A Fusion of Educational Research and Fuzzy Information: A Pragmatic Approach

Muhammad Naeem Butt
Institute of Education & Research
University of Peshawar, Peshawar, Pakistan
&
Muhammad Shafiq
Institute of Numerical Sciences
Kohat University of Science & Technology, Kohat, Pakistan

Abstract

Results and recommendations, based on data interpretation, in the research of social sciences are generated on precise measurements. However, advancement of numerical sciences showcases the unattainability of precise measurement of continuous phenomena due to the recent conception of fuzziness. Measurement results establish through educational research are obtained mostly by employing classroom experiments, survey-questionnaires and/or standardized-questionnaire; whereas, available literature also suggests the occurrence of fuzziness in classical measurement and questionnaire responses instead of preciseness. Therefore, in educational research it is recommended to employ latest measurement technique, i.e. fuzzy numbers instead of precise numbers, which will authenticate the reliability and appropriateness of obtained results.

Keywords: Education; Fuzzy Number; Non-Precise Data; Real Measurement Results

Introduction

Research in social sciences circumambient all the areas pertaining to qualitative and quantitative features. Data collection, analysis, and interpretation are central characteristics of research in social sciences. To deal with data in more systematic, effective, and efficient ways, various qualitative and quantitative tools have been developed, among which questionnaire is a prime tool; yet the issue of vagueness is unattended in it. Research in social sciences predominantly relies on questionnaires and interviews for data collection. These research instruments often encounter linguistic vagueness; hence, the issue is regarded as a challenging task for the social scientists to deal with. Certainly, most of the social sciences' variables are difficult to categorize precisely. For instance, if a question item solicits the honesty of a person; it is always difficult to obtain exact degree of the answer. It would either be in Yes/No or Agree/disagree and would not tell the precise level of agreement or disagreement of the respondent. The rationale here is to lesson, if not eliminate, the factor of fuzziness in questionnaire or any other research instrument. The more precise answer we obtain; the more accurate analysis will be generated, which would help in reducing the element of subjectivity in data collection and interpretation. It is a buzz and newly established phenomenon in research of social sciences to pursue enhanced and integrated research tools in addressing the problem of vagueness or fuzziness. This phenomenon also promotes the element of preciseness especially in the content of the questionnaires (Arfi, 2010).

Furthermore, in the real world of measurements one cannot get a precise measurement of irregular phenomenon like depth of a river because of its water level fluctuation. In the same way one cannot find a precise criterion between high or low temperature, good or poor student, effective or ineffective teacher etc. keeping in view the same Viertl (2006) concluded that there are two types of uncertainties in measurements: variation among the observations and imprecision of individual observations called *fuzziness*.

All classical models cover variation among the observations, and ignore imprecision of individual observations. According to Zadeh (1965) another method of modeling was necessary to consider the imprecision of a single measurement. To overcome this problem, the idea of fuzzy sets was first introduced by Zadeh in 1965. According to Viertl (2011) some preliminary concepts of fuzzy set theory are explained below

Fuzzy Number

Let f^* represents a *fuzzy number*, determined by the *characterizing function* denoted by
$\psi(\cdot)$, which is a real function of real variable satisfying the following conditions:

1. $0 \leq \psi(\cdot) \leq 1 \;\forall f \in \mathbb{R}$
2. $\psi(\cdot)$ has bounded support, i.e.
   $$\text{supp}[\xi(\cdot)] := \{ f \in \mathbb{R} : \xi(f) > 0\} \subseteq [a\,,b].$$
3. For all $\delta \in (0\,,1]$ the so-called δ-cuts, i.e. $C_\delta(f^*) := [\, f \in \mathbb{R} : \xi(f) \geq \delta]$ is a finite union of non-empty and compact intervals
   $$C_\delta(f^*) = \cup_{j=1}^{k_\delta}[a_{\delta,j}\,;\; b_{\delta,j}] \neq \emptyset \qquad \forall\, \delta \in [0\,,1].$$

Construction Lemma

Let $\mathcal{A}_\delta \;\cup_{j=1}^{k_\delta}[a_{\delta,j}\,;\; b_{\delta,j}]$ shortly denoted as $(\mathcal{A}_\delta;\, \delta \in (0\,,1])$ be a nested family of non-empty subsets of $\mathbb{R}$. Then the characterizing function of the generated fuzzy number is given by

$$\psi(f) = \sup\{\delta \cdot I_{\mathcal{A}_\delta}(f) \colon \delta \in (0\,,1]\} \qquad \forall\, \delta \in [0\,,1].$$

For details compare (Viertl and Hareter, 2006).

