From Conceptual Frameworks to Testable and Simplified Models of Language Learning Strategy Assessment

Indika Liyanage, Fiona Bryer, and Peter Grimbeek
Griffith University, Queensland, Australia

In the study of how students learn a second language, inventories have provided a readily available methodological instrument used to assess strategies. Items describing ways of learning a new language have been written, added to similar items, and organised into categories of types and subtypes of strategies. In the construction of one inventory for learning Russian and Spanish (Chamot, O'Malley, Kupper, & Impink-Hernandez, 1987), items were linked to conceptually defined types of strategies (specifically, metacognitive, cognitive, and social-affective approaches to learning foreign languages). A 63-item version of this inventory has been adapted to assess strategies of learning English as a Second Language (ESL), and strategic preferences of ESL high school students from Sri Lanka and Japan have been linked to learner variables such as ethnicity, mother tongue, religion, gender, and personality type (Liyanage, 2004, 2005). More recent scrutiny of participant responses to this language learning strategies inventory (LLSI), has exposed limitations in its notional classification of strategies and basic design. Factorial analysis of the responses of large heterogeneous samples did not support preconceived links between items, strategies, and types of strategies. Moreover, the different numbers of items for the three types of strategies affected measurement properties. Re-analysis involving content analysis of inventory items, together with exploratory and confirmatory factor analysis utilising a large data set, revealed a much simplified model of language learning. Clear differences between what LLSI scales were supposed to test and outcomes based on analyses of (a) item content and (b) responses elicited by these items indicate the need for caution when using inventories to measure strategies for learning a second language.

Conceptual frameworks for instrument design
It seems a contradiction in terms but the expertise and prior knowledge of instrument designers could be viewed as a kind of blind spot. Tversky and Kahneman (1973) discussed the idea that the decision-making of experts can be swayed by the availability (salience) of specific aspects of objects under scrutiny in such a manner that they make judgments based on subjective rather than objective probabilities. For instance, medical doctors with years of training and clinical practice can misdiagnose medical conditions based on the salience of certain symptoms. It follows that instrument designers can similarly misjudge the usefulness of an instrument and the meaningfulness of test material. More specifically, precisely because the expert instrument designer is familiar with the material, the expert tends to "see" items as naturally agglomerating to form certain conceptually based scales and to measure certain latent variables. However, these
items, scales, and latent variables might not be quite so salient to a naïve viewer who lacks awareness of the supposed importance of these features and does not share the values and expertise of the test designer. It should not be surprising, therefore, that this blind spot can lead to difficulties in reconciling the scales and latent variables originally reported by the instrument designer with outcomes based on empirical analysis of the factor structure of an instrument after its administration to a large sample.

Consultations with panels of content experts do not mitigate this problem because the expert knowledge of these panel members can be blind for much the same reason. Another option for checking the psychometric properties of a newly developed instrument is a small-scale trial of that instrument. Yet, small-scale trials also are unlikely to challenge prevailing perceptions or preconceptions about what the instrument measures because the sample size is rarely large enough to enable more than a cursory analysis of the ways in which items cluster to form scales, let alone providing insights about the latent variables expressed by these items and scales.

Given the cost and effort involved in administering a large scale testing program of such an instrument, an instrument designer would be better placed if more fully able to anticipate the likely clustering of items into scales and the likely emergence of latent variables. In that respect, the technology of automated text analysis (e.g., Leximancer: Smith, 2005) provides such a mechanism. Tversky and Kahneman (1973) concluded amongst other things that machine analysis might be better able to accurately diagnose a patient's medical condition than a live doctor. In this paper we show that machine-based text analysis has the potential to provide a view of test items and scales that is equivalent to that obtained by large-scale administrations of the instrument and unlike that of the test designer. That is, automated text analysis can be used to identify the distinctive features of text based on objective probabilities as opposed to what might have appeared salient and available (i.e., what seems subjectively probable) in the mind of the instrument designer.

With the prospect in mind that modern content analysis can contribute to the discussion about objective and subjective probability in instrument design, one such instrument at present in use in applied linguistics—the Language Learning Strategies Inventory (LLSI)—is outlined, and the instrument designers' account of that instrument is compared with outcomes based on content analysis and factor analysis.

Inventorying strategies for learning language

The definition and classification of language learning strategies (LLS) have been problematic. However, two taxonomies (O'Malley & Chamot, 1993; Oxford, 1994) have had a major influence on recent work. These taxonomies have provided a rich resource of ideas for research about how different strategies affect processes involved in second language learning. Oxford consolidated 62 strategies that included every strategy previously cited in the language learning literature. This taxonomy has provided a widely used platform from which to generate items for a questionnaire and, thus, to assess how learners have used strategies in second language learning. Chamot and colleagues (Chamot, O'Malley, Kupper, & Impink-Hernandez, 1987; see, also, Chamot & O'Malley, 1993; O'Malley & Chamot, 1993; O'Malley, Chamot, Stewner-Manzanoares, Kupper, & Russo, 1985),
However, proposed that some theoretical justification would improve discriminative classification of strategies.

