational stress data sets.

4.3 Discussions

Based on results shown in Table 1, Table 2 and Table 3, we have the following findings:

1. In both scenarios, transformation of scale data based on E. J. Snell scaling procedure does not do much in making the transformed data sets pass the normality test. However, the transformed data sets, in general,

conform better to normality.

2. The statistics calculated for factor analysis show that the transformation does not incur much difference in exploring and interpreting factors. Still, scale items strongly connected tend to be grouped together in both scenarios.

3. The Pearson χ^2 tests in the first scenario led to the same decisions for both data sets.

4. Not surprisingly, since the data sets in both scenarios do not pass the normality test, many ANOVA computations failed the tests for variances homogeneity. Having said that, the transformation does have marginal effects on age groups.

As highlighted in section 2, the scaling procedure has its limitations and assumes that the theoretical probability function is logistic. The transformation does not guarantee that the derived data set will pass normality tests.

5 CONCLUSIONS

Likert scale is a popular instrument to collect subjective data such as attitudes, images and opinions in the field of social sciences. Applying parametric procedures to Likert-scale data relies on assumptions about distribution properties of data, which are often accepted to be true, or are considered irrelevant. And this convenient approach to either accepting or ignoring statistical assumptions for parametric methods is considered to be too optimistic. Nonetheless, assigning integer scores successively to scale categories is often criticised for being not realistic and not conforming to the assumptions required by parametric methods.

The scaling procedure proposed by E. J. Snell is a candidate procedure, as pointed out by Fleiss (1981), for determining the scores for the categories of subjective scales. In this paper we have introduced the concepts and the computation procedure for determining scores for scale categories based on Snell's scaling procedure. The computation procedure has linear complexity in terms of time and space requirements. The study on the two surveys presented in this paper shows that the distribution of the transformed data sets tilts more toward normality; and the results from applying the selected data analysis techniques are very much the same for both data sets in each survey scenario. If management practice allows, we would avoid assigning successive integers to scale categories because integer scoring easily attracts doubts and criticism.

There are other approaches, such as transformations based on fuzzy theory and random-effects models (Fielding, 1999), to dealing with the scoring of subjective categories. In the future, we may consider realizing other transformation procedures and conduct more thorough comparisons based on real world survey data. Furthermore, we may also consider developing the algorithm and system to obtain the iterative estimates of a_i and x_j for E. J. Snell's scaling procedure.

