

CHAPTER 5 – A GENERAL-USE ECONOMIC LIVING STANDARD INDEX

The purpose of this chapter is to describe how the items in the generic scale were used to develop the Economic Living Standard Index or ELSI. This scale has the same theoretical basis as the generic scale provided by the CFA model, and has essentially the same properties – that is, a continuous scale that assigns scores that reflect a position on the underlying latent continuum. However, it does not rely on the indicator regression coefficients from the CFA model to calculate the scales. Analysis of the responses to individual items was used to identify a set of items that were: (a) individually desirable to a substantial proportion of the sample, have high discrimination, and invariant properties (described below) across sub-groups within the population; and (b) as a set, have the same distributions across sub-groups of the number of items regarded by respondents as desirable and the number regarded as important. These conditions were intended to ensure that the ELSI scale did not have a bias caused by differences in preferences and values between subgroups.

The items of the resulting set were combined to give a score which can have a minimum of 0 and a maximum of 60 where a higher score implies a higher standard of living. Two sets of standard score intervals are provided for ease of use and interpretation – a 14-fold classification and a seven-fold classification. Finally, as a method of establishing the construct validity of the derived scale, a set of distributions of ELSI responses from various sub-groups of the population are presented (sole-parent families, Pacific Island people, Māori, couple-only families, and older New Zealanders).

Background

In the preceding chapter, it was shown that a group of items measuring ownership restrictions, social participation restrictions, economising behaviours, and self-assessments of standard of living and adequacy of income could be combined into a single scale. Confirmatory factor analysis revealed that this scale had a single underlying factor as evidenced by good indices of fit and low error estimates. The confirmatory factor analysis estimates the latent variable by means of a regression equation.

Although there are a number of advantages in using a regression equation to produce the score (e.g., treatment of measurement error, simultaneous testing of general and specific effects, overall tests of multifaceted variables, etc.), there are also several disadvantages. Most notably is the complexity of the statistical process from which the various parameters are estimated and the latent variable is calculated. In addition, the task of providing an interpretative framework is made more difficult if the scores do not incorporate some type of standardisation. Accordingly, the next stage in the development of the living standards measure was to develop and evaluate an

alternative to the generic scale using a subset of items from this scale. This alternative is based on an observable, total score approach.

In the total score approach, the individual indicators of a scale are combined in a linear fashion to produce a total score. The principle advantages of this approach are the conceptual and data analytic simplicity (Carver, 1989). In addition, once computed, total scores can be analysed in relation to any outcome that is of interest.

The development of the general use version of the scale was guided by the following criteria:

1. The measure should have a simple procedure for computing total scores to enable its easy use by a wide range of researchers and survey practitioners.
2. The measure should be robust, replicable, and useful in future surveys from which separate samples are drawn.
3. The measure should be suitable for sub-group analysis (including examination of individual groups and comparisons between groups). It is important that the given item set does not unduly bias the results of one sub-group but not another.
4. Finally, the measure should still maintain the properties of the CFA generic scale.

Thus, the development of the ELSI was guided by the need to produce a living standard scale suitable for research and policy advice that was based on a solid theoretical foundation. The remainder of this chapter is concerned with the development of a scale that satisfies these criteria.

Item Analysis

In developing the total score scale, the choice of items used in the scale is revisited. This is because in CFA modelling, it is possible to account for systematic measurement error by allowing the residual matrix of a given set of items to correlate. The final model adopted for the total population, in Figure 4.3 in the preceding chapter, allowed for correlated errors between the variables ownership restrictions and social participation restrictions, and between standard of living and adequacy of income. However, such a procedure is not possible when a total score approach is used. Items that are unreliable or variant across different sub-groups within the population will poorly measure the construct under investigation.

Approach Used in the Item Analysis

The properties of the items were examined using an approach taken from Item Response Theory (IRT), although the approach used could not be described as IRT analysis and did not employ the specialised statistics distinctive to IRT that have been developed in that area.

Whereas the CFA models used in the previous chapter account for the covariance between test items, IRT models account for participant item responses. To accomplish this, IRT models stipulate a nonlinear monotonic function (called an items response function) to account for the relation between the respondents' level on the latent variable and the probability of a particular item response (Lord, 1980). The basic assumptions in IRT modelling are that the item responses are unidimensional and locally independent. Unidimensional implies that the set of items assess a single underlying latent dimension; local independence means that if the level of the latent variable is held constant, the test items are pairwise uncorrelated (Reise et al., 1993). In the present case, the suitability of each item was assessed by examining a number of different facets of each item. Broadly, these relate to the item's 'desirability' and 'importance'; its response patterns across the score range of the CFA generic scale; its discrimination power; and whether its properties in relation to the above are invariant across sub-groups.

Item's Desirability and Importance

The first step in the scale development is to identify a set of items which are desired equally across the sample. Failure to do so may result in the inclusion of items which are, for example, only desired by those who are able to afford them. In practical terms, this means that the probability of the participants endorsing whether they would like a given item should be the same across different standards of living and sub-groups.

A second relevant aspect is the item's importance ratings. Like the desirability ratings, the probability of participants endorsing how important they perceive a given item should be the same across different levels of standard of living and sub-groups. Using items that are only regarded as important by some groups will introduce noise into the measurement of the construct as judgements of importance may impact upon item attainment. A desired item that is expensive is more likely to attract an individual's scarce resources if it is important to them, potentially causing them to forego the attainment of several other items that are less expensive but less important. By contrast, if it is desired but unimportant, it may be foregone in favour of a number of other items being attained.

The third relevant aspect concerns the item's response pattern across the continuum defined by the latent variable. This was done by generating 'item trace' curves showing the probability of respondents at different parts of the scale recording a particular type of enforced lack, or for an economising behaviour, a specified degree of economising (either 'a lot' or 'a little or a lot'). Items were assessed on the basis of whether the traces indicated a monotonic relation to the generic

scale score (i.e., fell progressively over the range, rather than showing rises and falls), had steep gradients across some part of the generic scale, and were approximately invariant across sub-groups.

The procedure used to examine importance, desirability and response pattern was as follows. The generic scale score was calculated for each respondent (using the CFA regression coefficients), enabling respondents to be ordered on the generic scale score (for lowest to highest) and divided into 14 ordered groups (analogous to decile groups), defined by 14 score intervals on the generic scale. These score intervals were chosen with a view to enabling the data to be used efficiently in the item analysis¹. For each of the ownership and participation items, a set of four item trace curves was produced showing the properties of the item in relation to its importance to the respondent, whether it was desired, and its propensity to generate an enforced lack. The traces for an item were generated by obtaining the following statistics for the respondents of each score interval:

- the proportion in the score interval who rated the item as very important
- the proportion in the interval who rated the item as fairly important or very important
- the proportion in the interval who wanted the item (those who wanted the item were operationally defined as those either having the item, or not having it but reported wanting it and gave cost as the reason for not having it)
- the proportion in the interval who had an enforced lack of the item (i.e., reported not having the item but wanting it and giving cost as the reason for not having it)

For each of these, a graphical trace was plotted, thereby obtaining a graphical representation of the way in which the proportion varied across the generic score deciles.

This was done for the complete group of respondents and each of eight sub-groups, which were defined by age (aged 18-64, aged 65+) ethnicity (Māori, non-Māori), whether part of a couple or non-partnered, and whether responsible for a dependent child (no child, one or more children). The properties of each of the ownership and participation items were thus represented by a set of 36 trace curves.