The strategies in both taxonomies were defined in terms of the learner's actions, such that learners have been said to use strategies deliberately, consciously, and selectively (i.e., in a way that reflects preferences in how to learn a new language). Oxford (1990, p. 8) described "specific actions taken by the learner to make learning easier, faster, more enjoyable, more self-directed, more effective and more transferable to new situations" (p. 8). Chamot et al. (1987), however, argued that descriptions of the influence of LLS on learning and memory processes of learners need a theoretical model for the learner's actions. Therefore, O'Malley and Chamot (1990) described strategies as "the special thoughts or behaviours that individuals use to help them comprehend, learn or retain new information" (p. 1).

Chamot et al. (1987) chose a cognitive theory proposed by Anderson (1981, 1996, 2000) to classify strategies in their 48-item Learning Strategy Inventory (LSI). They described LLS in terms of how and at what level—metacognitive, cognitive, and social-affective—learners process new information. Detailed discussion of this information-processing approach has been provided elsewhere (Chamot et al., 1987; Liyanage, 2004). The study by Chamot et al. (1987) involved both quantitative and qualitative means of data gathering at two stages: Quantitative procedures with the LSI in the first stage were followed by qualitative procedures with post-task interviews in the second stage. The LSI used in the first stage was designed to gather 16 different strategies for learning Spanish and Russian. On completion of the second stage, Chamot et al. discovered 10 additional strategies used by the participants for which no questions had been framed in LSI.

Unlike Oxford's pragmatic approach to describing and labelling strategies, the cognitively based definition and classification of LSI items enabled prioritisation of strategies of most importance to learning. This elaborate scheme also included clustering of items into substrategies within the three major categories, with clearly specified non-overlapping boundaries between some subcategories based on the information-processing model. According to O'Malley and Chamot (1990), metacognitive strategies would serve to facilitate the learner's planning for learning tasks and, thus, help the learner to oversee how well the plan is implemented. Strategies in this metalevel category could also help learners to evaluate their success at task completion. At a more basic level, cognitive strategies would serve to perform three tasks (i.e., repetition, rehearsal, & elaboration). Repetition refers to imitating language models exactly by activities such as repeating a list of words, silent practising, or copying. Rehearsal refers to the practising of language, with attention to meaning, for a written or oral task. Elaboration refers to the linking of new information to prior knowledge, either linking personal associations to the new information or linking different parts of new information. O'Malley and Chamot included a third category of social-affective strategies in their taxonomy. This category gave recognition to the interactional exchanges involved in learning a second language and the social context within which information is processed. It was speculated that strategies of this kind serve to facilitate the interaction between a learner and a speaker of another language, and items were written to capture this aspect of language learning.
Cultural strategies for learning language

A series of two studies by Liyanage (Liyanage, 2004; Liyanage, Birch, & Grimbeek, 2004) examined aspects of the influence of personal (e.g., personality) and contextual (e.g., cultural background) variables affected language learning. Liyanage and colleagues used an adapted version of LSI (Chamot et al., 1987) to measure how 16- and 17-year-old high school students from four ethnicities were learning English. Liyanage (2004) adapted the original 48-item LSI to embrace the 10 additional strategies identified by O'Malley and Chamot (1990). In total, Liyanage added 16 questions, making a total of 63 items and 26 strategies in his adaptation of the Chamot et al. inventory (see appendix). In the first study, Liyanage (2004) tested a representative sample of 948 high school students learning English as a Second Language in six government schools in Colombo, Sri Lanka. The sample of Sinhalese (n = 301), Tamil (n = 283), and Muslim (n = 299) students represented the three dominant subcultures in the country with their differing ethnicity, first language, and religion. Liyanage translated the LLSI into the first language of each subsample. When Liyanage (2004) explored the association between LLS and learner variables (e.g., personality type, ethnicity, gender, religion, and mother tongue) for those 883 participants who had no missing data, he found extremely significant associations between these variables (p < 0.000).

Second, Liyanage et al. (2004) examined whether ethnicity or religion was more important in determining choice of strategies when students were learning English as a second language (ESL). The LLSI used with the Sri Lankan study (Liyanage, 2004) was translated into Japanese by a competent translator and administered to a group of Japanese students studying English in Queensland. The rationale for adding this Japanese sample (N = 134) was that, despite being ethnically different to the Sinhalese sample, they have a similar religious identity (Buddhism). Therefore, Liyanage et al. were able to compare the two religiously similar but ethnically different groups (Sinhalese-Buddhist and Japanese-Buddhist) with another two groups (Tamil-Hinduism and Muslim-Islam) with closely related ethnic and religious identities.

It was hypothesised that similarities between the Japanese and Sinhalese groups would indicate that religious identity had superior influence on the learners' choice of language learning strategies. On the other hand, marked differences between these two groups would indicate the superior influence of ethnicity. Similarities in the responses of Sinhalese and Japanese students supported the view that the religious identity of learners was more important than ethnic identity in determining the selection of learning strategies. The specificity or generality of this apparent conclusion, however, remains to be determined through replication studies in various other contexts (e.g., with ethnically different Muslim and ethnically different Hindu students).

Although earlier researchers could and did introduce new schemes to classify and describe language learning strategies, they could not justify any particular scheme describing preferred strategies over any other way of assessing strategies (Chamot, 1987). Both the atheoretically accumulated Oxford taxonomy of strategies and the theoretically organised O'Malley and Chamot taxonomy served to stabilise the listing of strategies used in research, but language researchers using the LSI were interested in measured strategy preference as a basis for studying language characteristics and for improving
ESL instruction. They did not seek to empirically validate the information-processing model of strategy classification.