Fuzzy Vector

Let $\Psi(\cdot,\cdots,\cdot)$ represents *vector-characterizing function* of the k-dimensional fuzzy vector $\underline{f}^*$. Which is a real function of k real variables $f_1, f_2, \dots, f_k$ obeying the following conditions:

1. $\Psi(\cdot,\cdots,\cdot): \mathbb{R}^k \to [0,1]$
2. The support of $\Psi(\cdot,\cdots,\cdot)$ is a bounded set
3. Its δ-cuts $C_\delta\left(\underline{f}^*\right) := \left\{\underline{f} \in \mathbb{R}^k : \Psi\left(\underline{f}\right) \geq \delta\right\} \quad \forall\, \delta \in (0,1]$ is non-empty, bounded, and a finite union of simply connected sets.

Extension Principle

It's the generalized form of an arbitrary function $\mathcal{F}: \mathcal{N} \to \mathcal{M}$ for fuzzy argument value x^* in $\mathcal{N}$ with membership function $\eta: \mathcal{N} \to [0,1]$.

Then the fuzzy value $y^* = \mathcal{F}(x^*)$ is the fuzzy element in $\mathcal{M}$, and its membership function $\vartheta(\cdot)$ is defined by

$$\vartheta(y) = \begin{cases} \sup\{\eta(x): x \in \mathcal{N}, \mathcal{F}(x) = y\} & if\ \exists\, x: \mathcal{F}(x) = y \\ 0 & if\ \nexists x: \mathcal{F}(x) = y \end{cases}$$

$\forall\, y \in N.$

For details see (Klir and Yuan, 1995).

Regarding fuzziness in educational research only few references can be found like using fuzzy logic in educational measurement: The case of portfolio assessment by Fourali (1997) explained the idea of fuzzy logic in decision making through various examples. Furthermore, he illustrated his procedure through numerical examples. In a study by Cole and Persi-chitte (2000) generalized the idea of "Pressley and McCormick" and "Kosko" for the cognitive mapping through graphical representation. As learning is not a precise criterion, therefore; for the cognitive mapping instead of classical techniques fuzzy mapping gives more suitable results. According to Bassey (2001), some of the procedures of prediction in educational fields under fuzzy environment. He explained his ideas through the scientific and social experiments in such a way that scientific experiments give same result under same environment, but this cannot be hold in social behaviors. This study focuses on developing a framework or a strategy, which covers the element of fuzziness of the responses in a questionnaire/inventory or tool and makes the data more authentic and precise.

Descriptive Statistics and Fuzzy Data

The first step in educational research is to draw a histogram of the frequency distribution. For precise measurements it is very common and almost every

package has the option to draw it, but in case of fuzzy responses instead of classical histogram fuzzy histogram are more suitable.
Let $x_1^*, x_2^*, \dots, x_n^*$ be n fuzzy observations, for these observations there are n classes, i.e.
$K_j,\ j = 1(1)n$. Then the characterizing function of j^{th} class for the relative frequency distribution is obtained by the δ-cuts through mentioned construction lemma.

$$C_\delta\left(h_n^*\ (K_j)\right) = \left[\underline{h}_{n,\delta}\ (K_j),\ \ \overline{h}_{n,\delta}(K_j)\right] \quad \forall\ \delta \in (0\,,1] \qquad (1)$$

where$\underline{h}_{n,\delta}\ (K_j)$ represents lower end, while $\overline{h}_{n,\delta}(K_j)$ is the upper end of the corresponding
δ-cuts, which are obtained as:

$$\underline{h}_{n,\delta}\ (K_j) = \frac{\#\ \{x_i^* : C_\delta(x_i^*) \subseteq\ K_j\ \}}{n} \qquad (2)$$

and

$$\overline{h}_{n,\delta}(K_j) = \frac{\#\ \{x_i^* : C_\delta(x_i^*) \neq \emptyset\ \}}{n} \qquad (3)$$

See (Viertl, 2011).

