Assessing the Answer Selections of SILL

Amanda Black
Northern Arizona University
Abstract

The Strategy Inventory for Language Learners is a widely employed instrument both domestically and internationally to examine the strategy use of language learners world-wide. Though this instrument has been assessed for validity and reliability several times (Oxford et al., 2014; Oxford & Burry-Stock, 1995; Russel, 2010; and Yeşilbursa & İpek found, 2012), the answer choices provided have never been closely examined. As Tourangeau et al. (2013) suggest, it is imperative for participants to understand the statements they are responding to if we as researchers are to make determinations about their results. Thus providing clear answers along with an option indicating a lack of understanding could be a useful way to obtain more reliable results. The present study thus aimed to determine whether adding clarity to the answer selections along with an “I do not understand this statement” option would impact the results on SILL. Results indicate that altering the answer choices on SILL to reflect more specific time options did not yield significantly different results; however, providing the “I do not understand this statement” did appear to have an impact.
Assessing the Answer Selections of SILL

Background

The Strategy Inventory for Language Learning (SILL), created by Rebecca Oxford in 1990, is a widely employed instrument for research on language learning strategies (LLS) (Yeşilbursa & İpek, 2013). As such, it has had a major impact on this field of research for nearly 30 years domestically as well as internationally with translations existing in Chinese, French, German, Japanese, Korean, Russian, Spanish, Thai, Ukrainian and Turkish (Demirel, 2009; Hsiao & Oxford, 2002). Researchers have relied on SILL, or some modification of it, for reasons including overall measures of language learning strategy use, determining relationships between LLS and learner beliefs, and determining relationship between LLS and language proficiency (Yeşilbursa & İpek, 2013). Therefore, SILL has been assessed several times for validity, reliability, utility, and even fakability. However, the answer choices presented in SILL are arguably vague and offer no option to demonstrate a lack of understanding leading to a gap in the analysis of the instrument. Thus the present study proposes to conduct an analysis of answer choices to determine if clarification factors will alter results.

Motivation for the present study comes first and foremost from research that has been conducted on language learning strategies, or tools that a language learner purposely relies on to actively enhance the acquisition, retrieval, and storage of the language being learned (Oxford et al., 2014). LLS first became an intriguing phenomenon in the field of language learning as a response to Krashen’s acquisition-learning hypothesis which made a distinction between acquisition and learning; Within this hypothesis is also the argument that language is unique to acquisition and thus cannot be learnt indicating that conscious learning of language strategies would yield little to no effect (Griffiths & Parr, 2001; Krashen, 1976); however, researchers such
as Rubin observed that “good learners” were relying on strategies that seemed to enhance their ability to learn language (Griffiths & Parr, 2001; Plonsky, 2011). Thus, subsequent research on LLS was dedicated to determining characteristics of “good learners” and transferring those characteristics to less successful L2 learners by means of strategy instruction (SI). This “teachability” of LLS lends to the idea that learners make deliberate choices in terms of their own learning which is in direct contrast to Krashen’s Monitor Theory and is also the reason that “contemporary educators and researchers are increasingly keen to harness the potential” of LLS (Griffiths & Parr, 2001, p. 249).

A shift in focus away from attributes of a good learner and towards the importance of LLS in the teaching and learning process soon emerged and several classifications of LLS were devised. The most notable were those by Rubin (1987), Oxford (1990), and O’Malley & Chamot (1990). The initial classification by Rubin (1987) divided LLS into the three categories of direct, communication and social strategies which centered around strategies directly involved with learning the language, strategies to communicate a message, and strategies associated with helpful behaviors for learning the language respectively. The classification by O’Malley & Chamot included metacognitive strategies involving higher mental functions such as evaluation and planning, cognitive strategies involving direct manipulation of the information being learnt, and social/affective strategies involving emotional control and interaction behaviors. Oxford’s (1990) is arguably the most definitive with two overarching categories, direct and indirect, and six subdivisions: memory, cognitive, metacognitive, compensatory, social, and affective (Yeşilbursa & İpek, 2013). These divisions are similar to those described above with the exception of memory and compensation strategies which are unique to Oxford’s classification.
Once these taxonomies were devised, ideology about LLS evolved to include self-regulation and learner autonomy with proponents in the field of educational psychology suggesting that learning strategies promote self-regulatory behaviors (Oxford et al., 2014). Thus the view that learning strategies play a prominent role in learner self-management and involves a deep connection to the learner’s beliefs and understanding of the information being learned became prominently liked with research on LLS (Rubin, 2001).