Although the Liyanage studies did point to contextual influences on language learning strategies and, also, to personality influences (Liyanage, Grimbeek, & Bryer, 2006), the complex links between items, specific substrategies, and the three types of strategies in the adapted inventory posed difficulties for data interpretation. That is, inspection of test items raised methodological questions about what the inventory of strategies was measuring. The arrangement of the scale allocated a disproportionate numbers of items and strategies to the three types of strategies measured within these scenario-based settings. The adapted version collated by Liyanage contained (a) metacognitive learning strategies with 7 substrategies and 20 items, (b) cognitive strategies with 15 substrategies and 34 items, and (c) social-affective strategies with 4 substrategies and 9 items. Given the disparity in items per subscale, one might expect measures of cognitive learning strategies, for instance, to be more reliable than measures of social-affective strategies. One surmise is that the instrument designer overlooked the uneven distribution of items relative to scales representing metacognitive, cognitive, and social-affective strategy types because of the salience or availability (Tversky & Kahneman, 1973) of these aspects of the instrument seemed of greater significance. In mitigation, the third author, in his role as research methodologist, has observed conceptually sensible agglomerations of items under such circumstances.

The ability to access large and varied samples permitted examination of the measurement properties of the LLSI through quantitative analysis of item responses and latent variables. Grimbeek, Bryer, Liyanage, and Birch (2005) re-analysed ESL learners' responses to LLSI items. They reported two distinct analyses, using both the Sri Lankan (Sinhalese, Tamil, Buddhist) and Japanese samples. In order to identify a statistically viable subset of the 63 items, data sets from the two samples were analysed from differing starting points. For the Sinhalese sample, the aim of re-analysis was to identify a statistically viable subset regardless of conceptually defined strategies, strategy subtypes, and contexts. The Sinhalese starting point was defined by the outcomes of an iterative series of exploratory factor analyses, with the resulting factor structure used as the starting point for confirmatory factor analysis (CFA). In the case of the Japanese sample, the starting point was taken to be the conceptual model of three types of learning strategies (i.e., metacognitive, cognitive, and social-affective), and this model was refined iteratively. That is, the theorised three-strategy classification of the inventory was maintained as the standard, conceptually bound model against which each item was tested (i.e., by CFA). The findings from these analyses indicated that a priori allocation of items to strategies and categories of strategies, while conceptually appealing, could be empirically blinding.

These two analyses resulted in CFA models with a range of statistically acceptable goodness-of-fit estimates. Using the Sinhalese subset as a starting point resulted in a 2-factor 9-item structural equation model (SEM). Methodological limitations of this study included the necessary exclusion from the modelling process of a number of linearly dependent variables in the Sri Lankan dataset. However, this early decision did not prevent the use of the Sri Lankan dataset to identify a two-factor model for learning
strategies (Grimbeek et al., 2005). Although this model did not adhere strictly to the conceptual presuppositions regarding the relationships between items, contexts, and types of strategies, it did constitute a short test of two types of learning strategies (loosely described as "Looking ahead" and "Checking up") that were shown to have application across the four ethnic subgroups in this study.

Using the Japanese sample as a starting point resulted in a 7-item SEM with a single factor. Specifically, the items loading this factor could be described in terms of cognitive strategies of learning language. Thus, the use of the Japanese sample to identify a single-factor solution provided support for a second short test of learning strategies that again was shown to have application across the four ethnicities tested by Liyanage. Although this single-factor solution did not take account of context, it adhered to the conceptual presuppositions regarding items and strategy types. Model comparisons via the chi-square difference test indicated that, statistically speaking, the single-factor 7-item model was significantly better than the 2-factor model. For that reason, the present examination of these data has focused on outcomes based on the Japanese sample.

**Aims**

Theoretical and methodological issues in test construction made it worthwhile to undertake further work in order to determine whether the inventory could be improved. Specifically, the present study aims to extend the Grimbeek et al. (2005) re-analysis of the Japanese data set by comparing the results of content analysis of the 63 items in the LLSI with an abridged report of the factor analysis based on the responses of Japanese students learning English. The question addressed via these analyses was the extent to which the somewhat elaborate information-processing framework of scenarios, strategy types, and strategies of LLS measurement employed in the LLSI design would be reflected in either the qualitative item analyses or the quantitative confirmatory factor analyses.

**Methods**

**Sample**

This study revisited samples collected in two studies (Liyanage, 2004; Liyanage et al., 2004). In those studies, participants completed the 63-item LLSI questionnaire as part of a larger research agenda. In the first study, Liyanage (2004) surveyed an approximately equal numbers of males and females in each subgroup: Sinhalese (M = 150; F = 151), Tamil (M = 145; F = 138) and Muslim (M = 141; F = 158). In the second study of a smaller Japanese sample (N = 179), the responses of Japanese students (n = 134; M = 58; F = 76) excluded adult data from Japanese teachers (n = 38) who accompanied them to Australia.