¹ Two of the generic scale score intervals were larger than the rest. They encompassed the comparatively long tail of the distribution containing people with very low living standards. Each of these intervals contained approximately 5 percent of the sample. The other intervals were of equal size, obtained by dividing the remained of the score range into 12 equal-sized parts. This procedure was judged to give the largest number of score intervals compatible with ensuring that all intervals contained sufficient numbers for the items analysis to be satisfactorily carried out.

For the economising behaviours, the information collected was limited to three-point respondent ratings of the degree of economising (using the response categories of: not at all; a little; and a lot). The traces for an economising item were generated by obtaining for each decile:

- the proportion in the decile who recorded economising a lot
- the proportion in the decile who recorded economising a little or a lot

As previously, traces were produced for all respondents and each of the eight sub-groups, giving rise to 18 trace curves for each economising item.

For the self-ratings, a parallel procedure was used. For example, for the self-rating of standard of living, a set of traces was generated relating to the proportion recording a high standard of living, the proportion recording a fairly high or standard of living, and so on.

Traces were obtained for three self-ratings, which related to standard of living (five response categories), satisfaction with standard of living (five response categories), and adequacy of income to meet everyday needs (four response categories). Each of the first two ratings generated 36 trace curves, while the third one generated 27.

Discrimination Power

In addition to an item being regarded as equally desirable and important across the population, a second key characteristic is the item's ability to discriminate between those with a low standard of living and those with a high standard of living. Generally, some items will be good discriminators of low living standard as only those with low living standards will *not* be able to obtain the item because of cost. Other items will be good discriminators of high living standards as only those with a high living standard will be able to obtain the item as cost will prevent all but those with the highest standard of living from obtaining the item. Item discrimination was examined through the inspection of the relevant item traces; an item's discriminating power was reflected in the curve's gradient at the steepest point. It was also assessed by having regard to the correlation between item score and the generic scale score.

Results

Tables 5.1 to 5.3 show the summary results of the analysis of the individual items. Judgements about the suitability of an item were determined on the basis of visual inspection against the set criteria (shown in the tables). As an illustration, the function for an item regarded as a good measure (a good pair of shoes), and an item regarded as a bad measure (dishwasher), are included in Appendix D. Summary results are not given for the three self-ratings which performed uniformly well across the criteria.

Table 5.1: Enforced lacks of ownership: item analysis

	Discrimination power for population	No desirability differences between sub-groups	No 'importance' differences between sub-groups	No discrimination differences (in shape or power) between subgroups	No discrimination power at the upper end of the living standard range	Other features	Overall judgement regarding inclusion in scale
Telephone	✓	✓	✓	✓	×		✓
Secure locks	✓	✓	✓	✓	×		✓
Microwave	✓	✓	?	?	×		×
Washing machine	✓	✓	✓	✓	×		✓
Clothes dryer	✓	✓	×	×	×		×
Waste disposal unit	✓	✓	×	×	×	×	×
Dishwasher	✓	✓	×	×	×		×
Food processor	✓	✓	×	×	×		×
Heating available in all main rooms	✓	✓	✓	✓	×		✓
A good bed	✓	✓	✓	✓	×		✓
Warm bedding in winter	✓	✓	✓	✓	×		✓
A warm winter coat	✓	✓	✓	✓	×		✓
A good pair of shoes	✓	✓	✓	✓	×		✓
A best outfit for special occasions	✓	✓	✓	✓	×		✓
Pay television	✓	✓	?	?	✓		✓
Video player	✓	✓	✓	✓	×		×
Stereo	✓	×	×	×	×		×
Personal computer	✓	✓	×	×	✓	✓	✓

Key for Discrimination Criteria

- ✓ = item met criteria
- ×
- ? = item almost met criterion

Key for Other Features Criteria

- ✓ = there is an 'other' feature that is positive
- ×
- [] = there is no 'other' feature

Key for Overall Judgement

- ✓ = include item
- ×

Table 5.1 Continued : *Enforced lacks of ownership: item analysis*

	Discrimination power for population	No desirability differences between sub-groups	No 'importance' differences between sub-groups	No discrimination differences (in shape or power) between subgroups	No discrimination differences (in shape or power) between subgroups	Discrimination power at the upper end of the living standard range	Other features	Overall judgement regarding inclusion in scale
Access to the Internet	✓	✓	×	×	✓	✓	✓	
Home contents insurance	✓	✓	✓	✓	×		✓	
Boat	✓	×	×	×	✓		×	
Car	×	×	×	×	×		×	
Holiday home, bach, or crib	?	×	×	×	✓	✓	×	
Television	×	✓	✓	✓	×		×	
A pet	×	✓	✓	✓	×		×	
An inside lavatory or toilet	×	✓	✓	✓	×		×	
Running water in house	×	✓	✓	✓	×		×	
Mains electricity	✓	✓	✓	✓	×	✓	✓	
Hot running water in the house	×	✓	✓	✓	×		×	

Key for Discrimination Criteria
 ✓ = item met criteria
 × = item did not meet criterion
 ? = item almost met criterion

Key for Other Features Criteria
 ✓ = there is an 'other' feature that is positive
 × = there is an 'other' feature that is negative
 [] = there is no 'other' feature

Key for Overall Judgement
 ✓ = include item
 × = exclude item

Table 5.2: Enforced lacks of social participation: item analysis

	Discrimination power for population	No desirability differences between sub-groups	No 'importance' differences between sub-groups	No discrimination differences (in shape or power) between subgroups	Discrimination power at the upper end of the living standard range	Other features	Overall judgement regarding inclusion in scale
Participant in family (whanau) activities	✓	✓	✓	✓	x		x
Give presents to family or friends	✓	✓	✓	✓	x		✓
Visit the hairdresser once every three months	✓	?	✓	✓	x		✓
Have a holiday away from home every year	✓	✓	x	x	✓		✓
Have a holiday overseas at least once every three years	✓	✓	?	x	✓		✓
Have a night out at least once a fortnight	✓	?	x	x	x		✓
Have family or friends over for a meal at least once a month	✓	✓	x	x	x		✓
Have a special meal at home at least once a week	✓	?	x	?	x		x
Have enough room for family to stay the night	✓	✓	✓	✓	x		✓

Key for Discrimination Criteria
 ✓ = item met criteria
 x = item did not meet criterion
 ? = item almost met criterion

Key for Other Features Criteria
 ✓ = there is an 'other' feature that is positive
 x = there is an 'other' feature that is negative
 [] = there is no 'other' feature

Key for Overall Judgement
 ✓ = include item
 x = exclude item

Table 5.3: Economising behaviours: item analysis

	Discrimination power	No desirability differences between sub-groups	No 'importance' differences between sub-groups	No discrimination differences (in shape or power) between subgroups	No other negative features	Discrimination power at the upper end of the living standard range	Other relevant features	Overall judgement regarding inclusion in scale
Less/cheaper meat	✓	✓	✓	✓	✓	✓	✓	
Less fruit and vegetables.	✓	✓	✓	✓	✓	✓	✓	
Bought second-hand clothing	✓	?	✓	✓	✓	×	✓	
Kept wearing worn clothing	✓	?	✓	✓	✓	×	✓	
Put of buying new clothing	✓	?	✓	✓	✓	✓	✓	
Relied on gifts of clothing	✓	?	✓	✓	✓	×	✓	
Continued to wear worn out shoes	✓	✓	✓	✓	✓	×	✓	
Put up with feeling cold	✓	✓	✓	✓	✓	×	✓	
Stayed in bed for warmth	✓	?	✓	✓	✓	×	✓	
Postponed/put off visits to doctor	✓	✓	✓	✓	✓	×	✓	
Postponed/put off visits to dentist	✓	×	✓	✓	✓	×	×	
Went without glasses	✓	?	✓	✓	✓	×	✓	
Not picked up prescription	✓	?	✓	✓	✓	×	✓	
Cut back/cancelled insurance	✓	✓	✓	✓	✓	×	×	
Cut back on visits to family/friends	✓	✓	✓	✓	✓	×	✓	
Cut back on trips to shops	✓	?	✓	✓	✓	✓	✓	
Spent less time on hobbies	✓	?	✓	✓	✓	×	✓	
Not gone to funeral/tangi	✓	✓	✓	✓	✓	×	✓	