APPENDIX:Detailed Algorithm

```
Procedure SnellTransformation (#OfRespondents, #OfScaleItems, #OfCategories)
BEGIN
declare integer array scaleData [#OfRespondents+1][#OfScaleItems+1],
   integer frequencies [#OfScaleItems+1][ #OfCategories+1],
   integer array totalForNj [#OfCategories+1],
   integer array totalForGi [#OfScaleItems+1],
   integer array accumulativeFrequencies [#OfScaleItems+1][ #OfCategories+1];
declare real array accumulativeProportions [#OfScaleItems+1][ #OfCategories+1],
   real array interval[#OfCategories],
   real array X [#OfCategories+1],
   real array scores[#OfCategories+1];

Initialize all declared array elements to 0.
Populate scaleData array.

// Compute the frequency of each category (nij) for each scale item.
// Let frequencies[i][j]=nij.
   for (r=1; r<=#OfRespondents; r++)
      for (s=1; s<=#OfScaleItems; s++)
         begin
           response=scaleData[r][s]
           if (response >=1) and (response <=#OfCategories) then
              frequencies[s][response]++
         end

// Compute the total frequency (Nj) for each of response category.
// Let totalNj[j] = Nj.
   for(c=1; c<=#OfCategories; c++)
      for (s=1; s<=#OfScaleItems; s++)
         totalForNj[c]=totalForNj[c]+frequencies[s][c]

// Compute the total response frequency (Gi) for each scale item.
// Let totalForGi[i]=Gi.
   for(s=1; s<=#OfScaleItems; s++)
      for(c=1; c<=#OfCategories; c++)
         totalForGi[s]=totalForGi[s]+frequencies[s][c]

// Compute the accumulative frequencies of response categories (fij) for
// each scale item. Let accumulativeFrequencies[i][j]=fij.
   for(s=1; s<=#OfScaleItem; s++)
      for(c=1; c<=#OfCategories; c++)
```

```
         if ( c is equal to 1) then
            accumulativeFrequencies[s][c]=frequencies[s][c]
         else
            accumulativeFrequencies[s][c]=accumulativeFrequencies[s][c-1]+
                                          frequencies[s][c]

// Compute the accumulative proportions for response categories (pij) for
// each scale item. Let accumulativProportions[i][j]=pij.
   for(c=1; c<=#OfCategories; c++)
      accumulativeProportions[s][c]=accumulativeFrequencies[s][c]/totalForGi[s]

// Compute the intervals corresponding to the non-extreme response categories.
   for (c=#OfCategories-1; c>1; c--)
      if (c is equal to #OfCategories-1) then
         interval[c]=F1(frequencies,accumulativeProportions,
                        totalFoNj,#OfScaleItems,#OfCategories,c)
      else
         interval[c]=F2(frequencies,accumulativeProportions,totalFoNj,
                        interval[c+1],#OfScaleItems, #OfCategories, c)

// Compute interval boundaries (xj) for non-extreme response categories.
   let X[0]=(minimum int value), X[1]=0, and X[#OfCategories]=(maximum int value)
   for (c=2; c<#OfCategories; c++)
      X[c]=X[c-1] + intervals[c]

// Compute scores for response categories.
   for (c=2; c<#OfCategories; c++)
     scores[c]=(X[c]+X[c-1])/2

   if the average of the observed proportions in the first category is
      less than 0.10 then
         core[1]=X[1]-1.0
   else
         score[1]=X[1]-1.1

   if the average of the observed proportions in the last category k is
      less than 0.10 then
         score[k]=X[k-1]+1.0
   else
       score[k]=X[k-1]+1.1
END

Function F1(frequencies, accumulativeProportions, totalFoNj, #OfScaleItems,
            #OfCategories, c)
BEGIN
   Return a value representing the interval corresponding to the last
   but one category using formula 4.
END

Function F2 (frequencies,accumulativeProportions,totalFoNj,
             interval[c+1],#OfScaleItems, #OfCategories, c)
```

```
BEGIN
   Return a value representing the interval corresponding to a category in between
   the second and the last but two categories using formula 5.
END
```

ACKNOWLEDGEMENTS. The author is obliged to Ms. Mei-Lan Liao for her assistance in administering the survey for the Melancholy study for this research.

References

[1] Agresti, A. (2002) *Categorical Data Analysis,* NY: Wiley-Interscience.

[2] Aho, A., J. Hopcroft and J. Ullman (1976) *The Design and Analysis of Computer Algorithms,* Bell Telephone Laboratories.

[3] Fielding, A. (1999) Why Use Arbitrary Points Scores?: Ordered Categories in Models of Educational Progress, *Journal of the Royal Statistical Society*, **162**(3), 303-328.

[4] Fisher, R. (1938) *Statistical Methods for Research Workers,* Edinburgh: Oliver and Boyd.

[5] Fleiss, J. (1981) *Statistical Methods for Rates and Proportions,* N.Y.: John Wiley and Sons.

[6] Li, C. (2002) The Study of Occupational Stress Reflecting on Recreational Participation in Taiwan. *NSC Report No: 90-2413-H-003-053,* National Science Council (Taiwan).

[7] Likert, R. (1932) *A Technique for the Measurement of Attitudes,* NY: Archives of Psychology.

[8] Montgomery, D. and G. Runger (2006) *Applied Statistics and Probability for Engineers,* NJ: Wiley, John and Sons.

[9] Neter, J., W. Wasserman and G. Whitmore (1992) *Applied Statistics,* London: Allyn and Bacon.

[10] Snell, E. (1964) A Scaling Procedure for Ordered Categorical Data, *Biometrics,* **20**(3), 592-607.

Received: June 13, 2007