**Instrument characteristics**

The 63-item LLSI scale was arranged in five subscales associated with distinct scenarios (e.g., "The teacher has assigned a short composition or paragraph to be written entirely in English. This might be to write a report or to describe a picture or a personal experience") and with items rated on a 4-point Likert response category scale (between
the extremes of "Almost never true of you" and "Almost always true of you"). The five scenario-based subscales were labelled Listening in class (14 items), Speaking in class (12 items), Listening and speaking outside of class (9 items), Reading English (16 items), and Writing in English (12 items). It has been noted that this arrangement of the scale allocated a disproportionate numbers of items and strategies relative to the three types of strategies measured within these scenario-based settings.

**Content analysis and results**

Analysis of the item descriptors was the first method used to probe what was being measured in the LLSI. Smith and Humphries (in press) have demonstrated the utility of Leximancer for textual analysis. The highly automated Leximancer software produces readily interpretable output from default settings. Thus, this kind of analysis is particularly beneficial in self-assessment of activities such as the writing of item descriptions, because it provides a fairly unbiased and objective method of reviewing complex text and a clear process of justifying decisions about text selection.

In Leximancer, content analysis commences with a conceptual analysis (thematic analysis) in which the program detects and quantifies predefined concepts within the text. Content analysis continues into a relational analysis (semantic analysis) in which the program quantifies relationships between identified concepts within text (via a stochastic exercise in cognitive mapping from the multidimensional concept array to a 2-dimensional map). The reporting of results utilises a series of descriptive analyses, mostly immediately available through the medium of the concept map. A concept map (with various optional versions of that map resulting from analytic decisions) provides a visual summary of concepts and their co-occurrences.

During Leximancer’s initial conceptual phase, scanning of the text identifies frequently used terms (concept seeds) from which Leximancer generates a thesaurus of terms. Scanning also identifies names (e.g., start-of-sentence). It excludes nonlexical and weak semantic information (e.g., 9, &) and nontextual material such as menus. As part of this phase of content analysis, Leximancer identifies frequently used terms around which other terms cluster. For example, "fleas" and "bite" cluster around "dog", "hound", and "puppy." An iterative process ensues in which some of the potential concepts are eliminated. This process converges on a stable state containing most highly relevant concepts, in which a shortlist of concepts is defined by a longer list of thesaurus terms (weighted for relative closeness to the concept).

In the second phase of Leximancer, relational analysis measures the co-occurrence of concepts within text. It does so by specifying a set length of words or sentences (called a window). The program moves this window sequentially through text, noting co-occurring concepts (usually in discrete, nonoverlapping three-sentence blocks). Results are stored in the co-occurrence matrix, which stores the frequency of co-occurrence of all concepts against all others. The results of this analysis can be accessed via a spreadsheet. The final step of relational analysis represents information visually for comparison via a 2-dimensional concept map.

With the concept map, Leximancer can provide information about the results of the content analysis in a number of ways, accessible via three slide bars (i.e., concepts, theme,
The concepts slide bar allows the viewer to vary the number of visible concepts to show only the most frequent or to include, by incremental adjustments of the slide bar, those less or least frequently used. The rotation slide bar allows the viewer to rotate the array of visible concepts to optimise their interpretability, usually by aligning one of the most frequent—and conceptually relevant—concepts with the horizontal or vertical axis. The theme slide bar allows one to identify what might be described as prototypical concepts (i.e., highly frequent concepts around which others cluster). The size of thematic circles can be varied from (a) a minimum setting (Figure 1; only concepts visible); through (b) intermediate settings where circles identify a number of locally distinct concepts (Figure 2, left hand); through to (c) a maximum setting in which one or two thematic circles encompass all concepts (Figure 2, right hand).

Figures 1 and 2 illustrate the concept mapping of LLSI text analysis. Smith and Humphries (2005) outlined guidelines for interpreting the visual display in a concept map, with high frequency concepts shown as brighter and larger dots and concepts with stronger relationships positioned more closely. Thematic circles around clusters of related concepts can be added (with the theme slide bar) in order to mark graphic boundaries (see Figure 2). The most frequent concept was "words", which was aligned along the horizontal axis by rotation. In Figure 1, sentence blocks and threshold for learning were set to a single sentence. Number of concepts was set to 30. A few concepts (i.e., "word-words", "speak-talk", and "read-reading" were merged, and "give" was excluded. The theme setting was set at minimum in Figure 1, so that the array displayed a global array of prototypical concepts and their co-occurrences.

The first approximation arising from Figure 1 was that concepts emergent from the 63-item LLSI could be examined relative to the horizontally aligned concept words, by using clusters of concepts in the four quadrants as a guide to analysis. In those terms, this highly frequent concept co-occurred with first language familiarity ("hear", "Sinhala", "write", "dictionary": Top left quadrant), reading for meaning ("read", "plan", "passage", "remember", "meaning": Bottom right hand quadrant); contexts for language use ("written", "material", "questions", "class": Bottom right hand quadrant), and the formalities of second language use ("understand", "speak", "English", "teacher" "help", "rules, grammar": Top right hand quadrant).
Figure 1.
Concept map for LSS items, with all concepts shown and with the most frequent concept of "words" aligned with the horizontal axis.
This preliminary understanding was supplemented by examining concepts with thematic qualities (i.e., nominated as thematic labels, which are superimposed upon the relational analysis). This analysis maintained the previous settings for sentence blocks and threshold for learning, number of concepts, and mergers and exclusions. In addition, two different thematic settings reflected the most encompassing theme (left hand) and a set of intermediate (right hand) themes. Figure 2 (left hand) shows the single most encompassing theme to be "words." That is, the concept of words was central to the entire array of concepts plotted in Figure 2. Figure 2 (right hand) shows that this central theme was qualified with reference to seven locally distinct subthemes (right hand) related to "read", "material" (written, material), "mentally", "help", "dictionary", and, to a lesser extent indicated by spatial separation, "think."

