Predicting Purchase Quantities: Further Investigation of the Juster Scale

Peter Seymour, Mike Brennan and Don Esslemont

The Juster Scale is an eleven-point purchase probability scale that has been found to be more accurate than purchase intention scales for predicting rates of purchase of a variety of consumer goods and services. However, in addition to knowing the rate of purchase, that is, the number of people who will purchase, it is often necessary to know purchase levels, that is, the quantity of goods that will be purchased. To date, two studies have examined the effectiveness of three methods of applying the Juster Scale to estimate purchase levels, with promising results. However, the accuracy of predictions varied across methods and products, and there was evidence that some respondents were confused by the procedures. This paper presents the findings of a further investigation of three methods of using the Juster Scale to estimate purchase levels. The study examined the impact of age and educational level on respondent understanding and predictive accuracy. Recommendations are made regarding possible applications and limitations of the methods.

Keywords: Juster Scale, purchase quantities, probabilities, constant sum, multiple question, intentions

Introduction

Juster's purchase probability scale (see Figure 1) has been administered in several different ways to predict, quite successfully, the purchase rates of both durable and non-durable goods (Juster 1966; Clancy & Garsen 1970; Gabor & Granger 1972; Gan, Esslemont & Gendall 1986; Gendall, Esslemont & Day 1988; Day, Gan, Gendall & Esslemont 1991;). These studies all predicted the proportion of individuals, or households, that were likely to purchase a product over a certain period of time; that is, they estimated the purchase *rate*.

However, for fast moving consumer items, such as toothpaste, eggs and butter, a household is likely to buy several units of the product even over a relatively short period of time. Therefore, simply estimating the percentage of households that will buy a frequently purchased item is of limited value to practitioners. As Day *et al*. (1991) noted, what is required is an estimate of the total purchase *level* over that time period. To date, only two studies have addressed this issue; the 1989 study by Hamilton-Gibbs and the 1991 study by U. Both of these studies used two different methods of administering the Juster scale to forecast the purchase rates of fast-moving consumer goods. These methods were labelled, by U (1991), the Multiple Question Method and the Constant Sum Method.

In the Multiple Question Method, respondents are required to estimate the probability, using the Juster scale, of buying zero units of an item, then one unit, two units and so on until a zero probability value is reached. Purchase levels are then calculated by, first, multiplying the purchase amount by the probability that the respondent will buy that amount, then adding these numbers together to give the predicted purchase level for each respondent.

Figure 1. The Juster Purchase Probability Scale

10	Certain, practically certain	(99 in 100)
9	Almost sure	(9 in 10)
8	Very probable	(8 in 10)
7	Probable	(7 in 10)
6	Good possibility	(6 in 10)
5	Fairly good possibility	(5 in 10)
4	Fair possibility	(4 in 10)
3	Some possibility	(3 in 10)
2	Slight possibility	(2 in 10)
1	Very slight possibility	(1 in 10)
0	No chance, almost no chance	(1 in 100)

In the Constant Sum Method, respondents are provided with 10 tokens and a grid, printed on a card on which the tokens can be arranged. The rows of the grid represent different quantities or numbers of items, from 0 to 12. Respondents are required to distribute ten tokens, each representing a probability of 0.1, to show the probability of buying different quantities of an item. For example, if there is an equal probability (0.5) that a respondent will buy three or four units of the product, they would place five counters in the row representing three items, and five counters in the row representing four items. Purchase levels are then calculated in a similar way as for the Multiple Question Method.

Hamilton-Gibbs (1989) attempted to predict the purchase rates of eight frequently purchased items. These were toothpaste, margarine, butter, eggs, spaghetti, chicken, ice cream, and cheese. He found that the Constant Sum Method overestimated purchases whereas the Multiple Question Method underestimated purchases for all of the items tested. However, both methods were accurate predictors of the actual purchase levels for seven of the eight items tested. Hamilton-Gibbs' results also showed the Constant Sum Method to be a superior predictor of consumer purchase behaviour in all but two of the cases. Although the accuracy of the Constant Sum Method was not very much greater than that of the Multiple Question Method, Hamilton-Gibbs concluded that, because of the more consistent predictions, the Constant Sum Method was superior for predicting the purchase rates of fast moving consumer goods. Hamilton-Gibbs also concluded that the Constant Sum Method was easier and less tedious for respondents and interviewers to utilise.