After collection of data for the statistical inference descriptive statistics has prime importance.

For fuzzy observation $x_1^*, x_2^*, \dots, x_n^*$ arithmetic mean can be defined as:
Let $C_\delta\left(x_i^*\ (K_j)\right) = \left[\underline{x}_{i,\delta}\ ,\ \ \overline{x}_{i,\delta}\right] \quad \forall\ \delta \in (0\,,1],\quad i = 1(1)n$, representing the δ-cuts then the corresponding arithmetic mean can be obtained in the following way

$$\overline{x^*} = \frac{1}{n}[x_1^* +\ x_2^* +\ \dots +\ x_n^*]$$

$$\overline{x^*} = \frac{1}{n}\sum_{i=1}^{n} x_i^* \qquad (4)$$

Simply, lower and upper level δ-cuts can be obtained as,

$$C_\delta\left(\overline{x^*}\right) = \left[\frac{1}{n}\sum_{i=1}^{n}\underline{x}_{i,\delta}\ ,\ \ \frac{1}{n}\sum_{i=1}^{n}\overline{x}_{i,\delta}\ \right] \quad \forall\ \delta \in (0\,,1]$$

and characterizing function is obtained through construction lemma.
For skewed or ordinal data median is the appropriate measure of central tendency, according to (De S´aa et al., 2015) for fuzzy data median ($Med(\ \widetilde{x^*}\ )$) based on the δ-cuts can be obtained in the following way

$$C_\delta\left(Med(\ \widetilde{x^*}\ )\right) = \left[\ Med\left(\underline{\tilde{x}}\right),\quad Med\left(\overline{\tilde{x}}\right)\ \right] \quad \forall\ \delta \in (0\,,1] \qquad (5)$$

and characterizing function of the fuzzy estimate of median can be obtained through the mentioned construction lemma. It is obvious that only central tendency cannot represent the data very well so in addition to this variation of the data is also necessary for data representation. Viertl (2014) states that based on fuzzy observations; fuzzy estimates of the sample standard deviation are denoted by s^* and is given as

$$C_\delta(s^*) = \left[\min_{x_1,x_2,\dots,x_n \in C_\delta[x_1^*] X C_\delta[x_2^*] X \dots X C_\delta[x_n^*]} s \;,\; \max_{x_1,x_2,\dots,x_n \in C_\delta[x_1^*] X C_\delta[x_2^*] X \dots X C_\delta[x_n^*]} s\right] \quad \forall\, \delta \in (0\,,1] \qquad (6)$$

Where

$$\underline{s}_\delta = \min_{x_1,x_2,\dots,x_n \in C_\delta[x_1^*] X C_\delta[x_2^*] X \dots X C_\delta[x_n^*]} s$$

and

$$\overline{s}_\delta = \max_{x_1,x_2,\dots,x_n \in C_\delta[x_1^*] X C_\delta[x_2^*] X \dots X C_\delta[x_n^*]} s$$

Inferential Statistics and Fuzzy Data

Estimation and testing of hypothesis is core of the statistical procedures. The most common techniques which are generalized for fuzzy data are presented as: Wu (2009) states that based on fuzzy measurements, confidence interval estimation for the fuzzy data is presented.

The lower and upper ends can be defined as:

$$\left[\overline{\underline{x}_{i,\delta}} - z\alpha/_2 \frac{\underline{s}_\delta}{\sqrt{n}} \;,\; \overline{\underline{x}_{i,\delta}} + z\alpha/_2 \frac{\underline{s}_\delta}{\sqrt{n}}\right] = 100(1-\alpha) \quad \forall\, \delta \in (0\,,1]$$

and

$$\left[\overline{\overline{x}_{i,\delta}} - z\alpha/_2 \frac{\overline{s}_\delta}{\sqrt{n}} \;,\; \overline{\overline{x}_{i,\delta}} + z\alpha/_2 \frac{\overline{s}_\delta}{\sqrt{n}}\right] = 100(1-\alpha) \quad \forall\, \delta \in (0\,,1]$$

According to Filzmoser and Viertl (2004) for fuzzy observations the fuzzy test statistics and fuzzy p-value are generalized in the following way:

