



OPEN ACCESS

EDITED BY

G. Sue Kasun,
Georgia State University, United States

REVIEWED BY

Yupika Maryansyah,
Universitas Muhammadiyah Bengkulu,
Indonesia

*CORRESPONDENCE

Akihiro Saito
✉ aki.saito@rs.tus.ac.jp

RECEIVED 15 December 2025

REVISED 18 February 2026

ACCEPTED 05 March 2026

PUBLISHED 24 March 2026

CORRECTED 31 March 2026

CITATION

Saito A (2026) Dataset for the adapted
Strategy Inventory for Language
Learning.
Front. Educ. 11:1768093.
doi: 10.3389/feduc.2026.1768093

COPYRIGHT

© 2026 Saito. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](#). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these
terms.

Dataset for the adapted Strategy Inventory for Language Learning

Akihiro Saito*

Institute of Arts and Sciences, Tokyo University of Science, Noda, Japan

This Data Report documents a dataset that was collected for the purposes of adapting and validating the SILL–JP/STEM instrument, a Japanese-language version of Oxford’s Strategy Inventory for Language Learning (SILL) that has been validated in a sample of science and engineering students. The original 50-item SILL was adapted through collaborative expert review, resulting in an item pool that was administered online to undergraduate science and technology students. Following a two-stage psychometric analysis, this 50-item pool was reduced to a 22-item instrument featuring six strategy subscales. The publicly archived dataset, drawn from a combined sample of 599 students enrolled in English as a foreign language courses, contains responses to the adapted 50 items. The survey was distributed via Google Forms, and responses were mandatory, ensuring that no substantive item responses were missing. The dataset is deposited in the Zenodo repository: <https://doi.org/10.5281/zenodo.18678940>.

KEYWORDS

English as a foreign language (EFL), instrument adaptation, Japanese learners, language learning strategies, scale validation, strategy inventory for language learning (SILL)

1 Introduction

Research on language learning strategies (LLSs) has played a central role in efforts to understand why some learners appear to be more adept at learning languages than others (Griffiths, 2018). Early work has suggested that proficient learners tend to flexibly deploy a diverse repertoire of strategies in response to task demands, while later developments in educational psychology and applied linguistics have linked such behaviours to elements of self-regulation, autonomy and agency (Thomas and Rose, 2019). While the precise definition and classification of LLSs have been debated (e.g., Hsiao, 1997; Oxford and Cohen, 1992), Oxford’s (1990) Strategy Inventory for Language Learning (SILL) emerged as one of the most widely used instruments in the field. The SILL distinguishes between six strategy domains—memory, cognitive, compensation, metacognitive, affective and social strategies—and has been employed across a wide range of contexts.

Notably, however, some researchers have raised concerns about the factorial structure and contextual appropriateness of the original SILL in particular educational settings, such as Japanese university English-as-a-foreign-language (EFL) classrooms (Robson and Midorikawa, 2001). Prior work has shown that the six-factor structure is not always cleanly replicated and that specific items either may be difficult for learners to interpret or may not reflect their actual strategy usage. Such concerns, alongside broader calls for the context-sensitive adaptation of major second-language (L2) instruments (Rose et al., 2018), have motivated several SILL adaptation and shortening projects across different countries and education levels (e.g., Ardasheva and Tretter, 2013; Danko and Dečman, 2019; Guillemain et al., 1993; Saks and Leijen, 2018).

The dataset covered by this report underpins a study that was conducted with Japanese university students in the science and engineering disciplines. The

associated research, reported by Saito (2025), followed a two-stage protocol inspired by Ardasheva and Tretter (2013), who adapted, calibrated and subsequently validated an initial item pool based on the original SILL, resulting in a reduced set of items. During Stage 1, as described by Saito et al. (2023), a contextually adapted 50-item Japanese SILL was administered. This was then reduced to a 22-item model through a series of exploratory and confirmatory factor analyses. During Stage 2, as reported by Saito (2025), this 22-item model was evaluated in a validation sample as well as a combined dataset, ultimately yielding the SILL-JP/STEM instrument.

This Data Report centres not on the analytical results but on the dataset itself. It documents how the data were collected, which variables are available, how the data were cleaned and prepared for archiving, and how they may be reused moving forward. The dataset has been deposited in Zenodo as a fixed, open resource. By making these data publicly available and clearly documented, the report aims to support transparency, replication and further research on LLSs in EFL contexts, both within the disciplines of science and engineering and in other fields of study.