U's study differed markedly from that of Hamilton-Gibbs. While Hamilton-Gibbs tested the accuracy of predictions for product classes, U attempted to predict the purchases of product brands; Coca-Cola, Campbell's Red and White Label canned soup, and a new product, Tasti Fruit Splits.

U's results were less consistent than those of Hamilton-Gibbs. For Coca Cola, purchases were slightly underestimated by the Multiple Question Method, and slightly overestimated by the Constant Sum Method. In contrast, both methods produced a very large overestimation of purchases for Campbell's soup. Estimates could not be produced for the Tasti Fruit Splits, as the product was not launched onto the market in time.

The studies of both U (1991) and Hamilton-Gibbs (1989) suggest that the Constant Sum Method produces more accurate results, on average, than the Multiple Question Method, although the accuracy of the two methods is quite similar. However, different conclusions were drawn regarding the ease of use of the methods. Hamilton-Gibbs suggested that the Constant Sum Method was the easier method to use and understand for both respondents and interviewers. U, on the other hand, found that both respondents and interviewers had difficulty with the Constant Sum Method, and suggested it required further development.

Several possible reasons can be proposed to account for these differing conclusions regarding the ease of use of the two methods. Firstly, U's study involved branded products, including a new product. This may have made the respondents' task more difficult, and contributed to a lower understanding of the more difficult concepts in the Constant Sum Method. Secondly, Hamilton-Gibbs and U used slightly different versions of the Constant Sum Method. In Hamilton-Gibbs study, respondents were required to stack the tokens, whereas in U's study the tokens were laid flat. Thirdly, Hamilton-Gibbs administered the questionnaires himself, whereas U's interviewing was undertaken by minimally trained students. Interviewer training and interviewer understanding are often very important to obtaining good results, particularly when complex concepts are involved in the study. As a consequence, the limited training of interviewers in U's study may well have contributed to the problems encountered with the Constant Sum Method.

A further possible reason for the anomalies could be differences in the age and educational levels of respondents in the two studies. Interviewers in U's study noted difficulties with the understanding of the Constant Sum Method among the less well educated and among the older respondents. This raises the possibility that different methods of presenting the Juster Scale may be required for different sample segments. Clearly, there are a number of important issues to resolve before the Juster Scale can be used with confidence to make predictions of purchase levels.

The purpose of the present study was two-fold. Firstly, it attempted to develop procedures for employing the Juster scale that would maximise the ease of use and understanding of the scale for both respondents and interviewers. Specifically, the study examined the effect of respondents' age and education on the understanding and interpretation of three methods of employing the Juster scale. Secondly, the study sought verification of the findings of Hamilton-Gibbs (1989) and U (1991) regarding the relative accuracy of the three methods of using the Juster Scale to predict purchase levels.

Method

Sample

Since one of the objectives of the study was to assess the effect that age and level of education had on respondents' ability to understand the three methods of presenting the Juster Scale, and on the accuracy of predictions, an equal number of respondents was obtained with different age and education levels. Respondents were categorized into three age groups (see Table 1), and two educational groups: those with less than four years secondary school education were described as having a "low" education level; those with four or more years of secondary schooling were described as having a "high" education level.

To achieve a balanced design, quota samples were taken from selected areas in Palmerston North. Areas were chosen that had a particularly high number of residents from either the high or low end of the socio-economic scale. It was assumed that no correlation would exist between a respondent's propensity to be at home and the effect that the respondent's age and education would have on understanding and predictive accuracy. For this reason, and because the study aimed to compare categories rather than generalise to the overall population, no callbacks were made. Houses where the main grocery shopper was not at home were simply disregarded, and an interview was sought with the main grocery shopper from the next house on the right.

Procedure

Ninety face-to-face interviews were conducted with main grocery shoppers in selected households in Palmerston North, between July 11 and August 12, 1993. These 90 respondents were assigned to one of the three methods of applying the Juster Scale, balanced by age and educational level (see Table 1).