Given the prominent role of LLS in the fields of applied linguistics and educational psychology, much research has been conducted on the subject. Several studies have examined the effect of learning strategies on listening and reading comprehension (Anderson, 1991; O’Malley et al., 1989), have assessed strategy use (Chamot, 2007; Oxford 2011 b), have examined frequency of strategy use (Ungureanu & Georgescu, 2012), and have examined variables related to LLS (Griffiths, 2008; Oxford, 2011; Oxford et al., 2014; Oxford &Nyikos, 1989; Salashour et al., 2013). While there are many instruments that can be utilized in research examining LLS, a good majority of major studies have relied on the Strategy Inventory for Language Learning (SILL) which is explained below (Oxford & Burry-stock, 1995).

**The Strategy Inventory for Language Learning.**

The structure of the original instrument is based on a division of LLSs into two overarching categories: direct and indirect (Oxford, 1990). These two classes are subsequently divided into three subcategories each. The direct category is divided into memory, cognitive, and compensation strategies whereas the indirect category is divided into metacognitive, affective, and social strategies. These six categories are presented to learners as alphabetical sections A-F with a differing number of questions for each section. The total survey is comprised of 80 questions with sections B (cognitive strategies), D (metacognitive strategies), and A (memory
Answer Selection of SILL

strategies) containing the highest counts of 25, 16, and 15 questions respectively. Sections C (compensation strategies), E (affective strategies) and F (social strategies) contain 8, 7, and 9 questions respectively.

For all questions, participants are asked to, “mark the response (1, 2, 3, 4, or 5) that tells how true the statement is in terms of what you actually do when learning the new language” (Oxford, 1990). Participants are then provided with the following list of options: 1. Never or almost never true of me; 2. Generally not true of me; 3. Somewhat true of me; 4. Generally true of me; 5. Always or almost always true of me and asked to mark each statement accordingly. Thus, the language learning strategies are measured on a five point Likert ordinal scale.

SILL is an instrument constructed to measure language learning strategies or, “operations used by learners to aid the acquisition, storage, and retrieval of information” (Oxford & Nyikos, 1989). According to Oxford (1990), it specifically measures cognitive, memory, metacognitive, social, affective, and compensation strategies and systematically covers the four language skill areas of listening, reading, speaking, and writing. The goal is to determine which strategies learners rely on as evidence suggests that “good” learners depend on strategies appropriate to their own stage, context, and purpose for learning the language (Plonsky, 2011). The theoretical underpinnings of this instrument are derived from Oxford’s own conceptual framework of language learning strategies which define LLS as direct, or those strategies directly involved with the language being learned (cognitive, memory, and compensation), and indirect, or those strategies that are helpful for learning the target language (metacognitive, social, and affective) (Hsiao, & Oxford, 2002). Specifically, memory strategies are those used to remember the target language more effectively, cognitive strategies are those which require the use of mental processes, compensation strategies are ways in which learners compensate for missing
knowledge, metacognitive strategies are those used for organizing and evaluating learning, affective strategies are used to manage emotions, and social strategies are those used when learning with others (Russel, 2010). The division of LLS into such precise factors has led to its wide application in many research studies domestically as well as internationally. As such, several studies have also assessed the instrument to ensure its validity and reliability.

**Assessing SILL**

According to Oxford and Burrystock (1995), “There exists evidence that endorses both the reliability of SILL and its construct, content, concurrent, and predictive validity.” In the mid-nineties it was estimated that more than 50 studies had depended on SILL to conduct their analyses with more than 8000 participants completing the questionnaire; now, more than 20 years later, SILL is still the most widely used strategy instrument and appears to be the only language learning strategy instrument that has been extensively checked for reliability and validated in multiple ways. (Oxford et al., 2014; Oxford & Burry-Stock, 1995). In particular it has been assessed for utility, reliability, validity, and underlying factorial structure.

An assessment conducted by Oxford & Burrystock (1995) determined that SILL has utility based on “people around the world who have employed it”; is reliable based on many Cronbach Alpha scores above .91 for the English version and above .85 for translations; is valid in terms of construct, criterion, and fakability; and is relevant as several factorial analyses revealed relevant items between 4 and 9 factors depending on context of application.

An assessment completed as a thesis by Adam Russel (2010) corroborates much of the findings reported above with statistical support of the validity and reliability of SILL. A more recent study conducted in 2012 by Yeşilbursa & İpek found a four factor model with 16 items to be useful for their purposes. However, the researchers presented a concern that many studies
which have used SILL lack a report about the validity of SILL for their specific research contexts. Additionally, though several studies have assessed SILL for validity and reliability, the majority of these studies conducted factor analyses and have never examined the effectiveness of the answer choices. Though seemingly less important than the test items, the answers are where the analyzed data are actually held; therefore, ensuring that the answers are perceived and understood as intended is imperative to the reliability and validity of the retrieved data.