Let $t^* = \mathcal{G}\,(x_1^*, x_2^*, \dots, x_n^*)$ is denoting fuzzy test statistics having characterizing function $\psi(t)$, then their corresponding δ-cuts are defined as:

$$C_\delta(t^*) = [t_{1,\delta}\,,\; t_{2,\delta}] \quad \forall\, \delta \in (0\,,1]$$

For a precise p-value the decision based on fuzzy test statistic is given as

(a) $p = P\,[T \le t = \max(supp\psi(\cdot))\,]$ (b) $p = P\,[T \ge t = \min(supp\psi(\cdot))\,]$

Using δ-cuts of the fuzzy test statistics t^* the corresponding δ-cuts of the fuzzy p-value (p^*) is obtained in the following way

$$C_\delta(p^*) = [P(T \le t_{1,\delta})\,, \quad P(T \le t_{2,\delta})] \quad \forall\, \delta \in (0\,,1]$$

Or

$$C_\delta(p^*) = [P(T \ge t_{2,\delta})\,, \quad P(T \ge t_{1,\delta})] \quad \forall\, \delta \in (0\,,1].$$

Comparison of two or more than two population parameters are one of the prominent aspects of educational research. For this purpose, one cannot use pairwise comparison of the populations means. According to Wu (2007). The best suited technique is analysis of variance (ANOVA), for fuzzy data analysis of variance (ANOVA).

Let we have T_{ij}^* the observation on i^{th} treatment and j^{th} replicate, where $i = 1(1)r, j = 1(1)n_i$, and $n = \sum_{j=1}^{n_i}$, then the corresponding ANOVA model for fuzzy observations can be written as

$$T_{ij}^* = \mu_i^* + \epsilon_i^*$$

This model can be simply written as
Total Sum of Squares = Sum of Squares of Treatment + Sum of Squares of Error

$$SSTot = SSTr + SSE$$

Consider the hypothesis that

$$H_0^*: \mu_1^* = \mu_2^* = \cdots = \mu_r^*$$
$$H_1^*: Not\ all\ means\ are\ equal$$

The corresponding lower and upper δ-cuts of the fuzzy observations T_{ij}^* are denoted by $T_{ij\delta}^L$ and $T_{ij\delta}^U$ respectively.
Total of the i^{th} treatment and grand total is obtained in the following way

$$T_{i.}^* = \bigoplus_{j=1}^{n_i} T_{ij}^* \text{ and } T_{..}^* = \bigoplus_{i=1}^{r} \bigoplus_{j=1}^{n_i} T_{ij}^*$$

Similarly, lower and upper δ-cuts of the i^{th}treatment and grand total is obtained as:

$$T_{i.\delta}^L = \sum_{j=1}^{n_i} T_{ij\delta}^L \text{ and } T_{i.\delta}^U = \sum_{j=1}^{n_i} T_{ij\delta}^U \qquad \forall\ \delta \in (0\,,1]$$
$$T_{..\delta}^L = \sum_{i=1}^{r}.\ \sum_{j=1}^{n_i} T_{ij.\delta}^L \qquad \text{and} \qquad T_{..\delta}^U = \sum_{i=1}^{r}.\ \sum_{j=1}^{n_i} T_{ij.\delta}^U \qquad \forall\ \delta \in (0\,,1]$$

The lower and upper ends of δ-cuts of corresponding sum of squares are obtained in the following way

$$SSTot_\delta^L = \sum_{i=1}^{r}.\ \sum_{j=1}^{n_i} \left[T_{ij\delta}^L\right]^2 - \frac{[T_{..\delta}^L]^2}{n}$$
$$SSTot_\delta^U = \sum_{i=1}^{r}.\ \sum_{j=1}^{n_i} \left[T_{ij\delta}^U\right]^2 - \frac{[T_{..\delta}^U]^2}{n}$$
$$SSTr_\delta^L = \sum_{i=1}^{r} \frac{[T_{i.\delta}^L]^2}{n_i} - \frac{[T_{..\delta}^L]^2}{n}$$
$$SSTr_\delta^U = \sum_{i=1}^{r} \frac{[T_{i.\delta}^U]^2}{n_i} - \frac{[T_{..\delta}^U]^2}{n}$$