Table 1. Sample characteristics

		Age		
Method	**Education**	**<30**	**30-60**	**>60**
Multiple Question Method	low education	Group 1	Group 2	Group 3
	high education	Group 4	Group 5	Group 6
Flat Sum Method	low education	Group 7	Group 8	Group 9
	high education	Group 10	Group 11	Group 12
Stacked Sum Method	low education	Group 13	Group 14	Group 15
	high education	Group 16	Group 17	Group 18

Note: For each group, n = 5.

For six frequently purchased grocery items (toothpaste, butter, margarine, eggs, spaghetti, and ice-cream), respondents were asked to indicate the product size most commonly bought, the probability that they would buy any of that product in the next four weeks, the most likely quantity of each item they would buy during that time period, and the purchase probabilities (from the Juster Scale) associated with various quantities of each product, using one of the three methods under consideration. Respondents were also asked whether they would agree to be reinterviewed at a later date. Interviews took around fifteen minutes to complete when using the Multiple Question Method, and around twelve minutes when either of the two Constant Sum Methods (Stacked Sum or Flat Sum) was employed.

Twenty-eight days after the initial interview, respondents were recontacted and the actual purchase amounts were obtained for the six grocery items. If respondents could not be contacted on the appropriate day, they were telephoned every following day until contact was made. If respondents did not have a telephone but agreed to further questions, face-to-face interviews were conducted to obtain the required information. On average, the telephone interviews took about four minutes to administer. Of the original 90 respondents, 84 were successfully re-interviewed by telephone, 28 for each method. Thus the response rate for each of the three methods was 93.3 %, which was also the overall response rate (see Table 2).

Table 2. Response rates

	Multiple Question Method	**Flat Sum Method**	**Stacked Sum Method**
Interviewed	30	30	30
Reinterviewed	28	28	28
Refusals	1	1	2
Wrong numbers	-	1	-
No telephone [1]	1	-	-
Response rate	28/30 = 93.3%	28/30 = 93.3%	28/30 = 93.3%
Total response rate	84/90 = 93.3%		

Note: 1. Respondents who had no phone and did not provide an address could not be contacted.

The aim of the study was not only to collect the data to evaluate the accuracy of predictions using the three methods, but also to evaluate respondents' ease of use and understanding of the methods used to obtain the purchase prediction data. To accomplish this, the Belson double-back method (Belson, 1986) was employed after every interview to help assess respondents' understanding. That is, respondents were asked "In your own words, what did you think I meant when I asked....?". Respondents' level of understanding was rated on a seven-point scale on the basis of such aspects as how well the questions were understood, whether respondents grasped the relationship between the probability questions, and how much explanation was required.

Instruments

Three methods for presenting the Juster Scale were employed in this study: the Multiple Question Method and two forms of the Constant Sum Method: that used by Hamilton-Gibbs

(1989), renamed the Stacked Sum Method, and that used by U (1991), renamed the Flat Sum Method.

Multiple Question Method

Respondents were asked a preliminary question:

> *"How many tubes of toothpaste are you most likely to buy over the next four weeks? "*

This number was labelled "n". The interviewer then asked the respective probabilities of buying n, n-1, n-2, and so on until a zero purchase probability was obtained. The interviewer then asked the probability that the respondent would buy n+1, n+2, and so on, continuing until a zero purchase probability was found, for example:

> *"Taking everything into account, what are the chances of you personally buying <n> tubes of toothpaste in the next four weeks? What are the chances that you will buy <n-1>? What about <n-2>? <n-3> ... <n+1>, <n+2>...?".*

It was hoped that asking the preliminary question, to determine the most likely number of purchases, would help respondents think about their most probable purchase quantities.

This approach differed from that used previously by both Hamilton Gibbs and U, where respondents were asked:

> *"Taking everything into account, what are the chances of you personally buying no tubes of toothpaste in the next four weeks? What are the chances that you will buy 1? What about 2? 3?.......".*

A major problem with this earlier form of the Multiple Question Method was that it became very laborious if the likely purchase quantity was high. As the probability of purchase was only asked for purchase quantities close to the most likely quantity of each item, the modified approach using the preliminary question facilitated the reduction of redundant questions.