Key for Discrimination Criteria
 ✓ = item met criteria
 × = item did not meet criterion
 ? = item almost met criterion

Key for Other Features Criteria
 ✓ = there is an 'other' feature that is positive
 × = there is an 'other' feature that is negative
 [] = there is no 'other' feature

Key for Overall Judgement
 ✓ = include item
 × = exclude item

This procedure revealed that a total of 14 ownership restriction items, seven social participation items and 16 economising items provided good assessments of the living standards construct. Of the initial set of ownership restrictions, more than half (15) were excluded. Most commonly this was because of differences in the item's importance ratings between sub-groups or differences in the item's discrimination power between sub-groups. Typical items that were excluded were enforced lacks of: a clothes dryer; waste disposal unit; dishwasher; and holiday home. Two items were excluded from the social participation restriction set - enforced lacks of participating in family (whanau) activities and a special meal at home at least once a week. Two items were excluded from the economising behaviour set – put off going to the dentist and cut back/cancelled insurance.

The ELSI scoring procedure

The items from the generic scale that met the selection criteria for the general use form comprised: two self-ratings (of standard of living and adequacy of income to meet everyday needs), 14 ownership restriction items, seven social participation items and 16 economising behaviours. An examination was made of possible ways of combining the items to produce a general use form of the scale that would closely mimic the generic scale. However, it proved difficult to find a simple and sensible procedure that did not increase (at least marginally) the generic scale's lean to the right. It was found that this problem could be dealt with by including a third self-rating: satisfaction with standard of living. Using the generic scale's distribution as the benchmark, the distributions were found to correspond closely. At a conceptual level the third self-rating measured a similar construct to the other self-rating items. An inspection was made of the correlation of the satisfaction rating with the generic scale, the two other self-ratings and the indicator variables for ownership restrictions, participation restrictions and economising. These correlations (which were all substantial) offered no indication that inclusion of the rating would undermine the unidimensionality of the scale.²

An examination was then made of how best to specify a scoring procedure for the augmented set of 40 items. The challenge here is arrive at a simple procedure that is easy to describe and to apply in research and policy contexts but does not result in a significant loss of information or measurement precision.

There are two key issues to be addressed: The first concerns a consideration of the different response categories for each of the scale's indicators. For the ownership and social participation restriction items, the responses are given in a dichotomous format. The responses to all other items are given on a polytomous format: the economising items are given on a three-point scale, while the self-rating items are given on a four- and five-point scale. Therefore, responses to each of these items need to be standardised in some way for the purposes of interpretation. The second issue is

² The satisfaction with standard of living variable is highly correlated with the latent variable ($r = .58$), and with the other indicators ($r = .51$ for adequacy of income, $r = .51$ for standard of living, $r = .40$ for ownership restrictions, $r = .44$ for social participation restrictions, and $r = .53$ for economising).

concerned with differences in the number of items for each set of indicators. The standardisation procedure employed will need to consider adjustments to account for the length of the scales.

Based on these issues, a number of different approaches were investigated for the best way to combine the item set together. These approaches were based on two goals: (a) that the procedure should be as simple as it can be made without (b) compromising the essential properties of the scale. With respect to the latter consideration, the critical considerations were that there should be a very high correlation ($r > .95$) between ELSI and the generic scale, and that ELSI should preserve the shape of the distribution of the generic scale.

Early attempts to combine the items were centred on dichotomising all 40 items (including the self-ratings). The paired categories thus created were scored 0 and 1, and the dichotomous values simply added together to produce the living standard score. The appeal of this procedure lay in the parallel it provided with the well-tested scoring method used for many psychometric tests. Unfortunately, it proved to be unsatisfactory. The unweighted sum of dichotomies had an unacceptably low correlation with the generic scale, indicating an appreciable loss of power, and inspection of the scale's distribution revealed extreme compression at the top end of the continuum. The most extreme case of this was given by an option that excluded self-ratings and expressed the remaining 37 items as dichotomies. When the score range was divided into equal score intervals, with the distribution expressed as the proportions of the populations falling into the intervals, the modal interval was found to be extremely right leaning.

This led to an investigation into increasing the contribution of some of the individual indicators. In deciding which indicators should have their contribution increased, consideration was given to the number of items in the indicators, and to increasing the sensitivity of the scale at the high end of the continuum. These considerations led to the development of a procedure that doubled the contribution of the social participation and self-rating indicators. This produced a scale that had a high correlation with the generic scale ($r = .98$), and hence no appreciable loss of information. In addition, an inspection of the scale's distribution showed that it closely matched that of the generic scale. In particular, a close fit was obtained at the top end of the ELSI distribution. The details of this procedure are given below.

Procedure for Obtaining ELSI Score

The procedure used to compute the ELSI score was:

1. Items were coded as follows:

Ownership Restrictions

Enforced lack – coded as 0

Ownership or not wanting – coded as 1

Social Participation Restrictions
Enforced lack – coded as 0
Participation or not wanting – coded as 1

Economising Behaviours
A lot – coded as 0
A little – coded as 1
Not at all – coded as 2

Self-rating – Standard of Living
Low – 0
Fairly low – 1
Medium – 2
Fairly high – 3
High – 4

Self-rating – Satisfaction with Standard of Living
Very dissatisfied – 0
Dissatisfied – 1
Neither dissatisfied nor satisfied – 2
Satisfied – 3
Very satisfied – 4

Self-rating – Adequacy of Income
Not enough – 0
Just enough – 1
Enough – 2
More than enough – 3

2. Responses to the items were combined into a single score using the following formula:

$$(\Sigma \text{ Ownership restrictions}) + (2 \times \Sigma \text{ Social participation restrictions}) + (\Sigma \text{ Economising behaviours}) + (2 \times \Sigma \text{ self-rating: standard of living}) + (2 \times \Sigma \text{ self-rating: satisfaction with standard of living}) + (2 \times \Sigma \text{ self-rating: adequacy of income})$$