And

$$SSE_{\delta}^{L} = \sum_{i=1}^{r} . \sum_{j=1}^{n_i} \left[T_{ij\delta}^{L}\right]^2 - \sum_{i=1}^{r} \frac{[T_{i.\delta}^{L}]^2}{n_i}$$
$$SSE_{\delta}^{U} = \sum_{i=1}^{r} . \sum_{j=1}^{n_i} \left[T_{ij\delta}^{U}\right]^2 - \sum_{i=1}^{r} \frac{[T_{i.\delta}^{U}]^2}{n_i}$$

Now the relation can be written as

$$SSTot^{L} = SSTr^{L} + SSE^{L}$$
$$SSTot^{U} = SSTr^{U} + SSE^{U}$$

Now the mean square is obtained as

$$MSTr_{\delta}^{L} = \frac{SSTr_{\delta}^{L}}{r-1} \text{ and } MSTr_{\delta}^{U} = \frac{SSTr_{\delta}^{U}}{r-1} \qquad \forall\, \delta \in (0\,,1]$$

And

$$MSEr_{\delta}^{L} = \frac{SSEr_{\delta}^{L}}{n-r} \text{ and } MSEr_{\delta}^{U} = \frac{SSEr_{\delta}^{U}}{n-r} \qquad \forall\, \delta \in (0\,,1]$$

Now the required lower and upper ends of the δ-cuts of F-test are obtained as

$$F_{\delta}^{L} = \frac{MSTr_{\delta}^{L}}{MSEr_{\delta}^{L}} \text{ and } F_{\delta}^{U} = \frac{MSTr_{\delta}^{U}}{MSEr_{\delta}^{U}} \qquad \forall\, \delta \in (0\,,1]$$

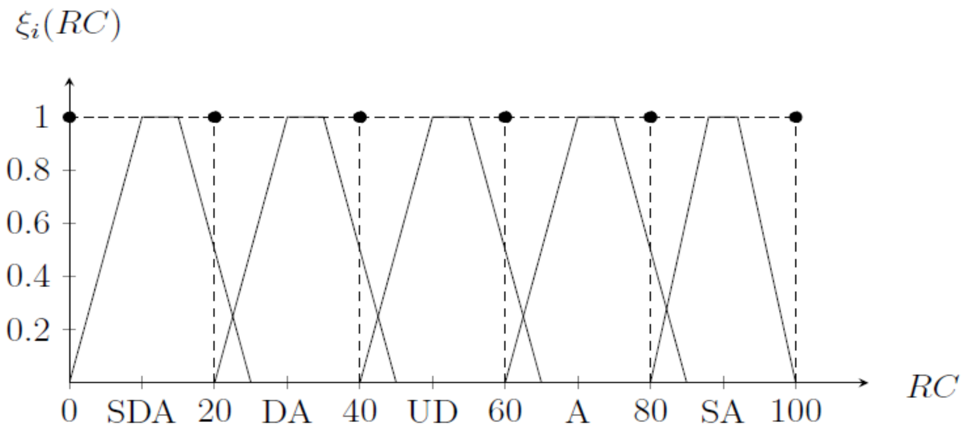
From the above equations corresponding lower and upper ends of the generating family of intervals, and the characterizing function is obtained through construction lemma. Furthermore, to quantify the associated factors of the dependent variable multiple regression model is a suitable choice. For this purpose, Bargiela et al., (2007) is a significant contribution for fuzzy data.

For Ordinal Responses

Research in social sciences is as indispensable as in pure sciences. It is normally conducted in social setting and undertake the areas of human constructs and traits like behavior, attitude, feelings, emotions, IQ, EQ etc. Standardized questionnaires/inventory/scale by Baron, Hamilton, Beck and way more are frequently used in the research of social sciences.

Methodology, in research, plays a pivotal role in data collection, which is usually based on interviews and/or questionnaires. Mostly closed-ended items are used, which could either be standardized or self-developed questionnaires. Sometimes a questionnaire is slightly modified in the light of standardized questionnaire/scale/inventory. Standardization is a procedure of designing and employing procedural and technical criteria or benchmark and aids to enhance reliability, validity, and excellence. The standardized research instrument for data collection normally solicits the respondents on 3, 5, 7, 9, or 11-point rating scale responses. For instant, in 5-point rating scale (SDA, DA, UD, A, SA) a respondent has the provision to opt for one of the response options, which reflects his/her level of confidence in that specific question item, e.g. if a respondent selects strongly disagree (SDA) for a certain item that carries 1-20 % of the total weightage then determining the level of confidence of the respondent for that specific item will be difficult as if its 1% or 20 % or in between the two limits. This reflects the fuzziness of the response, which affects the results of the collected data.