Constant Sum Methods

Respondents were required to place the ten counters (each representing a purchase probability of 0.1) onto a grid, to show the probability of buying different numbers of items. Thus, if a respondent was equally likely to buy two or three items in the given period, they would arrange the tokens as follows. For the Flat Sum Method, the respondent would place five counters in the row representing two items, and place five counters in the row representing three items. For the Stacked Sum Method, they would stack five counters in the cell representing two items, and stack five counters in the cell representing three items.

As with the Multiple Question Method, respondents of both Constant Sum Methods were asked a preliminary question to identify the most likely number of items to be bought in the next four weeks. This was, again, in attempt to compel respondents to concentrate their answers around the quantities of purchase that were most probable.

Results

Effect of Age, Education and Method of Presentation on Respondent Understanding
At the conclusion of each interview, the respondent's level of understanding was assessed. A seven point scale was utilised for this assessment, with "1" signifying an excellent understanding level and "7" a poor one. The mean level of understanding was calculated for each group (see Table 3), and the significance of the effects on understanding of respondent age, education level and method of presentation tested using analysis of variance (see Table 4).

Table 3. Mean level of understanding for age/education groups

		Age			Mean level of Understanding for Method
Method	**Education**	**Under 30**	**31 to 60**	**Over 60**	
Multiple Question Method	low	1.8	2.4	2.4	2.7
	high	4.0	3.2	2.4	
Flat Sum Method	low	2.8	2.6	1.4	3.3
	high	5.0	4.0	3.8	
Stacked Sum Method	low	2.6	2.4	3.2	3.3
	high	3.8	4.4	3.4	

Notes: 1. Education level

The mean ratings of understanding for the 18 groups ranged widely from 1.4 to 5.0, while the means for the three methods were ranged from 2.7 to 3.3, suggesting that, overall, the two Constant Sum Methods were better understood than the Multiple Question Method (see Table 3).

The analysis of variance revealed that the ratings of understanding of respondents differed significantly ($p < 0.01$) between the 18 groups shown in Table 3. Although respondents' age, education and the method all affected the level of understanding, education accounted for the largest variance ($p<.01$) in the understanding of respondents (see Table 4). That is, the more highly educated respondents consistently understood the questionnaire, for each of the three methods of presentation, better than their less educated counterparts, particularly among the younger age groups.

Table 4. Effect of method used and respondents' age and education level on respondent understanding

Source of Variation	Sum of Squares	Degrees Freedom	Mean Square	F- Ratio	Probability of F
Main effects	58.645	7	8.378	5.487	.000
Method	7.229	2	3.615	2.367	.103
Agegroup	7.030	2	3.515	2.302	.109
Education	46.023	3	15.341	10.048	.000
2-way Interactions	19.683	16	1.230	.806	.674
Method/ Agegroup	3.765	4	.941	.617	.652
Method/ Education	10.898	6	1.816	1.190	.324
Agegroup/ Education	5.598	6	.933	.611	.720
Explained	91.206	30	3.040	1.991	.012
Residual	90.083	59	1.527		
Total	181.289	89	2.037		

It is important to note that the ratings of respondent understanding are based on the interviewer's subjective judgements. It was difficult to judge the understanding of many respondents, particularly the elderly, as they claimed they were sure of how many items they would buy, and consistently gave purchase probabilities of ten. Whether these high probability responses were due to a lack of understanding of the questionnaire or whether respondents actually were certain of their purchases is hard to judge at the initial interview stage. Furthermore, the interviewer reported that, although many of the older respondents did understand the questionnaire quite well overall, they typically required more explanation than the younger respondents, particularly with the Constant Sum methods. The older respondents also had a tendency to dislike the counters and, in the Belson's pretest, one third of the older respondents actually stated they would prefer questions such as those used in the Multiple Question Method.

Respondent Difficulties in Understanding

Several of the difficulties respondents had with understanding the questionnaires were quite minor and were able to be largely solved through small changes to the questionnaire. Often the solution simply involved stressing certain words or providing a fuller explanation. For example, in the Multiple Question Method, several respondents initially confused the chances that they would buy exactly "n" items with the chance that they would buy at least "n" items. By stressing "exactly", this problem was greatly lessened.