Each of these seven subthemes characterised a specific concept. For example, the concept associated with "read" (reading passages, reading words, etc.) was used most frequently (77%, or n = 17 out of 22 co-occurrences with words). Concepts associated with other themes were less frequently in use. For example, words co-occurred four times with the concept, "mentally" (viz., mental practice) and the concept, "help" (e.g., getting help from rules [grammar], classmates, or words [phrases]). Similarly, words co-occurred three times, respectively, with "material" (going over [written] material) and with "dictionary" (using a dictionary to understand words, etc.). The concept, "think" (thinking about sentences, words, and situations) also co-occurred with words four times, although the word and think themes did not overlap.

In the inventory, therefore, item descriptions shared a common focus on processing of words. The subthemes appeared to be relatively unrelated to each other. That is, each
subtheme displayed links to the central *words* theme, but concepts within each subtheme displayed relatively few co-occurrences with concepts from other subthemes. To summarise this content analysis, the 63 LLSI items focused on speaking words (in English, in class, with a teacher) and reading words (word, phrases), with additional concerns about written words, mentally practising words, getting help with words, using the dictionary to look up words, and thinking about words.

**Factor analysis and results**

It has been noted that exploratory (SPSS) and confirmatory (AMOS) factor analyses were used to analyse responses to items from two samples (Sri Lankan students, Japanese students) as a supplement to the results of the LLSI content analysis reported in this study. The statistical software used to conduct these analyses was the Statistical Package for the Social Sciences (SPSS) and Analysis of Moment Structures (AMOS), both available from http://www.spss.com/. In this abridged report of how Grimbeek et al. (2005) analysed the responses of 141 Japanese student participants to the 63-item version of the LLSI questionnaire, these data were entered into a confirmatory factor analysis (CFA) that tested the theorised framework embedded in the instrument of three strategy types directly. Items were screened and excluded on the basis of (a) nonsignificant loadings on factors, (b) highly correlated errors and (c) significant cross-factor redundancy, with particular emphasis on this last criterion for item exclusion. That is, the preferential basis for exclusion of items from the model was the significant correlation of an item with either (a) other items or (b) the actual latent variables from one or more of the other subscales (see Table 1).

An iterative series of CFAs resulted in a 7-item single-factor solution based exclusively on items from the Cognitive Strategies subscale (see Table 1). It is important to note that this approach used only one of the four subsets of data to develop an acceptable model. Model comparisons then extended this analytic process by examining the extent to which the 7-item model based on the Japanese sample maintained the viability of its factor structure across three other ethnicities in the Sri Lankan sample (Sinhalese, Tamil, Muslim). The CFA was found to be equally statistically acceptable across these four ethnic subgroups.

**Table 1**

<table>
<thead>
<tr>
<th>Standard 7-item single-factor (Cognitive Strategies) solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>I guess at the meaning of unfamiliar words by using my knowledge of prefixes and suffixes.</td>
</tr>
<tr>
<td>After I listen, I try to summarise mentally what the teacher says to understand it better.</td>
</tr>
<tr>
<td>If I don’t completely understand what the other person says to me, I think about the words I did understand and try to guess what he or she might be saying.</td>
</tr>
<tr>
<td>I take notes when I read, listing the new words or phrases I find in the passage.</td>
</tr>
<tr>
<td>When I read new words, I try to think of what other situations they might be used in.</td>
</tr>
<tr>
<td>I use a monolingual (English-English) dictionary to understand other meanings of the words I read.</td>
</tr>
<tr>
<td>When I write, I replace words and phrases that I can’t recall with other words or phrases that have the same meaning.</td>
</tr>
</tbody>
</table>
For each of the four subsamples, SPSS AMOS provided a range of goodness of fit estimates (Byrne, 2001). These estimates include chi-square; chi-square corrected for model complexity ($\chi^2/df$; ideal 0-0.30); estimates of residual variance (root mean square residual [RMR]; root mean square error of approximation [RMSEA], where, in all cases, the ideal = <0.05); model comparisons with the baseline independence model (normed fit index [NFI], Tucker-Lewis index [TLI], comparative fit index [CFI], the relative fit index [RFI], with the ideal for each of these = >0.90); and a subset of other goodness-of-fit estimates (goodness of fit index [GFI] and adjusted goodness of fit index [AGFI], where again the ideal = >0.90).