2 Methods

2.1 Research context and design

The data were collected at two private four-year universities in Japan. Both institutions offer science and engineering degree programmes, with English as a required component of the undergraduate curriculum. Students in these programmes typically complete at least eight credits of EFL courses during their first two years. The overall research design followed a two-stage protocol. During Stage 1, an expert panel adapted the original 50-item SILL into Japanese and administered it to a calibration sample ($n = 271$) (Sample A). Based on this calibration sample, a series of exploratory and confirmatory factor analyses were conducted, leading to a 22-item model that replicated the six-strategy taxonomy in a form more suitable for Japanese learners of English. During Stage 2, this 22-item model was subjected to further confirmatory analyses using a validation sample ($n = 328$) (Sample B) and, later, the combined data from Samples A and B ($N = 599$). The dataset archived in Zenodo corresponds to the combined sample from Stages 1 and 2, featuring responses to the 50-item adapted SILL items used across both data-collection phases. Formal adaptation and psychometric validation procedures and results derived from these data are reported in detail elsewhere (Saito et al., 2023; Saito, 2025). The dataset should be interpreted as context-bound to Japanese undergraduate students in science and engineering disciplines enrolled in required EFL courses at private universities. While the instrument draws on a popular framework, secondary users should exercise caution in generalising findings outside comparable educational and disciplinary settings.

2.2 Participants and sampling

Participants were recruited using convenience sampling. The combined sample consisted of 599 undergraduate EFL learners

enrolled in EFL courses at the two institutions. All participants were science and engineering students taking EFL classes taught by the author and their collaborators (i.e., Saito et al., 2023). Although these students belonged to a diverse array of science and engineering programs, including mathematics, physics, life sciences, pharmaceutical sciences, mechanical engineering, electrical and electronic engineering, information science, civil and architectural engineering, and design, no information on their specific majors was collected. As participants were recruited from courses taught by the author and collaborators, it is possible that perceived instructor presence may have influenced response tendencies, despite explicit assurances of anonymity and voluntary participation. This should be considered when interpreting the dataset.

2.3 Instrumentation and adaptation process

The SILL was adapted by a three-person panel consisting of the author and two colleagues, both of whom held teaching positions at the participating universities. One collaborator was a former public-school English teacher and practising interpreter, while the other was a Japanese linguist. The panel first translated Oxford's (1990) SILL version 7.0 from English into Japanese. The Japanese version was then back-translated into English by a professional translation service. Discrepancies between the original and back-translated versions were discussed and resolved by the panel, giving way to a Japanese SILL that was linguistically equivalent to the source instrument.

Following the translation and back-translation, the panel examined the item content for contextual fit. Thirty of the 50 original items were overwritten or substantially rephrased to better reflect the learning behaviours and norms of Japanese university EFL contexts (see Saito, 2025; Saito et al., 2023). For example, the original item "I physically act out new English words" was rephrased to reflect more commonly reported strategies such as repeated contextual usage. Similarly, references to maintaining an English learning diary were adapted to reflect textbook review and note consolidation practices common in Japanese university contexts. These adaptations were intended to retain the original conceptual domains of the SILL while making individual items more meaningful and interpretable for this particular population.

Through the calibration analyses described by Saito et al. (2023), a 22-item model with six strategy subscales emerged, and this abbreviated instrument was subsequently validated in later work (Saito, 2025). The present dataset, however, preserves item-level responses to all 50 adapted items, thereby enabling replication of the published reduction as well as alternative model testing and revalidation (see Table 1 for an overview of strategy domains and item counts).

2.4 Data-collection procedures and period

The data were collected using Google Forms over two administration periods: December 2022 – January 2023, and July 2023 – August 2023. For each period, students were first

TABLE 1 Strategy domains and item counts.

Strategy domain	Code prefix	Items in adapted pool	Items in final validated scale
Memory strategies	MM	8	3
Cognitive strategies	CG	9	4
Compensation strategies	CM	6	3
Metacognitive strategies	MC	10	6
Affective strategies	AF	9	3
Social strategies	SC	8	3
Total		50	22

informed of the study's objectives during classes taught by the author and their collaborators as a means of promoting voluntary participation. The questionnaire link was then provided through the university's learning management system. Students completed the online questionnaire outside of regular class hours. Participation was voluntary, anonymous and unrelated to course assessment, and the students were assured that they could withdraw in the middle of the survey if they so desired without consequence. The Google Form required participants to read an information statement and provide informed consent before proceeding. All items were set as mandatory, ensuring that the dataset would feature no missing responses. Completing the survey typically took 10–20 min.