A number of respondents, particularly the elderly, initially had trouble grasping the counter concepts in the Constant Sum Methods. To help lessen this problem a second example was given to those who did not appear to understand the process. This did not totally solve the problem, but vastly improved respondents' level of understanding in many cases. Those who still did not totally understand were prompted as much as possible without biasing the results.

Another problem with the Constant Sum Methods was that some respondents thought that the counters represented quantities rather than probabilities. That is, they would place two counters beside/on the quantity two, three beside/on the three and so on. This problem was solved by more thoroughly explaining what the numbers on the grids represented and prompting respondents where necessary.

However, not all problems were so easily solved. Very few respondents realised that the probability of buying any items should relate inversely to the probability that they would buy zero items. On the occasions that the two probabilities did correspond correctly, it was often due to chance rather than a complete understanding of the concepts. Furthermore, few respondents who were administered the Multiple Question Method realised that the probabilities should add to one. This suggests that respondents do not usually think in probabilities out of ten. In fact, one of the modifications made to the questionnaires was to change the wording of the examples to stress the verbal probabilities of the Juster Scale rather than the numeric ones. This was done in order to assist respondent understanding, as most respondents related better to the verbal probability explanations. This was exemplified by the fact that many respondents would answer the questions with the written explanations rather than the numeric ones as required. That is, they would say "there is a fairly good possibility that I will buy five tubes of toothpaste" rather than there is a five in ten chance that I will buy.

Predictive Accuracy of Methods

The predictive accuracy of the three methods was examined by comparing their predictive error, that is, their absolute error as a percentage of actual purchases (Gendall et al., 1988; Hamilton-Gibbs, Esslemont & McGuinness, 1992; Brennan, Esslemont & U, 1993).

In the following analyses, both weighted and unweighted data are used. One of the consequences of using the Multiple Question Method, noted by U and evident in this study, is that the sum of the probabilities obtained by the Multiple Question Method typically sum to more that 1. While they are forced to sum to 1 with the Constant Sum Method, since 10 counters are used, with both methods the probabilities should really sum to the probability of buying any product at all. That is, if a respondent states that there is a .7 probability of buying any toothpaste, the sum of the probabilities of buying different quantities should sum to .7. If they do not, then the data will provide over-estimates of the respondents' purchases. To deal with this situation, the purchase level estimates were weighted by the probability of buying any of the product.

Overall Accuracy of Unweighted and Weighted Purchase Predictions

The main finding of this study was that the predictions of purchases were more accurate when weighted to take into account the chance of the respondents buying any of the product at all. However, regardless of whether weighted or unweighted data was used, the predictive accuracy of the three different methods of applying the Juster Scale was quite similar, and depended largely on the product (see Figures 2 and 3). The predicted and actual quantities purchased, and the weighted and unweighted predictive errors, are reported in Appendices A and B.

In order to assess the overall accuracy of the three methods, the mean absolute predictive error across all six products was calculated (see the rightmost column in Figures 2 and 3). For both the weighted and particularly the unweighted data, the overall mean predictive errors were similar for each of the three methods. With the unweighted figures, the Multiple Question Method had the smallest mean error of 25.5% while the Stacked Sum and Flat Sum Methods had slightly larger mean errors of 26% and 27%, respectively. When the weighted figures were used mean predictive errors were much smaller. The two Constant Sum Methods had smaller mean predictive errors, of 11.5% for the Flat Sum and 14% for the Stacked Sum, than the mean error of 16% for the Multiple Question Method. However, the differences were very slight, particularly for the unweighted data, and not significant at the ten percent level.

It should be noted that while the adjusted figures increase predictive accuracy in most cases, this trend does not always hold. By adjusting the data, the overestimates are lessened but the underestimates become greater. Therefore, unless all the predictions for a certain method are overpredictions or underpredictions, adjusting the data will have inconsistent effects on predictive accuracy across products. Nevertheless, the adjusted predictive errors are, on average, much smaller. For this reason, the adjusted figures will be used in the following analyses.

Accuracy of Methods for Each Product

As Figures 2 and 3 show, the purchase prediction errors differed widely across products. The Multiple Question Method was most accurate for one product, the Stacked Sum Method was most accurate for two products while the Flat Sum Method was the most accurate for three products.