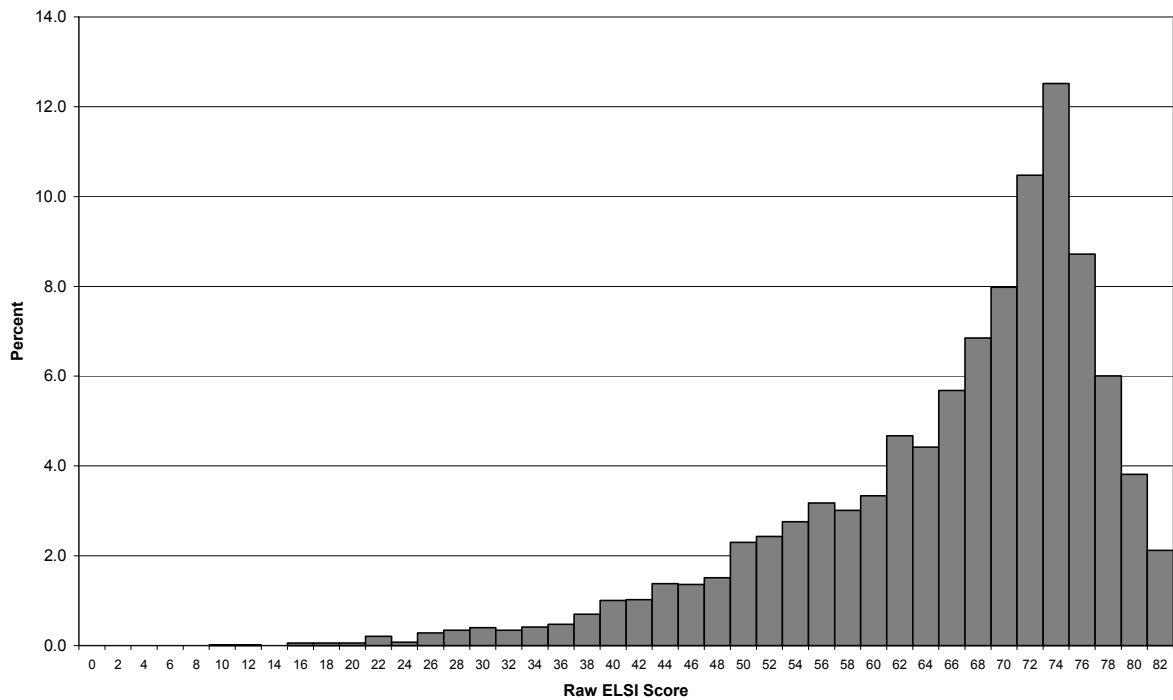
3. The application of this procedure meant that scores could theoretically range from 0 to 82. Such a range, however, is not necessarily useful as it produces a very long tail in the distribution without any significant gain in precision (see Figure 5.1 for an illustration showing the tail). As such, the minimum score that could be obtained was set as 22, and all responses less than this were coded as 22. The majority of responses in this category could reasonably be regarded as outliers.
4. Finally the range was 're-zeroed' by subtracting 22 from each score, so that the minimum was 0 instead of 22, and the maximum was 60 instead of 82. This procedure did not alter the essential properties; the purpose of this procedure was presentational, so that a respondent with the lowest possible standard of living had a score of zero.³

³ ELSI has no true zero, and scores are meaningless in the absence of information that provides a basis for their interpretation. This is not an inherent deficit and does not imply that the construct is imprecise, as many forms of measurement have no true zero (Kline, 1993).

5. It is the score produced by this procedure that is referred to from this point on as the ELSI score. For the purposes of score calculation ELSI can be most conveniently specified as follows

$$\begin{aligned}
 \text{ELSI} = & \sum^{14} \text{Ownership restrictions} \\
 & + 2 \sum^7 \text{Social participation restrictions} \\
 & + \sum^{16} \text{Economising behaviours} \\
 & + 2 \sum^3 \text{Self-ratings} \\
 & - 22 \text{ (with any total that is less than 0 being reassigned the value 0)}
 \end{aligned}$$

Figure 5.1: Distribution of ELSI scores without truncation



Scale Validity

What is the relationship then between the total score, derived from the procedures described above, and the latent variable estimate of the ELSI score, derived in the previous chapter? The correlation coefficients between the total score and the latent variable estimate are presented in Table 5.4, on the following page, for the total sample and for the sub-groups within the sample. These coefficients show a very high degree of association between both sets of scores, supporting the validity of the total score approach. Thus, for all practical intents and purposes, there is no substantial loss in information from using the total score scale instead of the latent variable model.

Table 5.4: Correlation coefficients between total score and latent variable score

Group	Correlation: Total Score and Latent Variable
Total Sample	-.98
Age 18-24	-.98
Age 25-44	-.99
Age 45-64	-.99
Māori over 65 years	-.98
Total over 65 years	-.97
Working-age Māori	-.98
Working-age non-Māori	-.99
Working-age people with children	-.99
Working-age people without children	-.99

Note: negative values are due to the generic scale and the ELSI score scored in opposite directions: Increasing generic scale scores imply decreasing standard of living; increasing ELSI scores imply increasing standard of living.

Specifying a standard set of score intervals

For many uses of a measure such as ELSI, it is convenient group results into score intervals. For example, when there is a need to present results indicating the distribution of a set of scores (e.g., of a group of a survey respondents), or to compare the distributions of two sets of scores, this is commonly done by means of a table or histogram showing the numbers of scores falling in specified score intervals.

It was considered useful to use the survey results on the distribution of ELSI scores to specify a set of score intervals likely to suit most purposes for which such intervals are likely to be put.

The following criteria were used to guide the task of specifying a standard set of score intervals:

1. not sacrificing useful discriminating power at the high living standard end of the scale
2. not getting undue bunching into the bottom couple of intervals for sub-populations with low overall living standards
3. having a fairly compact set of ranges, for example, 10-15 intervals for the primary set
4. having a secondary set of ranges that is more compact, for example, less than 10 intervals
5. having enough categories in the lower living standard region to permit debates and choices about where poverty thresholds might be specified
6. a bottom category (the low living standard end) that contains only a small proportion of the population (less than 10 percent).

Two sets of intervals were defined for the ELSI scale. The primary set, intended for use in fine-grained analysis of the distribution of ELSI scores, consists of 14 intervals. The secondary set consists of these intervals combined in pairs, giving seven broader intervals.

These seven broader intervals have been referred to as living standard levels, labelled from lowest to highest as level 1, level 2, etc. up to level 7. The lower and upper parts of level 1 have been labelled level 1L and level 1U, with the same convention applied across the range, so that the lower and upper parts of level 7 is are labelled level 1L and level 7U.

The scores that define the boundaries of the intervals are given in Table 5.5.

Table 5.5: Standard score ranges for the ELSI

ELSI Score Intervals	Primary Intervals (14)	Living Standard Levels (7)
≤ 11	Level 1L	1
12 – 15	Level 1U	
16 – 19	Level 2L	2
20 – 23	Level 2U	
24 – 27	Level 3L	3
28 – 31	Level 3U	
32 – 35	Level 4L	4
36 – 39	Level 4U	
40 – 43	Level 5L	5
44 – 47	Level 5U	
48 – 51	Level 6L	6
52 – 55	Level 6U	
56 – 59	Level 7L	7
60	Level 7U	

Note: L = Lower, U = Upper

Concurrent Validation of the ELSI Scale

In Table 4.9 (in Chapter 4) the correlations between the generic scale and various alternative factors associated with living standards were presented. Increasing equivalised income and housing adjusted equivalised income were associated with increasing living standard scores. Likewise, factors such as being unable to save, or unable to obtain \$5000 in an emergency were associated with lower living standard scores. To what extent do these findings generalise from the latent variable score to the total score scale developed in this chapter? Table 5.6 shows the correlations between the latent variable ELSI score, the total ELSI score, and the associated measures of standard of living that were gathered during the surveys. To reiterate, these measures included:

1. housing-adjusted equivalised disposable income
2. equivalised disposable income
3. for older people (age 65+): whether the respondent reported being unable to save on most months
4. whether the respondent reported being unable to find \$5000 in an emergency
5. for older people (age 65+): whether the respondent reported health-related financial stress in the past 12 months
6. for older people (age 65+): whether the respondent reported being in possession of a Community Services Card
7. for older people (age 65+): Material Well-being Scale (MWS)

Table 5.6 shows that similar-sized correlation coefficients were obtained from the ELSI total score as were obtained from the generic scale – although the direction of the relationship is different due to each scale being scored in the opposite direction. ELSI scores were positively associated with increasing household adjusted equivalised income and increasing equivalised income. ELSI scores were negatively associated with being unable to save money most months, being unable to obtain \$5000 in an emergency, health-related financial stress, and possession of a Community Services Card. Thus, the pattern of associations observed for the generic scale is preserved by the ELSI scale.