2.5 Data cleaning and preparation for archiving

As all survey items were set as mandatory on Google Forms, the raw data for the strategy items were complete for all respondents. The dataset archived in Zenodo features responses to all 50 adapted SILL items, including those that were later excluded during the validation process. Although the published instrument was ultimately a 22-item model, all originally administered items are preserved in the public dataset to ensure transparency and facilitate any potential secondary analyses or revalidation work. The dataset does not contain any other data, such as additional demographic or proficiency variables.

The cleaned dataset was exported from Google Forms into a Microsoft Excel (.xlsx) file. An accompanying README file was prepared to document the dataset's structure, variable-naming conventions, response options and recommended citation format. It also includes a codebook-style summary specifying item identifiers, original strategy domain, item retention status in the final validated model (retained/excluded), response scale coding.

2.6 Repository and versioning

The dataset and accompanying documentation have been deposited in the Zenodo repository under the title "SILL-JP

Dataset" (DOI:10.5281/zenodo.18678940). The current deposit constitutes version 2.0 of the dataset and is intended to remain fixed. Any future modifications or extensions (e.g., additional cohorts, further variables) should be deposited as independent versions within the same Zenodo record, enabling users to differentiate between the versions while retaining access to the original data described in this report.

2.7 Permission to use and copyright

The dataset is based on an adapted version of Oxford's Strategy Inventory for Language Learning (SILL; Oxford, 1990). The original items were translated and modified for non-commercial academic research in accordance with established practices for instrument adaptation. The dataset does not reproduce the original SILL verbatim, and appropriate citation to the original source is provided. The publicly archived files contain item-level response data and abbreviated variable names only. The full adapted questionnaire is available from the author upon reasonable request.

3 Dataset description

The archived dataset consists of item-level responses to a 50-item adapted Japanese version of the SILL (SILL-JP), which are stored in Microsoft Excel (.xlsx) format. Each row corresponds to a single participant, while each column corresponds to a particular variable. The dataset contains no information on the backgrounds or attributes of the respondents, including only their responses to the 50 adapted SILL items. The Microsoft Excel (.xlsx) file contains abbreviated variable names, with variable-naming conventions and response options documented in the accompanying README file.

Each item is coded as a numeric variable along a five-point Likert scale, where 1 represents "hardly true of me", and 5 represents "very true of me". Variables are represented by a two-letter prefix indicating the strategy domain (e.g., MM for memory strategies, CG for cognitive strategies) alongside a numerical identifier that differentiates items within the same domain (e.g., MM1, MM2).

There are no missing values for any of the variables in this dataset. As a result, secondary users need not apply missing-data thresholds or imputation procedures when working directly with the archived file. To illustrate the general characteristics of the dataset, Table 2 presents descriptive statistics at the strategy-domain level for the combined sample. These descriptive summaries are provided for documentation purposes only and should not be interpreted as substantive findings or validation results, which are reported elsewhere (Saito, 2025).

The archived dataset enables examination of internal structure (e.g., dimensionality, reliability estimation, model comparison) at the item level. However, it does not permit evaluation of criterion-related or predictive validity, nor tests of measurement invariance across demographic or proficiency-based subgroups.

TABLE 2 Descriptive statistics of the SILL-JP/STEM: combined sample ($N = 599$).

Strategy domain	M	SD	ω	Range		Skewness	Kurtosis
				Potential	Actual		
Memory (3)	2.87	0.83	0.62	1–5	1.0–5.0	0.11	–0.27
Cognitive (4)	2.75	0.90	.72	1–5	1.0–5.0	0.22	–0.52
Compensation (3)	3.61	0.87	.72	1–5	1.0–5.0	–0.43	0.07
Metacognitive (6)	2.70	0.81	.82	1–5	1.0–5.0	0.23	–0.04
Affective (3)	2.06	0.85	.64	1–5	1.0–5.0	0.68	–0.04
Social (3)	2.63	0.93	.66	1–5	1.0–5.0	0.18	–0.43
Overall (22)	2.73	0.63	.90	1–5	1.04–5.0	–0.01	0.22

The numbers in the parentheses indicate the number of items in each strategy domain.

Reliability coefficients for subscales with three items should be interpreted with caution, as internal consistency indices tend to be attenuated for short scales. Values observed here are comparable to other abbreviated SILL adaptations reported in the literature.