The variation in predictive accuracy of the three methods was quite large. For the Multiple Question Method, the most accurate prediction was an underestimate of 6% for butter, while the largest error was a 56% overprediction for spaghetti purchases. While the Flat Sum Method was the most accurate overall, the predictive errors ranged from a 1% overestimate of butter to a large 21% underestimate of spaghetti purchases. The Stacked Sum Method consistently *over-estimated* purchases by between 2% and 19%. The only exception to this rule was the *under-estimate* of ice-cream of 24%.

With the adjusted figures, toothpaste, which originally had the greatest predictive error, is the most accurately predicted item, on average. When the adjusted figures are used, all the other products' mean predictive error decreases, apart from margarine whose mean error increases slightly from 18% to 20%. However, the purchase predictions still vary widely depending on which method is employed to make the predictions.

Figure 2. Comparative predictive errors: unweighted data

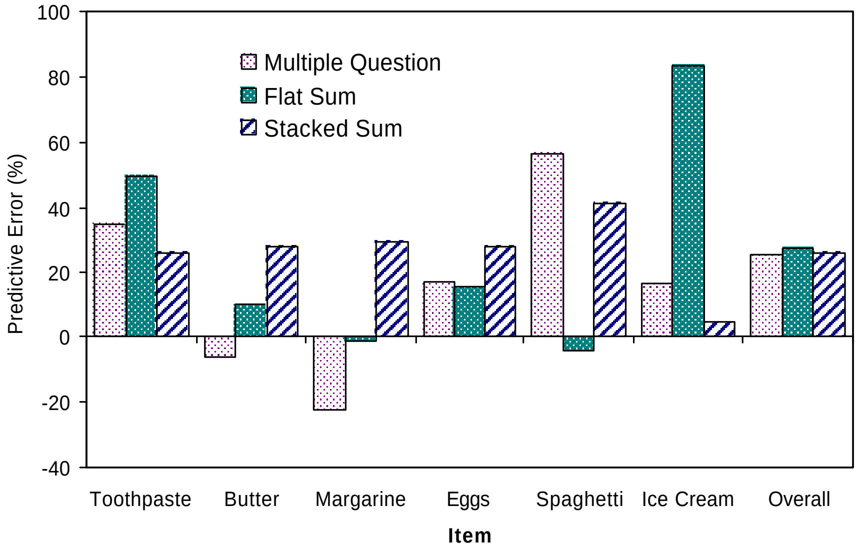
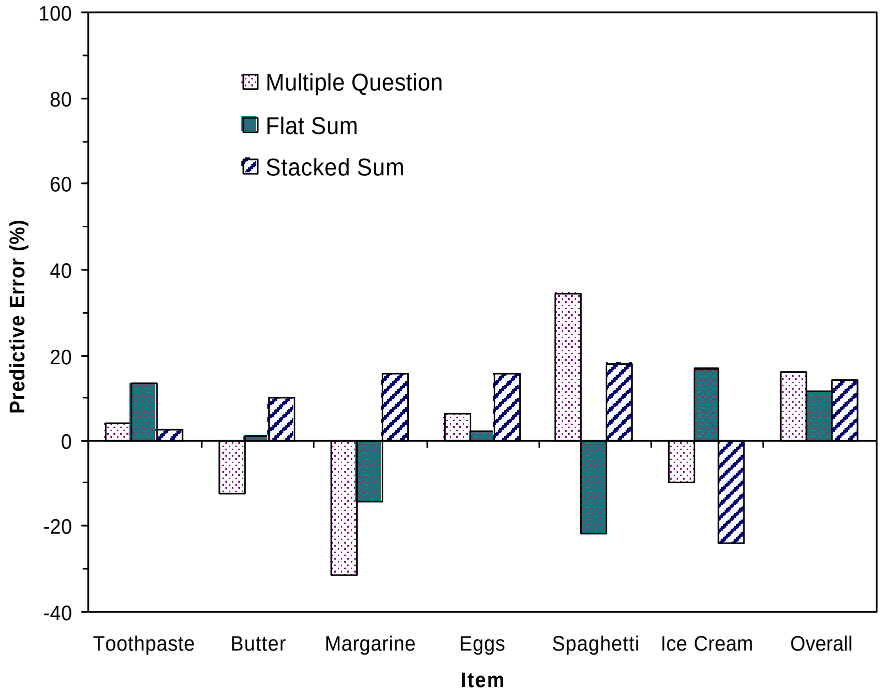