The idea of answer selections being clear and understandable is supported by Jenkins & Dillman (1995) who suggest that it is imperative that participants understand the selections that they are making as results will be meaningless otherwise. Additionally, other researchers have stated that answer selections should be unambiguous to ensure reliability of responses and that survey scales need to be practitioner friendly: short, easy to administer, and straightforward to interpret (Hamre & Cappella, 2015; Kosovich, Hulleman, Barron, & Getty, 2014). Furthermore, Tourangeau et al. (2013) suggests that the inclusion of no opinion, don’t know, and N/A response anchors should be considered in order to avoid misaligning conceptual knowledge with a lack of understanding. Based on the literature above, the answer selection on SILL should examined to ensure that they are clear and understandable which is the aim of the present study.

**The Present Study**

**Pilot Study**

Due to the specific question of the present researcher and lack of previous studies on similar subject, a pilot study was conducted. The pilot study was very small and included extensive interviews and completion of different versions of SILL by two highly advanced second language learners from the MA TESL program at NAU.
The interviews revealed that the addition of “I do not understand the statement” might greatly enhance the clarity of the answers. Additionally, the interviews helped the researcher determine how changing the answer choices to be more specific could be most beneficial.

Thus, in addition to the concerns stated above, results from the pilot study further indicate that the answer selections on SILL offer no recourse for students who may not understand the statement suggesting that answers may not be truly representative of strategy use. Therefore, it would be interesting to determine how answers are impacted when they are made more precise in terms of time and clarity. Based on the literature and pilot study, one efficient way to determine clarity may be to provide a do not understand choice as this could provide valuable information for the researcher and will prevent forced selections from the participants. Thus, the aim of the present research is to modify the answer selections on SILL to offer more precise time measurements and to provide a do not understand choice for purposes of clarity.

**Research Questions.** To guide the present analyses, the researcher relied on two questions of interest:

1) Will providing specific amounts of time in answer selections for SILL effect results?
2) Will providing an “I do not understand this statement” option impact results?

**Hypotheses.** To conduct the study the researcher will be working with the hypotheses below, attempting to reject the null hypothesis and accept the alternative.

H₀: There will be no impact on responses from changing the answer choices on SILL

Hₐ: Changing the answer selections on SILL will impact results
H₀: Providing an “I do not understand this statement” option will have no effect on results.

Hₐ: Providing an “I do not understand this option” will have an effect on results.

Methods

Participants

Participants for the present study included 20 English second language learners from the PIE at Northern Arizona University and international students enrolled in English 105. 11 participants were female, and 9 were male. L1 backgrounds were 10 Arabic speakers, 7 Chinese speakers, 2 Farsi speakers, and 1 Vietnamese speaker. The proficiency level of all participants was similar and judged by participants’ placement in 105 and courses at the PIE. The placement required students to score 45-56 (TOEFL) or 5-5.5 (IELTS) which correlates to levels B1-B2 intermediate/high intermediate in the Common European Framework.

Instruments

For the purposes of the present study, two versions of SILL were relied upon. The first was the original version created by Rebecca Oxford (1990) (see Appendix A). The second (see Appendix B) was identical in all aspects to the first with the exception of the answer selections which were modified based on results from the pilot study to:

1. I never do this (Never)

2. I do this less than once a month (Almost never)

3. I do not understand the statement
4. I do this 1-3 times a month (In the middle)

5. I do this about 3 times a week (Kind of A lot)

6. I do this almost every day (A lot)

Additionally, the questionnaires were shortened based on previous factor analyses and random selection. Specifically, the researcher consulted Russel (2010), and Yeşilbursa & İpek (2012) to determine which items were most relevant based on the combined results of their respective factorial analyses. As a result, the questionnaires were shortened from 80 questions to 27 questions.

Procedure

Once participants were recruited, they were divided evenly into four groups of 5 participants each. Each group was orally read the directions and given an opportunity to ask questions for clarification in the same classroom at the PIE. After the short directions session, groups were given a paper and pencil version of the instrument and were allowed an equal amount of time (20 minutes) to complete each survey. Group one was tasked with completing survey version A; group two completed version B; group three completed versions A and then B; and group four completed versions B and then A.