Table 5.6: Correlation coefficients between ELSI and validation measures

Measure	Generic Scale	ELSI
1. Housing adjusted equivalised disposable income	-.52*	.54*
2. Equivalised disposable income	-.43*	.44*
3. For older people (age 65+): Unable to save	.40*	-.42*
4. Unable to get \$5000	.52*	-.53*
5. For older people (age 65+) health-related stress	.43*	-.44*
6. For older people (age 65+) possession of Community Services Card	.33*	-.35*
7. For older people (age 65+) MWS	-.94*	.94*

Note: Generic scale scored so that an increasing score implies decreasing standard of living; ELSI scored so that an increasing score implies an increasing standard of living; * $p < .0001$.

Construct Validation of the ELSI Scale

Construct validation refers to an assessment of the extent to which the postulated attribute actually exists. It can be established using a variety of methods, many of which overlap with other methods of validation. For the purposes of the present research, it was decided to examine how the distributions of living standard scores differ across various sub-groups within the population. Groups for which a low standard of living might be expected include: sole-parent families, Pacific Island people, and

Māori. A higher standard of living could be expected amongst couples without children and older New Zealanders (which Fergusson et al., 2001, found to have better living standards on average than the population as a whole).

The distribution of ELSI scores amongst these groups is examined and compared with the ELSI distribution for the total population. It is expected that the construct validity of the ELSI scale will be established by showing that those groups who have consistently been shown to have lower living standards using a variety of indicators (e.g., sole-parent families, Pacific Island people, Māori) will have distributions skewed towards the low end on the ELSI scale, and the mean scores for these groups will be below the mean for the population. By contrast, groups that have been found to have higher living standards using different indicators (e.g., couples without dependent children, and people aged 65 and over) will have distributions that are skewed more towards the high end of the ELSI scale, and mean scores that are above the mean for the population.

This section begins with the ELSI distribution for the total population. This is followed by an analysis of the ELSI distributions for the groups discussed above. For the purposes of presentation, distributions are presented such that groups who have been consistently found to have lower living standards first, and groups with higher living standards presented second.

Figures 5.2 and 5.3 show the distribution of living standard scores grouped by the 14 categories (Intervals) and by the 7 categories (Levels). Both figures show that the distribution of living standard scores is substantially skewed, with a long left hand tail for a fairly small number of people (20 percent) who have relatively low living standards scores (levels 1, 2, 3). As will be discussed in more detail later, the skew should be interpreted as a greater degree of discrimination using the ELSI scale at the low end of the continuum (levels 1, 2, 3) than at the high end of the continuum (levels 6 and 7). The mean ELSI score for the population is 41.95 ($SD = 12.21$; $LCL = 41.49$; $UCL = 42.40$ for a 95 percent confidence interval).

Figure 5.2: Distribution of ELSI scores for total population – 14-fold classification

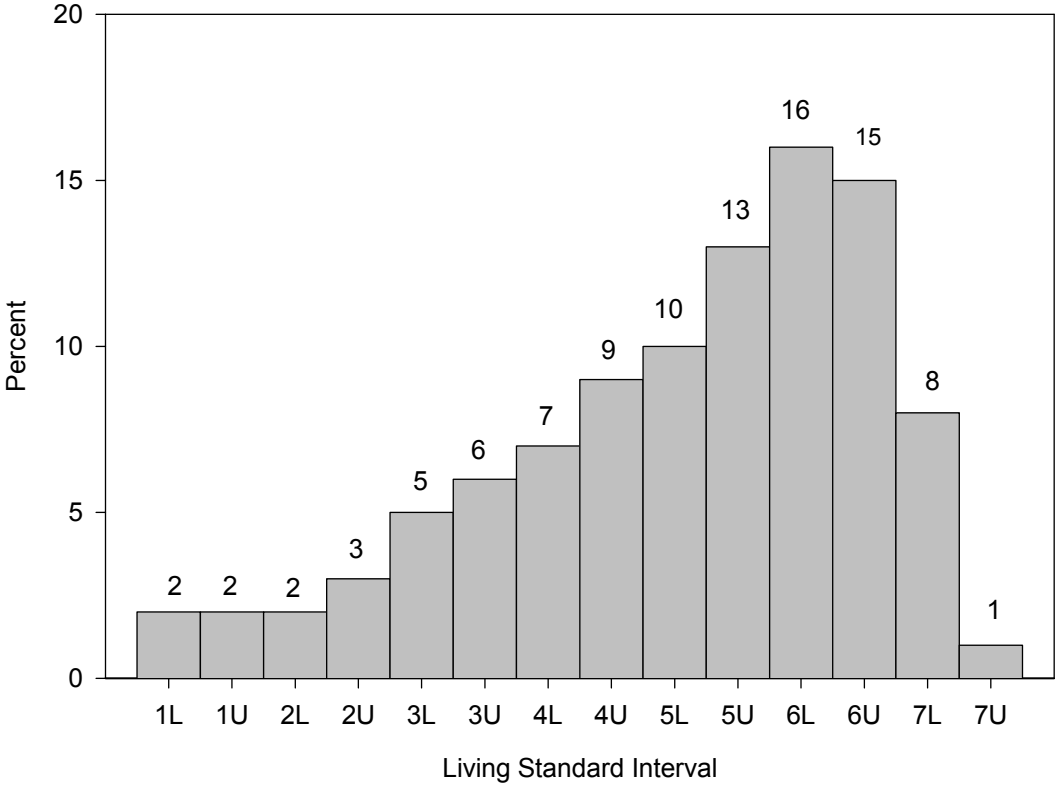
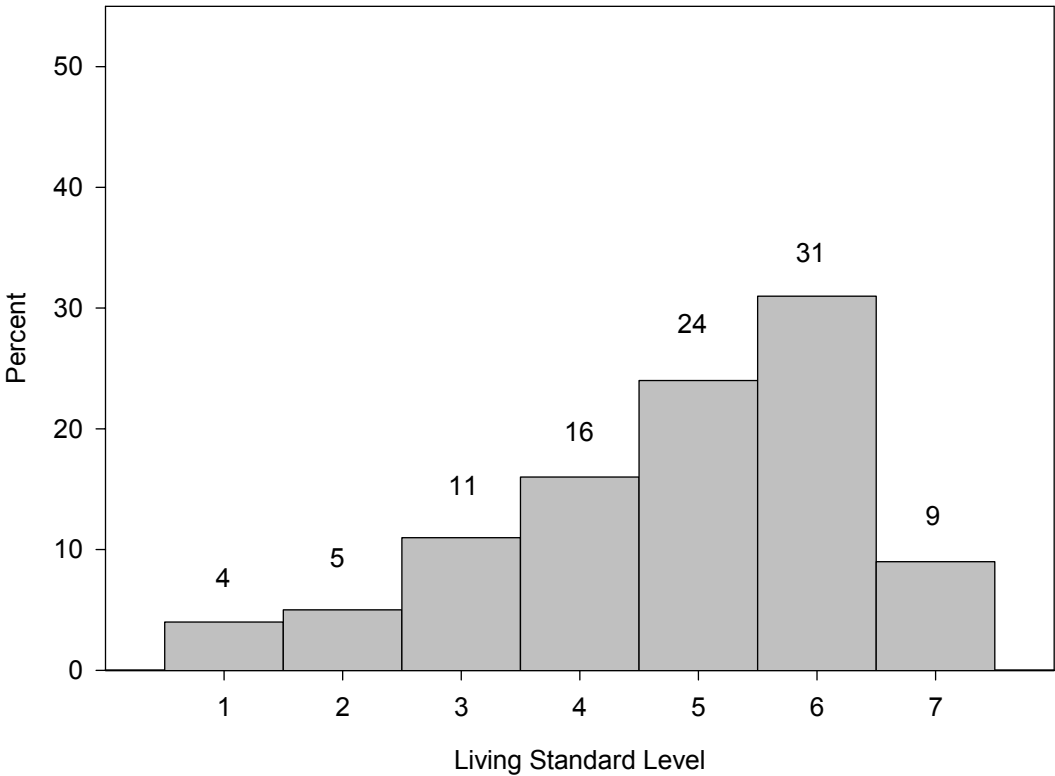


Figure 5.3: Distribution of ELSI scores for total population – 7-fold classification



Distribution of ELSI Scores Amongst Groups Expected to Show Lower Living Standards

The ELSI distributions in Figures 5.2 and 5.3 provide a basis for comparison with sub-groups within the population. ELSI distributions are presented for: sole-parent families (Figure 5.4), Pacific Island people (Figure 5.5); and Māori (Figure 5.6). Note that the majority of sole-parent families are benefit recipients.