Table 2

<table>
<thead>
<tr>
<th>Measure</th>
<th>Japanese</th>
<th>Sinhalese</th>
<th>Tamil</th>
<th>Muslim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-square ($\chi^2$)</td>
<td>16.802</td>
<td>41.855</td>
<td>40.370</td>
<td>11.472</td>
</tr>
<tr>
<td>df</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Probability</td>
<td>0.267</td>
<td>0.000</td>
<td>0.000</td>
<td>0.649</td>
</tr>
<tr>
<td>($\chi^2$)/df</td>
<td>1.200</td>
<td>2.990</td>
<td>2.884</td>
<td>0.819</td>
</tr>
<tr>
<td>RMR</td>
<td>0.046</td>
<td>0.047</td>
<td>0.042</td>
<td>0.022</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.039</td>
<td>0.081</td>
<td>0.082</td>
<td>0.000</td>
</tr>
<tr>
<td>NFI</td>
<td>0.860</td>
<td>0.740</td>
<td>0.833</td>
<td>0.919</td>
</tr>
<tr>
<td>RFI</td>
<td>0.790</td>
<td>0.610</td>
<td>0.750</td>
<td>0.879</td>
</tr>
<tr>
<td>TLI</td>
<td>0.958</td>
<td>0.702</td>
<td>0.884</td>
<td>1.031</td>
</tr>
<tr>
<td>CFI</td>
<td>0.972</td>
<td>0.801</td>
<td>0.881</td>
<td>1.000</td>
</tr>
<tr>
<td>GFI</td>
<td>0.966</td>
<td>0.960</td>
<td>0.960</td>
<td>0.989</td>
</tr>
<tr>
<td>AGFI</td>
<td>0.932</td>
<td>0.920</td>
<td>0.920</td>
<td>0.978</td>
</tr>
</tbody>
</table>

Note. Bolded numbers represent poor fit estimates between the tested model and uncorrelated latent variables model (i.e., unrelated items).

Examination of values listed in Table 2 indicated that the various goodness of fit estimates were statistically acceptable for most estimates across the four samples, with the exception of estimates comparing the tested model to the baseline model (i.e., NFI, RFI, TLI, CFI), where the baseline or independence model assumes that items are not associated with specific latent variables. That is, estimates of goodness of fit were generally statistically acceptable for the chi-square value adjusted for model complexity, estimates of residual variance, and a subset of more general goodness of fit estimates (GFI, AGFI). It is noted that, while the model was based on the Japanese sample the Muslim sample supported this model even more strongly with all except one estimate (RFI) meeting threshold statistical requirements.

Discussion

The present study set out to examine the extent to which a somewhat elaborate categorical framework of scenarios, strategy types, and strategies in LLSI measurement are reflected in the outcomes of qualitative item analyses and quantitative item response analyses. Based on these outcomes, it seems clear that the complex layering of substrategies inherent in the framework is not viable and that the notion of three
strategies is not supported. The single-factor model reported here adheres to the basic cognitive presuppositions about learner processing that underpinned the original design of LLSI items and strategy types, applies across a range of samples, and, most important, aligns with the outcomes of content analysis. Significantly, quantitative and qualitative approaches to analysis converge with one another but diverge from the complex conceptualisation of strategy operations initially envisaged. The simple model of language learning that has been identified can be used to assess a strategic approach to learning English, and it is readily amenable to further testing.

### Table 3
Standard 7-item single-factor (Cognitive Strategies) solution with highlighted text from content analysis

| I guess at the meaning of unfamiliar words by using my knowledge of prefixes and suffixes. |
| After I listen, I try to summarise mentally what the teacher says to understand it better. |
| If I don't completely understand what the other person says to me, I think about the words I did understand and try to guess what he or she might be saying. |
| I take notes when I read, listing the new words or phrases I find in the passage. |
| When I read new words, I try to think of what other situations they might be used in. |
| I use a monolingual (English-English) dictionary to understand other meanings of the words I read. |
| When I write, I replace words and phrases that I can't recall with other words or phrases that have the same meaning. |

A number of common threads (themes) emerge from this juxtaposition of the outcomes of content analysis based on the 63-item instrument with the outcomes of factorial analysis based on the Japanese study. As shown in Table 3, items loading on the Japanese-based factor correspond almost exactly to lexical themes related to reading words, writing words, mentally practising words, getting help with words (i.e., use rules to guess), using a dictionary to look up words, and thinking about words. The parallels between LLSI items and responses to items extend further, in that strategies with fewer items (7 metacognitive and 4 social-affective items) are not represented explicitly in either the item analysis or the CFA. In short, the disproportionate representation of cognitive strategy items in the instrument itself appears to be contaminating both the availability of content and analyses based on that instrument. That is, the heuristic of availability (Tversky & Kahneman, 1973), by convention, is thought to reflect the relative salience of items irrespective of the probability of their incidence (i.e., their subjective importance). However, in this instance, content analysis of items appears to reflect the objective availability of specific concepts regardless of the type of strategy or type of scenario conceptualised as underpinning the construction of the instrument.

The convergence of outcomes from textual analysis of items and CFA analysis of responses in the present results leads to several areas of speculation about educational and psychological assessment. First, it is possible that lexical item analysis could be utilised as an essential precursor to gathering data from live participants. Second, this convergence suggests that factor analytic techniques at best measure the latent variables introduced albeit unconsciously by the writers of such instruments. Finally, the
From Conceptual Frameworks to Testable and Simplified Models

A preponderance of cognitive measures as reflected via item and factorial analysis leaves open to further investigation the issue of whether the enterprise of learning languages benefits from applying social and emotional strategies. The few items written about this kind of strategy do not provide sufficient opportunity to test this notion. Moreover, in order to establish whether a learning community affects how learners approach learning a language outside that community (whether second, foreign, or dominant), linguistic issues in translation into different language and cultural issues in the communal values about learning may require closer scrutiny of a much wider range of item testing.