Given below in Figure-1 in which solid lines show the characterizing functions of fuzzy responses whereas, the dashed lines show the precise responses.



In the available literature researchers (De S´aa et al., 2015), (Lubiano et al., 2016a), (Lubiano et al., 2016b) suggest analyzing the data obtained on fuzzy rating scales, which are more appropriate to the realistic nature of data.

Conclusions

Recent advancement in numerical and measurement sciences has witnessed the unattainability of precise numbers on continuous scale due to the notion of fuzziness. Similarly, the responses on questionnaires (Agree or Disagree etc.) are always of linguistic nature and cannot be precisely measured. Therefore, use of fuzzy numbers in the research of social sciences would subside estimation error and establish more reliable and appropriate results. Rating scale techniques for data collection need to be modified on fuzzy scales and the analysis techniques developed for fuzzy data is highly recommended to be used to cover both types of variations: fuzziness and stochastic.

References

Arfi, B. (2010). *Linguistic Fuzzy Logic Methods in Social Sciences*. Springer. Berlin Heidelberg.

Bargiela, A., Pedrycz, W., & Nakashima, T. (2007). "Multiple regression with fuzzy data." *Fuzzy Sets and Systems*. 158 (19). Pp. 2169 -2188.

Bassey, M. (2001). "A Solution to the Problem of Generalization in Educational Research: Fuzzy Prediction." *Oxford Review of Education*. 27 (1). Pp. 5-22.

Cole, J. R. & Persichitte, K. A. (2000). "Fuzzy cognitive mapping: Applications in education." *International Journal of Intelligent Systems*. 15 (1). Pp. 1-25.

De S´aa, S. d. l. R., Gil, M. ´A., Gonz´alez-Rodr´ıguez, G., L´opez, M. T., & Lubiano, M. A. (2015). "Fuzzy Rating Scale-based Questionnaires their Statistical Analysis." *IEEE Transactions on Fuzzy Systems*. 23 (1). Pp. 111-126.

Filzmoser, P. & Viertl, R. (2004). "Testing Hypotheses with Fuzzy Data: The Fuzzy p-value." *Metrika*. 59 (1). Pp. 21-29.

Fourali, C. (1997). "Using Fuzzy Logic in Educational Measurement: The Case of Portfolio Assessment." *Evaluation & Research in Education*.11 (3). Pp. 129-148.

Klir, G. J. & Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice-Hall, New Jersey.

Lubiano, M. A., de la Rosa de Sa, S., Montenegro, M., Sinova, B., & ngeles Gil, M. (2016a). "Descriptive Analysis of Responses to Items in Question-naires. Why not Using a Fuzzy Rating Scale?" *Information Sciences*. 360. Pp. 131-148.

Lubiano, M. A., Montenegro, M., Sinova, B., de la Rosa de Sa, S., & Angeles Gil, M. (2016b). "Hypothesis Testing for Means in Connection with Fuzzy Rating Scale-based Data: Algorithms and Applications." *European Journal of Operational Research*. 251 (3). Pp. 918-929.

Viertl, R. (2006). "Univariate Statistical Analysis with Fuzzy Data." *Computational Statistics & Data Analysis*. 51 (1). Pp. 133-147.

-------. (2011). *Statistical Methods for Fuzzy Data*. Wiley, Chichester.

-------. (2014). "Measurement of Continuous Quantities and their statistical evaluation." *Austrian Journal of Statistics*. 44 (1). Pp. 25-32.

Viertl, R. & Hareter, D. (2006). *Beschreibung und Analyse Unscharfer Information: Statistische Methoden f¨ur Unscharfe Daten*. Springer, Wien.

Wu, H.-C. (2007). "Analysis of Variance for Fuzzy Data." *International Journal of Systems Science*. 38 (3). Pp. 235-246.

-------. (2009). "Statistical Confidence Intervals for Fuzzy Data." *Expert Systems with Applications*. 36 (2). Pp. 2670-2676.

Zadeh, L. A. (1965). "Fuzzy Sets." *Information and Control*. 8 (3). Pp. 338-353.