4 Reuse potential

The SILL-JP dataset offers a range of possibilities for secondary psychometric analysis. Researchers may use the data to replicate the factor structure reported in the validation article or to explore alternative psychometric models of strategy use. As the instrument builds on a well-known precursor, the dataset may also support comparative examinations of the SILL's internal structure across similar learner populations and educational contexts. From a methodological perspective, the dataset represents a transparent example of instrument adaptation, scale shortening, and open data archiving in a context where major L2 measurement instruments are being re-examined. It may therefore serve as a reference point for discussions of psychometric adaptation practices in applied linguistics and education.

Data availability statement

The dataset presented in this report is available in the Zenodo repository: DOI:10.5281/zenodo.18678940.

Ethics statement

The studies involving humans were approved by Ethics Review Committee for Life Science and Medical Research Involving Human Subjects at Tokyo University of Science (approval number 23067). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. No personally identifying information was collected, and the archived dataset contains only anonymised responses.

Author contributions

AS: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project

administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author declares that financial support was received for this work and/or its publication. This work was supported by JSPS KAKENHI Grant Number [22K00723].

Acknowledgments

The author thanks Junko Maeda and Mariko Iwasaki for their contributions, particularly their help in adapting the instrument and administering the survey. Their expertise and support were essential to the development of the SILL-JP/STEM instrument. Thanks are also due to the editing services performed by Cambridge Proofreading.

Conflict of interest

The author declares that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Correction Note

This article has been corrected with minor changes. These changes do not impact the scientific content of the article.

Generative AI statement

The author declares that generative artificial intelligence tools (ChatGPT, OpenAI) were used to assist with language editing of this manuscript. They were not used to generate original

research content, data, analyses, or interpretations. The author takes full responsibility for the accuracy, originality, and integrity of the manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

References

- Ardasheva, Y., and T. R. Tretter (2013). Strategy inventory for language learning–ELL student form: testing for factorial validity. *Mod. Lang. J.* 97, 474–489. doi: 10.1111/j.1540-4781.2013.12011.x
- Danko, M., and M. Dečman (2019). The strategy inventory for second language learning: tested, adapted, and validated in the Slovenian higher education context. *ESP Today* 7, 207–230. doi: 10.18485/esptoday.2019.7.2.5
- Griffiths, C. (2018). *The Strategy Factor in Successful Language Learning: The Tornado Effect*, 2nd Edn. Bristol, UK: Multilingual Matters.
- Guillemin, F., and Bombardier, C., and D. Beaton (1993). Cross-Cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J. Clin. Epidemiol.* 46, 1417–1432. doi: 10.1016/0895-4356(93)90142-n
- Hsiao, T. Y. (1997). A study of language learning strategies by college EFL learners: a theoretical perspective. *英語語言與文學學刊* 2, 29–46. doi: 10.6954/MJGE.200703.0235
- Oxford, R. L. (1990). *Language Learning Strategies: What Every Teacher Should Know*. New York: Newbury House Harper Collins.
- Oxford, R. L., and A. D. Cohen (1992). Language leaning strategies: crucial concept and classification. *Appl. Lang. Learn.* 3, 1–35.
- Robson, G., and H. Midorikawa (2001). How reliable and valid is the Japanese version of the strategy inventory for language learning (SILL)? *JALT J.* 23, 32–56. doi: 10.37546/JALTJ23.2-2
- Rose, H., Briggs, J. G., Boggs, J. A., and Sergio, L., and N. Ivanova-Slavianskaia (2018). A systematic review of language learner strategy research in the face of self-regulation. *System* 72, 151–163. doi: 10.1016/j.system.2017.12.002
- Saito, A. (2025). Measuring language learning strategies in STEM disciplines: validation and psychometric analysis of the SILL–JP/STEM instrument. *Cogent Educ.* 12(1):2547033. doi: 10.1080/2331186X.2025.2547033
- Saito, A., and Maeda, J., and M. Iwasaki (2023). Adapting the SILL for a foreign language learning setting: preliminary findings. *Proceedings of the 12th Malaysia International Conference on Languages, Literatures and Cultures*. p. 6–9.
- Saks, K., and Ä Leijen (2018). Adapting the SILL to measure Estonian learners' language learning strategies: the development of an alternative model. *Lang. Learn. J.* 46, 634–646. doi: 10.1080/09571736.2016.1191169
- Thomas, N., and H. Rose (2019). Do language learning strategies need to be self-directed? Disentangling strategies from self-regulated learning. *TESOL Q.* 53, 248–257. doi: 10.1002/tesq.473

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.