Previous research, including work by Liyanage (2004) and Liyanage et al. (2004), has relied on broad types of strategies with complex links via strategies to items and with disproportionate ratio of items-strategies to the three proposed types of strategies. Based on the outcomes of two factor analytic re-analyses of LLSI data reported here and in Grimbeek et al. (2005), this complex set of linkages can be collapsed into greatly simplified models that focus, at most, on one or two strategies. Hence, what this study has demonstrated clearly is that a priori allocation of items to learners’ preference for particular information-processing strategies and types of strategies can blind the LLSI test user to the actual meaningfulness of the strategies apparently being measured. Language researchers using educational tests and inventories are recommended to investigate and establish the content and statistical properties of tests that claim to measure learning strategies prior to their application to differential assessment of strategies and, hence, to differential planning of educational intervention to improve strategies.

One probably culturally defined aspect of the use of this instrument with Sri Lankan and Japanese students is that participant responses evidenced a bias towards positive responses. Similar biases have been noted in responses of research participants from mid-Eastern countries on various tests (P. Grimbeek, personal communication May 10, 2005). Participants in the LLSI studies tended to avoid the "never" response category for the 47 items remaining after excluding 26 items on the basis of linear dependency to the extent that only 2-10% of participants used this category (reversed for some items so that "always" category not used frequently) for more than half (53%, n = 25) of these items. The net effect of this tendency towards polite agreement was to contract the number of response categories in use by 90% of participants from four to three.

This cultural contraction of the range of response categories is troubling in that the four-response category scale in use in the LLSI is already at the minimum in terms of distributional reliability. Byrne (2001) indicated that further minimising the number of response categories increases the likelihood of high levels of skew or kurtosis (= >1). It follows that recommendation for future applications of this version of the LLSI instrument in settings such as Sri Lanka (and indeed other tests in these settings) must include both the use of more response categories and the selection of a set of response categories more likely to be used in full.

However, the most startling conclusion to be drawn from this study is that an internally focused content analysis of items has the capacity to reach an equivalent conclusion about the functionality of theoretically framed items and scales as would result from factor analytic approaches based on large samples. It follows that, at the very
least, content analysis has the capacity to enhance the process of item, scale, and instrument construction. In short, the habitual practice of diagnostic screening could gainfully be extended from quantitative to qualitative analysis. It has been shown that diagnostic screening of response measurement properties (Bond & Fox, 2001), with the aim of identifying either poorly ordered response categories or univariate or multivariate outliers for transformation or exclusion, is an appropriate preliminary phase of factor analysis. It now appears that diagnostic screening could extend to the lexical properties of the items to which learners are expected to respond, with the aim of identifying the latent, subjective biases already present in the written items. A corollary recommendation is that it would be very worthwhile to subject a broad range of instruments currently in use to such content analyses. Based on the present outcomes, one would not be surprised if the results of such content analyses assisted in the resolution of longstanding disputes about the ideal configuration of items, scales, and instruments.

More generally, many would agree that theoretical expectations about what a survey measures can go astray because of issues related to availability (Tversky & Kahneman, 1973) and that the developmental process needs to be supplemented by some type of empirical analysis. This present paper adds to these empirical analyses by considering the content of item descriptors (input) as well as the responses derived from administration of those same items (output). This combination increases the instrument developer’s assurance that the apparent objectivity of conceptually based instrument construction will be reflected in the items themselves (as per the results of content analysis) and in participant responses (as per the results of factor analyses).

A short inventory for LLS identification can serve various purposes in second and foreign language instruction. Given that the cognitive strategy inventory comprises a small number of items, it can be used as a tool for investigating strategy preferences of students within a short time either at the beginning of a programme of study or during it as a part of a student's needs analysis. The comprehensive nature of the inventory also makes it user-friendly, manageable, and with the boundaries of the identified strategies of learning more evident.

Acknowledgment
The collection of this data set by Indika Liyanage was supervised by his thesis advisers, Dr Gary Birch and Dr Maria Dobrenov-Major. Their contributions are gratefully acknowledged.

References


Appendix

Language Learning Strategy Inventory (LLSI): Adaptation (Liyanage, 2004) of the Learning Strategy Inventory (Chamot et al., 1987)

Students respond to each item using 4-point Likert response category scale with two anchor points (1) Almost never true of you > (4) Almost always true of you.

Listening in class

*Scenario*

In a typical class period your teacher uses English to: give directions, explain new material or review old material, and to ask the class questions. Remember to draw a circle around the number that tells how often you actually do what is described in each statement below.