Figure 3. Comparative predictive errors: weighted data



Effect of Age and Education on Accuracy

The effect of age and education on the accuracy of prediction was investigated using multiple regression for each of the six products. The Juster method used was also included in the regression analyses to determine the extent of the predictive error that it accounted for in comparison to the two demographic variables. Dummy variables were used to make this possible. Three levels of education were used in the analyses; less than four years secondary education, four or five years secondary education, and some tertiary education. The actual age of each respondent was used. As the adjusted predictive errors were smaller than the original errors in the majority of cases, it was these that were used in the regressions.

Education and age had no consistent effect on predictive accuracy across the different products (see Table 5). Only two of the 24 results in Table 5 are significant at the 5% level, which suggests that neither age, educational level nor Juster method greatly affects the accuracy of predictions.

Table 5. Effect of respondent age and education on predictive accuracy

	Predicted Error					
	T/paste	**Butter**	**Marg.**	**Eggs**	**Spag.**	**Icecream**
Method: Stacked						
Beta	.091	.167	.117	-.063	.479	-.164
Sig T	.591	.285	.379	.642	**.009***	.438
Method: Flat						
Beta	.010	.027	.099	-.063	.080	-.019
Sig T	.951	.861	.454	.653	.657	.929
Education						
Beta	-.215	-.082	-.200	.063	-.121	-.087
Sig T	**.152#**	.530	**.094+**	.594	.417	.652
Age						
Beta	.028	-.069	.334	.104	.279	-.304
Sig T	.850	.599	**.006***	.382	**.056+**	**.096+**
Sig F	.604	.751	.016+	.844	.011+	.458

Notes:
#. Significant at the 0.2 level
+. Significant at the 0.1 level
*. Significant at the 0.01 level

Discussion

This study supports the conclusions of Hamilton-Gibbs *et al.* (1989, 1992) that the Constant Sum method produces, on average, more accurate predictions than the Multiple Question Method. Furthermore, when experienced interviewers are used, it appears that both of the Constant Sum Methods (Flat Sum and Stacked Sum) are quicker and easier to administer and less tiresome for respondents than the Multiple Question Method.

The two Constant Sum Methods produced similar levels of overall predictive accuracy, and were understood by respondents to a similar degree. However, the Stacked Sum Method consistently overpredicted purchases, whereas the Flat Sum Method did not. The Flat Sum Method produced slightly more accurate predictions, although the differences in accuracy between the methods were small and not significant at the 10% level.

Although most respondents found the two Constant Sum Methods easier to use, fuller explanations of these methods of presentation were often required, particularly for the elderly respondents. Although, overall, respondents' age did not have a significant effect on their understanding of the methods, those in the older agegroup did display greater difficulty understanding these methods of applying the Juster Scale. There is some evidence that the more elderly respondents would prefer to be administered the Multiple Question Method, or an adaptation of this method. In spite of this, the overall predictive accuracy for the elderly respondents did not differ significantly across the three methods.

As might be expected, the more highly educated respondents appeared to understand the three methods of presentation better than their less educated counterparts. As the concepts in the methods of presentation, particularly the Constant Sum Methods, are quite complex, several examples are often required, along with a certain amount of prompting. Therefore, it is imperative that well trained, and well informed interviewers are used.

Although the respondent's level of education, and to some extent, age, did have an effect on the understanding of the three methods, these factors did not have a significant effect on the accuracy of predictions.

In conclusion, the findings suggest that, while all three methods are practicable, the most effective approach is probably to use the Flat Sum Method, provide respondents with several examples to illustrate how the method is used, and employ well trained interviewers.

References

Belson W A (1986). *Respondent understanding of questions in the survey interview*. London: London School of Economics, Survey Research Centre.

Brennan M; Esslemont D & U C (1993). *A comparison of two methods for predicting purchase rates and purchase levels using the Juster Scale*. Working paper, Marketing Department, Massey University.