The distribution of ELSI scores for people in single parent families is presented above in Figure 5.4. This histogram shows that this group contains a large percentage of people who have a low standard of living. Over one quarter of this group fall into levels 1 and 2. The mean ELSI score for this group is 29.7 ($SD = 12.9$; $LCL = 28.13$; $UCL = 31.20$).

Figure 5.5 shows that Pacific Island people are highly *over*-represented at the low end of the ELSI scale (levels 1 to 3) but highly *under*-represented at the high end (levels 6 and 7) when compared to the total population in Figure 5.2. In addition, the percentage of Pacific Island people in levels 1 to 5 is much more evenly dispersed than for the total population, where a much more pronounced pattern was observed. The mean ELSI score for this group is 32.85 ($SD = 14.12$; $LCL = 30.50$; $UCL = 35.20$).

Figure 5.6 shows the ELSI distribution for Māori. On average, scores for Māori were clustered around the middle of the ELSI distribution. Like Pacific Island people, Māori were over-represented at the bottom end of the scale, and under-represented at the top end of the scale. In addition, there was some evidence that ELSI scores for Māori have a mildly bimodal distribution: peaks were observed at the levels 3 and 5. The mean ELSI score for this group is 35.64 ($SD = 12.54$; $LCL = 34.30$; $UCL = 36.98$).

Figure 5.4: ELSI distribution of population in sole-parent families with dependent children