1. When I listen, I plan in advance to pay more attention to what the teacher is going to talk about in general than to specific words and details
2. I write down any new words, phrases or rules my teacher says so I’ll be sure to remember them.
3. I ask the teacher questions when I don’t understand what he or she is saying.
4. When I hear a new English word, I try to learn the teacher’s pronunciation by copying or imitating it.
5. When I hear a new English word that sounds like a familiar Sinhala/ Tamil word, I assume it has a similar meaning.
6. I find myself translating what the teacher says back into Sinhala/ Tamil so I can understand.
7. When I learn a new word or phrase, I play it back in my mind to remember it.
8. When listening to the teacher, I apply grammar rules to help myself understand.
9. When I hear a new word, I think of a sentence in which I might use it later.
10. When I don’t understand what the teacher says, I get help from a classmate.
11. I try to relate what I’m hearing to my own experiences or to information I already know.
12. I guess at the meaning of unfamiliar words by using my knowledge of prefixes and suffixes.
13. I pay more attention to some words and phrases than to others when the teacher is talking in English.
14. After I listen, I try to summarise mentally what the teacher says to understand it better.
Speaking In class

*Scenario*

The teacher requires class participation. This means that you have to speak English in class, including asking and answering questions, participating in oral drills, reading aloud and perhaps giving a short oral presentation. Remember to draw a circle around the letter that tells how often you actually do what is described in each statement below.

1. When the teacher calls on me in class, I plan my answer in my head before I say a word.
2. I listen carefully to what I say and correct myself when I make a mistake.
3. If I have to give a talk to the class, I give it to a friend first so he or she can tell me how it sounds.
4. If I have to give a talk to the class, I practise the talk several times paying attention to the meaning of the talk before I actually do it.
5. If I have to give a talk to the class, I mentally practise the talk before I actually do it to reduce anxiety.
6. If I can't recall a word or phrase when I speak in English, I try to use another word or phrase to replace it.
7. I think in Sinhalese/Tamil of what I want to say, and then I translate it into English.
8. When I speak, I am generally unaware of any mistakes I might be making.
9. I consciously apply the rules of grammar when I speak English.
10. I volunteer answers in class so I can practice using English.
11. I try to answer all questions mentally, even when the teacher is addressing someone else.
12. When I learn a new word, I say it in a sentence as soon as possible.

Listening and speaking outside of class

*Scenario*

You have an opportunity to speak English outside of class. For example, you meet several native speakers. Remember to draw a circle around the number that tells how often you actually do what is described in each statement below.

1. I listen especially for words or phrases that I already know to help me understand what is going on in a conversation.
2. I talk about the same sorts of things in English that I talk about in Sinhalese.
3. I ask native speakers the correct way to say things.
4. I try to talk with native speakers and keep the conversation going, because I get more practice that way.
5. If I don't completely understand what the other person says to me, I think about the words I did understand and try to guess what he or she might be saying.
6. I relate the English I hear in conversations to what I've learned in class.
7. If I don't understand what the other person says to me, I ask them to speak more slowly or to say it in a different way.
8. When I know I'm going to be around native speakers, I plan a few things to say.
9. I go home afterwards and think about what I said to see if I made any mistakes.

**Reading English**

*Scenario*

The teacher assigns a reading selection for homework. This may be a short story or an article from a newspaper, or a cultural passage. Remember to draw a circle around the letter that tells how often you actually do what is described in each statement below.

1. Before I read, I plan to pay more attention to the general meaning of the passage than to specific words, phrases and details.
2. Before I actually read (a passage or book), I arrange myself a treat to enjoy on completion of the task.
3. When I find the meaning of a new word, I read it over and over again to remember its meaning.
4. I take notes when I read, listing the new words or phrases I find in the passage.
5. I scan for special words, phrases or information to get the most important points when I read.
6. When I read, I organise information under different headings according to their attributes.
7. I try to guess the meaning of unfamiliar words by looking at the words in the rest of the sentence.
8. I get the major ideas of a reading selection by checking the comprehension questions before I begin reading.
9. When I read, I try to visualise what I read.
10. I first skim the material I must read to get the main idea and concepts.
11. I practice my reading skills by trying to read extra materials in English (such as newspapers, magazines, ads, etc.).
12. When I read new words, I think of what other situations they might be used in.
13. I try to relate what I'm reading to my own experiences or to material I already know.
14. I use a monolingual dictionary (Eng-Eng) to understand additional meanings of the words I read.
15. After I finish reading, I check my understanding by seeing if I can remember the main ideas of the passage.

16. After I finish reading, I try to summarise mentally, what I have read to understand it better.

Writing in English

Scenario

The teacher has assigned a short composition or paragraph to be written entirely in English. This might be to write a report or to describe a picture or a personal experience. Remember to draw a circle around the letter that tells how often you actually do what is described in each statement below.

1. Before I actually do a writing task (e.g., an essay or a letter) I arrange myself a treat to enjoy on completion of the task.

2. I use what I know about writing in Sinhala (structure, organization, etc.) to help write in English.

3. Before I write the actual assignment, I write it a few times to see whether it conveys the intended meaning.

4. When I write, I replace words and phrases that I can't recall with other words or phrases that have the same meaning.

5. I write the assignment first in Sinhala, and then translate it into English.

6. I consciously use grammatical rules when I write in English.

7. For accuracy, I ask a friend to read over what I've written.

8. I use a monolingual (English-English) dictionary or other English reference materials when I write in English.

9. I use my textbook and dictionary to look up spelling, verb conjugations, and gender agreement, etc.

10. I carefully reread what I've written to make sure there are no mistakes.

11. Before writing, I make a plan or outline of what I want to say.

12. While writing a first draft, I try to get all my ideas down instead of worrying about spelling and grammar.