Clancy K & Garsen R (1970). Why some scales predict better. *Journal of Advertising Research*, *10* (5), 33-38.

Day D; Gan B; Gendall P & Esslemont D (1991). Predicting purchase behaviour. *Marketing Bulletin*, *2* (5), 18-30.

Gan B; Esslemont D & Gendall P (1986). *A test on the accuracy of the Juster Scale as a predictor of purchase behaviour*, Market Research Centre Massey University.

Gabor A & Granger C (1972). Ownership and acquisition of consumer durables: Report on the Nottingham consumer durables project. *European Journal of Marketing*, *6* (4), 234-48.

Gendall P; Esslemont D & Day D (1988). A comparison of two versions of the Juster Scale using self-completion questionnaires. *Journal of Marketing Research Society*, *33* (3), 257-63.

Hamilton-Gibbs D (1989). *Predicting the demand for frequently purchased items*. Unpublished student research report, Massey University.

Hamilton-Gibbs D; Esslemont D & McGuinness D (1992). Predicting the demand for frequently purchased items. *Marketing Bulletin*, 3, 18-23.

Juster F (1966). *Consumer buying intentions and purchase probabilities*, Report no. 99, Columbia University Press.

U C (1991). *Predicting the demand for branded products using the Juster Scale*. Unpublished student research report, Massey University.

Mike Brennan and Don Esslemont are Senior Lecturers, and Peter Seymour was a BBS(Hons) student, in the Department of Marketing, Massey University.

Appendix A. Predicted and Actual Quantities Purchased

Table 1. Predicted and actual purchases: unweighted data

Product	Multiple Question Method		Flat Sum Method		Stacked Sum Method	
	Predicted	Actual	Predicted	Actual	Predicted	Actual
Toothpaste	36.4	27	35.9	24	30.2	24
Butter	59.0	63	71.5	65	84.5	66
Margarine	43.5	56	63.1	64	56.9	44
Eggs	52.6	45	65.6	57	61.5	48
Spaghetti	50.0	32	82.2	86	57.9	41
Ice-cream	19.8	17	23.9	13	33.5	32

Table 2. Predicted and actual purchases: weighted data

Product	Multiple Question Method		Flat Sum Method		Stacked Sum Method	
	Predicted	Actual	Predicted	Actual	Predicted	Actual
Toothpaste	28.0	27	27.2	24	24.6	24
Butter	55.0	63	64.4	65	72.5	66
Margarine	38.4	56	54.9	64	50.9	44
Eggs	47.8	45	58.2	57	55.4	48
Spaghetti	43.0	32	67.5	86	48.7	41
Ice-cream	15.3	17	15.2	13	24.4	32

Appendix B: Predictive Errors

Table 1. Comparative predictive error percentages: unweighted data

	Multiple Question Method	**Flat Sum Method**	**Stacked Sum Method**	**Mean Predictive Error: Product**
Product	Predictive Error(%)[1]	Predictive Error(%)	Predictive Error(%)	
Toothpaste	34.78	49.46	25.96	36.73
Butter	-6.33	10.03	27.94	14.77
Margarine	-22.29	-1.47	29.34	17.70
Eggs	16.91	15.16	28.08	20.05
Spaghetti	56.28	-4.40	41.45	33.94
Ice-cream	16.18	83.54	4.63	26.33
Mean Predictive Error: Method	**25.46**	**27.34**	**26.18**	

Notes: 1. Predictive error = ((predicted - actual / actual) x 100)

Table 2. Comparative predictive error percentages: weighted data

Product	Multiple Question Method Predictive Error(%)	Flat Sum Method Predictive Error(%)	Stacked Sum Method Predictive Error(%)	Mean Predictive Error: Product
Toothpaste	3.78	13.36	2.42	6.52
Butter	-12.60	0.95	9.80	7.78
Margarine	-31.41	-14.25	15.66	20.44
Eggs	6.20	2.04	15.38	7.87
Spaghetti	34.50	-21.61	18.18	24.97
Ice-cream	-9.88	17.00	-23.81	14.08
Mean Predictive Error: Method	**16.40**	**11.54**	**14.08**	