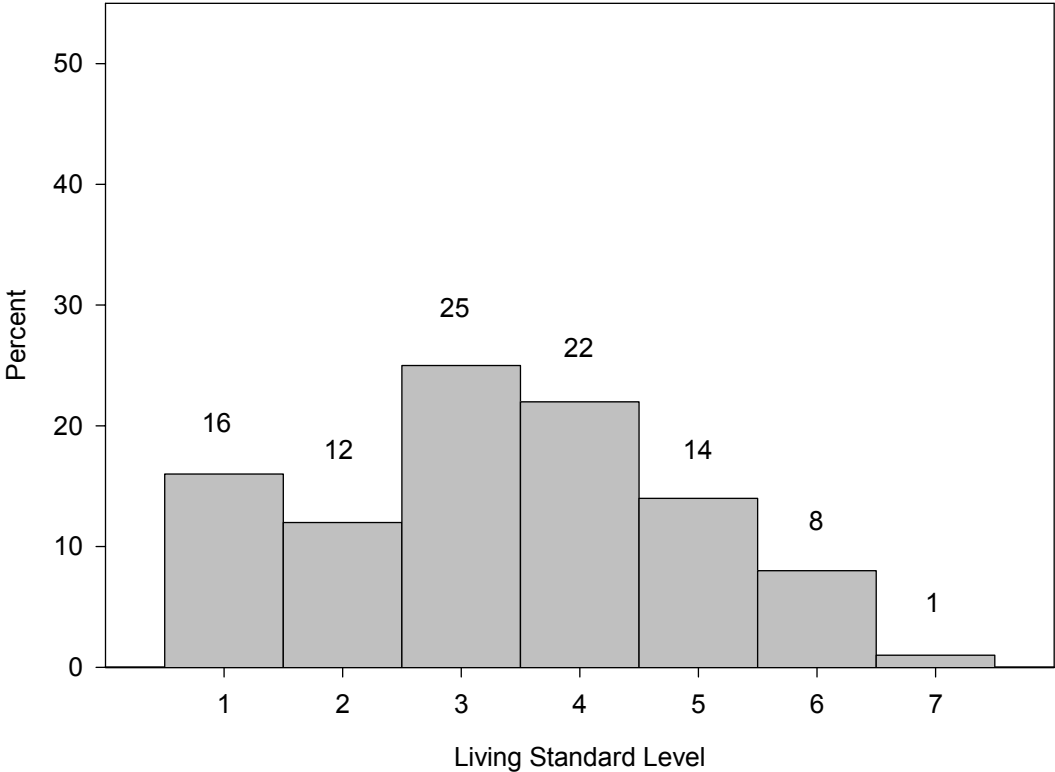


Figure 5.5: ELSI distribution for Pacific population

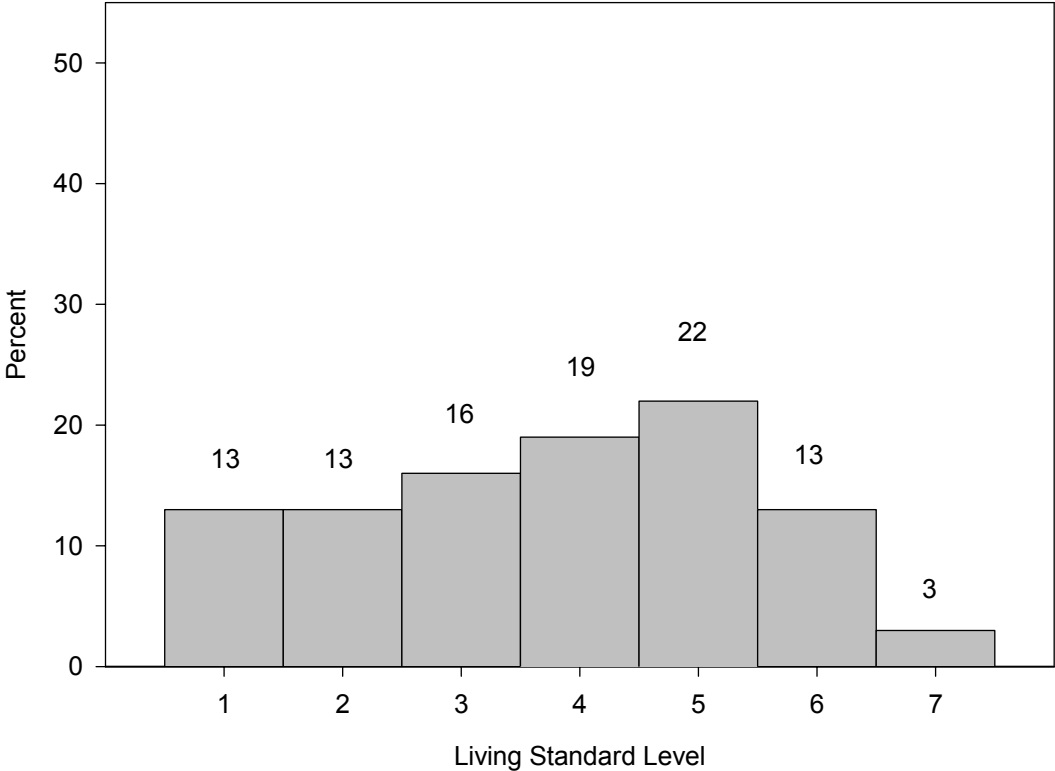
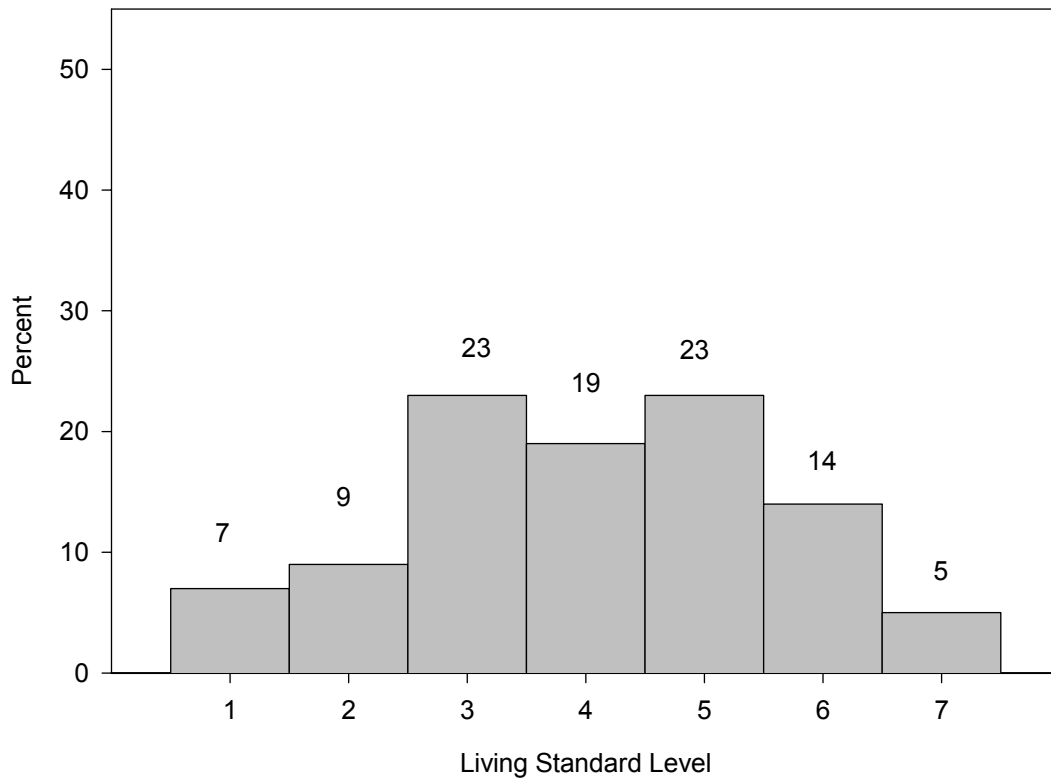


Figure 5.6: ELSI distribution of Māori population



Distribution of ELSI Scores Among Groups Expected to Show Higher Living Standards

As indicated earlier, it would be expected that above average scores would be found for couples without children and people aged 65 and over. This section examines the standard of living of each of these groups. As with the previous analysis, the comparative group is the total population.

Figure 5.7 shows the distribution of ELSI scores for couples without children. This histogram shows that this group has elevated ELSI scores when compared to the population. For this group, less people were in levels 1 to 5 than for the population, and a greater number of people were in levels 6 and 7. The mean ELSI score for this group is 46.60 ($SD = 10.07$; $LCL = 45.99$; $UCL = 47.22$).

Finally, the distribution of ELSI scores for older New Zealanders is presented in Figure 5.8. As expected, this group had a very high standard of living when compared to the total population. Older New Zealanders were under-represented at the low end of the ELSI scale (levels 1 to 4), and over-represented at the high end (levels 5 to 7). The mean ELSI score for this group is 47.42 ($SD = 8.72$; $LCL = 47.04$; $UCL = 47.80$).

Figure 5.7: ELSI distribution for population in couples without children

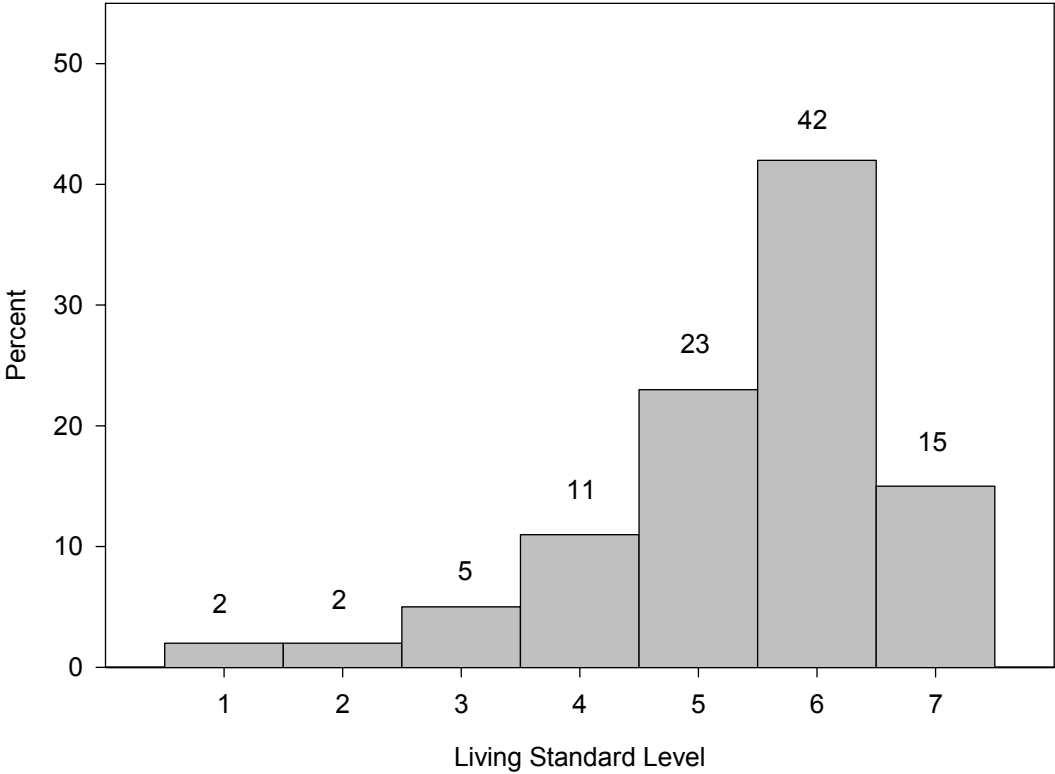
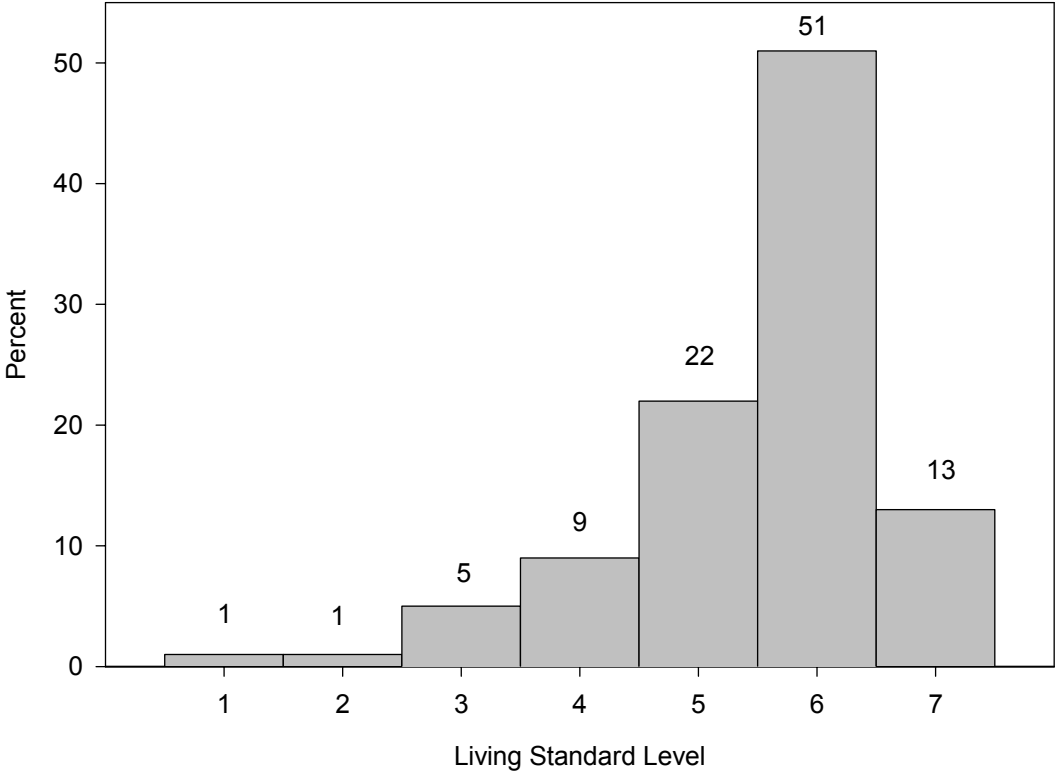