Analysis

Results on each survey were scored ordinally. Version A was scored on a 1-5 scale and version B was scored on a 1-6 scale. Individual scores to each item were recorded in Excel so that a Cronbach Alpha could be run to test for the instrument’s reliability. Descriptive statistics were also run to determine the central tendency of answers for each version.
In order to norm the scales for an accurate comparison of the means, z-scores were calculated and entered into SPSS for a between group comparison to determine whether survey versions were impacted by changes in options. Once distribution was determined, a T-test was run; results are reported below.

**Results**

The descriptive statistics of mean, median, and mode were run in Excel for each version of the test. Results for version A are reported in Table 1 below.

Table 1

*Descriptive Statistics for Version A*

<table>
<thead>
<tr>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.12</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

As the results show, the mean and median are around 4 which indicates that “usually true of me” is a commonly selected answer for the original version of SILL. The mode is 5 which suggests that the answer selected with most frequency is “always almost true of me.”

Results for version B are presented in Table 2 and demonstrate that the mean and median are close to 3 which indicate that the central tendency was for students to select “I do not understand the statement.” The mode was 4 which indicates that the most frequent selection was “I do this 1-3 times a month (In the middle).”
Table 2

*Descriptive Statistics for Version B*

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.23</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

To determine reliability of the two versions of the survey, a Cronbach Alpha was run for each in SPSS. Results for version A and B are presented in tables 3 and 4 respectively.

Table 3                                                                 Table 4

*Reliability Statistics*                                                                 *Reliability Statistics*

<table>
<thead>
<tr>
<th>Cronbach's Alpha Based on Standardized Items N of Items</th>
<th>.811</th>
<th>.795</th>
<th>27</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cronbach's Alpha</td>
<td>.555</td>
<td>.546</td>
<td>27</td>
</tr>
</tbody>
</table>

As the tables show, the Cronbach Alpha for version A is .811 while for version B it is only .555. This will be explored in the discussion section below.

To equivalently compare means of the two versions, Z-scores were calculated in Excel and entered into SPSS so an independent-samples T-test could be run. Results are presented in Table 5 below. As the table indicates, there is no significant difference between the mean Z-scores with $t$=-.17 and $p$=.928.
Table 5

t-tests for Two Versions

<table>
<thead>
<tr>
<th>Zscore</th>
<th>df</th>
<th>Mean Difference</th>
<th>Std. Error Difference</th>
<th>95% Confidence Interval of the Difference</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal variances assumed</td>
<td>.017</td>
<td>28</td>
<td>.986</td>
<td>-.00638</td>
<td>.36902</td>
<td>-.76229</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>-.017</td>
<td>27.998</td>
<td>.986</td>
<td>-.00638</td>
<td>.36902</td>
<td>-.76229</td>
</tr>
</tbody>
</table>

Relevance

Research Question 1

The first research question inquired as to whether providing specific amounts of time in answer selections for SILL affected results. As the T-test revealed no significant difference between the means of the two versions, the researcher must accept the null hypothesis and conclude that changing the answers selections to include to specific amounts of time does not significantly impact the results. While this is not what the researcher expected, these results are still extremely beneficial as they suggest that SILL may also be reliable in terms of the answer choices. This could potentially add positive results to the previous literature assessing SILL and further demonstrated that as an instrument it is reliable and valid.

Research Question 2

The second research question was designed to determine whether providing an “I do not understand this statement” option impacted results on the SILL. The first important thing to examine in light of the second research question is the Cronbach Alpha scores. As results indicate, version A remained pretty reliable with .81 while version B resulted in only .55.
Initially, this would appear as if version B were less reliable; however, it is important to consider that version B contained the “I do not understand this statement” which may also indicate that the Cronbach Alpha score is as a result of participants inconsistently relying on this option. It would be helpful to know which items participants were selecting the “I do not understand this statement” for on version B as this could shed light on items commonly not understood by participants, but this is outside the scope of the present research.

Additionally, the results from the descriptive statistics help answer research question 2 in that they indicate that participants are selecting “I do not understand this statement” more often than by chance when the option is available. This could mean that there are many items on SILL that participants truly do not understand, or it could indicate that they are selecting this option as a fallback. To be sure interviews with participants would need to be conducted, but again this fell outside of the scope of the present study. Based on the results, the researcher cannot explicitly answer research question 2 as further research should be conducted; however, thus far, it would appear that the “I do not understand this statement” option does impact results. While this cannot be taken conclusively as is, it does suggest that future research should continue to examine the answer selections of SILL, at least in terms of adding a don’t know option to the mix. This is in line with the implications from Tourangeau et al. (2013) who demonstrate that the inclusion of no opinion, don’t know, and N/A responses can help avoid confusing actual data for lack of understanding.
References


Appendix A

Strategy Inventory for Language Learning

Version 7.0 (ESL/EFL)


Directions: Please read each statement and select the number that best represents how often you complete the statement. Write the number that best represents how often you complete the statement on the empty line.

1. Never or almost never true of me
2. Usually not true of me
3. Somewhat true of me
4. Usually true of me
5. Always or almost always true of me

Part A

1. I connect the sound of a new English word and an image or picture of the word to help remember the word.

   _______________________

2. I remember a new English word by making a mental picture of a situation in which the word might be used.

   _______________________

3. I use rhymes to remember new English words.

   _______________________

4. I use flashcards to remember new English words.

   _______________________

Part B

5. I say or write new English words several times.

   _______________________

6. I try to talk like native English speakers.

   _______________________

7. I practice the sounds of English.

   _______________________

8. I use the English words I know in different ways.

   _______________________
1. Never or almost never true of me

2. Usually not true of me

3. Somewhat true of me

4. Usually true of me

5. Always or almost always true of me

9. I look for words in my own language that are similar to new words in English.

10. I try to find patterns in English.

11. I find the meaning of an English word by dividing it into parts that I understand.

12. I try not to translate word-for-word.

**Part C**

13. When I can't think of a word during a conversation in English, I use gestures.

14. I make up new words if I do not know the right ones in English.

15. If I can't think of an English word, I use a word or phrase that means the same thing.

**Part D**

16. I try to find as many ways as I can to use my English.

17. I notice my English mistakes and use that information to help me do better.

18. I plan my schedule so I will have enough time to study English.

19. I look for people I can talk to in English.

20. I look for opportunities to read as much as possible in English.

21. I have clear goals for improving my English skills.
1. Never or almost never true of me
2. Usually not true of me
3. Somewhat true of me
4. Usually true of me
5. Always or almost always true of me

Part E

22. I try to relax whenever I feel afraid of using English.

23. I give myself a reward or treat when I do well in English.

24. I write down my feelings in a language learning diary.

Part F

25. If I do not understand something in English, I ask the other person to slow down or say it again.

26. I ask English speakers to correct me when I talk.

27. I ask for help from English speakers.
Appendix B

Strategy Inventory for Language Learning

Version 7.0 (ESL/EFL)


Directions: Please read each statement and select the number that best represents how often you complete the statement. Write the number that best represents how often you complete the statement on the empty line.

1. I never do this (Never)
2. I do this less than once a month (almost never)
3. I do not understand the statement
4. I do this 1-3 times a month (in the middle)
5. I do this about 3 times a week (Kind of A lot)
6. I do this almost every day (A lot)

Part A

1. I connect the sound of a new English word and an image or picture of the word to help remember the word.

2. I remember a new English word by making a mental picture of a situation in which the word might be used.

3. I use rhymes to remember new English words.

4. I use flashcards to remember new English words.

Part B

5. I say or write new English words several times.

6. I try to talk like native English speakers.

7. I practice the sounds of English.

8. I use the English words I know in different ways.
9. I look for words in my own language that are similar to new words in English.
10. I try to find patterns in English.
11. I find the meaning of an English word by dividing it into parts that I understand.
12. I try not to translate word-for-word.

1. I never do this (Never)
2. I do this less than once a month (almost never)
3. I do not understand the statement
4. I do this 1-3 times a month (in the middle)
5. I do this about 3 times a week (Kind of A lot)
6. I do this almost every day (A lot)

Part C

13. When I can't think of a word during a conversation in English, I use gestures.
14. I make up new words if I do not know the right ones in English.
15. If I can't think of an English word, I use a word or phrase that means the same thing.

Part D

16. I try to find as many ways as I can to use my English.
17. I notice my English mistakes and use that information to help me do better.
18. I plan my schedule so I will have enough time to study English.
19. I look for people I can talk to in English.
20. I look for opportunities to read as much as possible in English.
21. I have clear goals for improving my English skills.
1. I never do this (Never)
2. I do this less than once a month (almost never)
3. I do not understand the statement
4. I do this 1-3 times a month (in the middle)
5. I do this about 3 times a week (Kind of A lot)
6. I do this almost every day (A lot)

Part E

22. I try to relax whenever I feel afraid of using English. __________
23. I give myself a reward or treat when I do well in English. __________
24. I write down my feelings in a language learning diary. __________

Part F

25. If I do not understand something in English, I ask the other person to slow down or say it again. __________
26. I ask English speakers to correct me when I talk. __________
27. I ask for help from English speakers. __________