Figure 5.8: ELSI distribution for older New Zealand population (65 and over)



Summary

This section sought to establish the construct validity of the ELSI through an examination of the distribution of ELSI responses for sub-groups of the population which have been consistently shown to have low living standards using a variety of measures. When compared to the total population, people in single-parent families, Pacific Island people, and Māori were all over-represented at the low end of the ELSI scale, but under-represented at the top end. Mean ELSI scores for these groups were significantly lower than for the total population. Couples with children and older people were all under-represented at the low of the ELSI scale, and over-represented at the top. ELSI scores for those in couple-only families and older people were significantly higher than the mean for the population. These results support the construct validity of the ELSI scale. Similar analysis of ELSI living standards distributions can be found in *New Zealand Living Standards 2000* (Krishnan et al., 2002).

ELSI Scale Limitations

The validation results provide some support for the usefulness of the ELSI scale as a measure of living standards. However, the scale has some limitations that could usefully be addressed by further research.

1. Lower discrimination at the top end of the scale: While there is broad agreement as to what constitutes a minimum standard of living it is more difficult to achieve agreement regarding a high standard of living. This presents a challenge when developing a scale to measure higher living standards: differences in personal preference mean that a large number of items are required to capture individual differences in preference; and it is difficult to control for the different costs of top-end items (cf. cost of overseas holiday vs. the cost of internet access). The present research attempted to increase the discrimination at the top end of the scale by including three self-ratings of standard of living. This attempt was somewhat successful: score differences in the upper part of the range appear to have discriminating power (as judged by calibration data and similar results from income and assets, which show a positive gradient with scale scores); although score values are probably closer together than they ideally should be. The most promising way of extending the top part of the range is to include items that make distinctions of quality. It seems reasonable to assume that in contemporary societies part of what distinguishes the poor from the rich is not the complete lack of certain amenities or consumer goods, but rather more subtle distinctions of quality. For many people, a pair of Nike sports shoes is not equivalent to a pair of bargain-basement shoes. The failure to incorporate this distinction into items does not mean that it is not captured in the scale scores to some extent. For instance, it is likely that people with high scale scores (e.g., a person who has overseas holidays and do not economise at all on shopping, or the pursuit of hobbies, etc.) will have better shoes on average than people with low scale scores. In other words, some items that contribute to high scale scores are probably functioning partly as

proxies for items that measure quality. Nonetheless, items of this type could be expected to increase both the scales discriminating power and validity.

2. Effects of differences between people in their preferences: The precision of the scale in measuring the living standards of a particular person can be affected by extent to which the person wants the things that form the basis of the scale. An item contributes information about a person's living standard only when it relates to something that the person wants. If all forty of the ELSI items are germane to a person, the item set would constitute a substantial sample of the population of potential items that could contribute to measuring the person's living standard. On the other hand, if only thirty of the items are germane to the person, those items offer a lesser amount of information about the person's living standard, and the possibility of biases in the score. In general terms, differences between people in the number of items that are germane can be expected to diminish the precision of the scale.
3. Potential for response bias: Many of the items have a subjective element, which creates a possibility of response bias. The subjective element is greatest with respect to the self-ratings: two people whose standard of living is essentially the same may rate themselves differently because they are more or less sensitive to deprivation and take more or less pleasure in the comforts they have. However, less conspicuously, there is also a subjective element involved when people indicate whether they engage in a particular economising behaviour 'a lot' or 'a little' or 'not at all', and indicate whether they want an item of property they don't have.
4. The effects of the social and historical context: The items in the ELSI have been selected as ones that provide useful information for a particular population. Because the scale was developed in a particular social and economic context, it is likely to diminish the utility as this context changes. For example, in the 1974 Survey of the Aged, access to a telephone was a scale item that was found to predict variation in living standards at the lower end of the scale. In the current research, this item did not prove to be informative as it was almost universally owned. While this is not a fundamental difficulty for the measurement theory, it means that – as with the Consumer Price Index – the item content is required to be reviewed periodically to ensure that the instrument's validity or sensitivity do not diminish. If this is not done, the specificity the item set and scoring to a particular social and economic context may diminish the appropriateness for time series analysis and for social monitoring.
5. The exclusion of housing: As noted earlier, accommodation problems were excluded from the CFA analysis as previous research had suggested that the inclusion of this variable would cause the model to be rejected – even though there is a statistical association between accommodation problems and living standards. Subsequent analysis to determine whether the inclusion of this item would have led to a significantly bad fit between the model and the

data was supported. It would be desirable to clarify the relationship between housing quality and the living standards construct measured by ELSI. Housing quality has commonly been regarded as an aspect of living standards. However, if it truly does not fit within a unidimensional construct containing the components of ELSI, then it may be desirable to develop a separate scale for housing, which in some contexts would be used in parallel with the ELSI scale.

6. Scope of ELSI: Finally, when planning research it will need to be kept in mind that economic living standards, as measured by ELSI, represents only one aspect of well-being. When the purpose of the research requires that consideration be given to other aspects, such as quality of life, life satisfaction, happiness, and so on, then it will be necessary to include separate measures of those things. Some purposes may be best served by using an extensive suite of measures, and sometimes it may be useful to combine several measures into a composite (although this would require an understanding of the characteristics of the new measure thereby created).

Concluding Comment

In this chapter, the approach used to develop a total score measure of economic living standards was presented. The ELSI scale has the same theoretical basis as the CFA generic scale developed in the previous chapter, and essentially has the same statistical properties although it does not depend upon CFA regression weights to estimate the scale scores. Rather, responses are combined according to a simple rule to produce a total score. Analysis of the responses to individual items was used to identify a set of 40 items that were suitable for measuring the living standards latent variable across different groups within the population. Further analysis was made to devise a rule for combining the items to produce a total score measure (ELSI) that closely approximated the CFA generic scale.

The ELSI scale was highly correlated with the generic scale and closely preserved its distribution. The construct validity of the scale was tested through a systematic examination of the distribution of responses of various sub-groups within the populations. As expected, the sub-groups with lower living standards, as assessed by other independent measures, were represented at the low end of the ELSI distribution; the sub-groups with higher living standards, using the same measures, were represented at the high end of the ELSI distribution. These results support the construct validity of the ELSI.

