

Second-language strategy instruction: Where do we go from here?

Luke Plonsky Northern Arizona University

Situating Strategy Use (SSU3) October 15, 2019

Modeling L2 development



Individual differences

(learner-internal variables associated with L2 development)

How strongly are these constructs associated with L2 proficiency / achievement (as shown via meta-analyses)?

Li (2016); Elahi Shirvan, Khajavy, MacIntyre, & Taherian (in press); Masgoret & Gardner (2003); Teimouri, Goetze, & Plonsky (2019); Linck, Osthus, Koeth, & Bunting (2014); Al-Hoorie (2018).



Strategy Instruction (L2SI)

Def. Explicit training on specific practices or techniques that can be employed autonomously to improve one's L2 learning and/or use (Chen, 2007; Ellis & Sinclair, 1989; Tudor, 1996; Taylor et al., 2006).

(Discussed in over 400 empirical, theoretical, and review article and books (see here)



Outline

Part I: State of the science / substantive findings of SI - The WHAT

Part II: Methodological issues
- The *HOW*

Part III: Recommendations

What do we know about the effects of L2SI?

Part I

Strategy Instruction

- Intuitive appeal
- Theoretical support
 - Strategic competence (e.g., Canale & Swain, 1980)
 - Learner-centeredness (Nunan, 1988; Tudor, 1996)
 - Developmental sequences (i.e., rate vs. route)
 - Autonomy / self-regulation / self-management (Gu, 2003; Rubin, 2005; Tseng, Dörnyei, & Schmitt, 2006)

"Teachers should not focus exclusively on the content of learning. Instead, attention should also be given to the process. For, to be selfsufficient, learners must know how to learn."

From, Toward a Theory of Instruction (Bruner, 1966)

Critiques of Strategy Instruction & SI Research

- Poor design (e.g., small sample sizes, non-random group assignment, exclusion of comparison groups)
- Unjustified selection of strategies
- Uncertainty of long-term effects
- Lack of valid and reliable instruments
- Incomplete reporting of treatments and results
- Absence of a comprehensive theory (C'mon SLA folks!)
- Cost/benefit ratio concerns
- "...what one must teach students of a language is not strategy, but language"

(Bialystok, 1990, p. 147).

(Chamot, 2005; Dörnyei, 1995; Kellerman, 1991; Macaro & Erler, 2007; McDonough, 1995; Macaro & Cohen, 2007; Rees-Miller, 1993; Rose et al., 2018)



Reviewing Strategy Instruction

(Chamot, 2005; Hassan et al., 2005; McDonough, 1995)

Positive effects for SI...

• Contexts

- second language, foreign language
- middle school, HS, university
- Children, adults
- beginner, intermediate, advanced
- class, lab
- Treatments
 - Strategies type: cognitive, metacognitive, socioaffective
 - Number of strategies: 1-99
 - Short-, long-term: 1 day 1 year
 - L1, L2; teacher- or researcher-delivered
- Outcomes
 - L2 skills: reading, writing, listening, speaking, vocabulary, grammar
 - others: autonomy, motivation, strategies use, general language ability

Reviewing Strategy Instruction

(Chamot, 2005; Hassan et al., 2005; McDonough, 1995)

Negative/mixed effects for SI...

• Contexts

- second language, foreign language
- middle school, HS, university
- Children, adults
- beginner, intermediate, advanced
- class, lab
- Treatments
 - Strategies type: cognitive, metacognitive, socioaffective
 - Number of strategies: 1-99
 - Short-, long-term: 1 day 1 year
 - L1, L2; teacher- or researcher-delivered
- Outcomes
 - L2 skills: reading, writing, listening, speaking, vocabulary, grammar
 - others: autonomy, motivation, strategies use, general language ability

Plonsky (2019)

A meta-analysis of the effects of L2SI

RQs: 1. How effective is L2 strategy instruction?

2. What is the relationship between the effectiveness of SI and different learning contexts, treatments, and outcome variables (e.g., skill areas)?

First, what is meta-analysis?

- Empirical approach to reviewing literature
- More systematic and objective than traditional reviews
- Origin? ("Necessity is the mother of invention")

Assumption: Developing scientific knowledge is a cumulative and corporate enterprise. First, what is meta-analysis?

Assumption: Developing scientific knowledge is a cumulative and corporate enterprise.

- THREE hallmarks (Mizumoto, Plonsky, & Egbert, in press; Plonsky & Oswald, 2015)
- 1. Exhaustive (vs. selective) searches (sample ≈ population)
 - \rightarrow Validity generalizability

2. Systematic coding for substantive features and effects (vs. subjectively or idiosyncratically interpreted)

3. Key component: effect sizes (e.g., *d*, *r*) more precise, stable, intuitive, and informative (vs. *p*)

• Consequently...



Lyster & Saito (2010)

- Q: Is it helpful to provide students with feedback when they make errors in class?
- A: YES, but it depends on what type of feedback you provide
- What about SI?

A meta-analysis of the effects of L2SI (Plonsky, 2019)

RQs:

1. How effective is L2 strategy instruction?

2. What is the relationship between the effectiveness of SI and different learning contexts, treatments, and outcome variables (e.g., the four skills)?

Method – Inclusion criteria

- Participants learning an L2
- Treatment that included L2 strategy instruction
- Data collected and compared in a control-experimental (between groups) design
- DV = quantitative measure of the effect of SI
- Sufficient data reported to calculate an effect size (Cohen's d)

Method - Sample

- 77 primary studies of the effectiveness of SI
- 112 unique samples / treatment groups
- 7,890 individual participants

Method – data collection and analysis

• Coded for...

(a) substantive and

(b) methodological features as well as

(c) estimates of treatment effects (Cohen's d).

• Analysis

- RQ1: Weighted average overall
- RQ2: Weighted average for subgroups created according to study features (i.e., potential moderators)

Results: RQ1

- Overall effect size: *d* = 0.66 [.62, .69]
- What does this mean?
 - Relative to SLA: "medium"

Effect Size	Small-ish 25 th percentile	Medium-ish 50 th percentile	Large-ish 75 th percentile
d	.40	.70	1.00



K = 346 primary studies and 91 metaanalyses of L2 research (N > 604,000) (Plonsky & Oswald, 2014)

Results: RQ1

- Overall effect size: *d* = 0.66 [.62, .69]
- What does this mean?
 - Relative to SLA: "medium"
 - Relative to L1 SI: *d* = .45 (Hattie et al., 1996)
 - Exp groups score on average 2/3 of an SD above control groups
 - Approximately 3/4 of EG participants outperform average CG participants (Lipsey et al., 2012)
 - Additional and practical considerations for interpretation
 - Teacher training
 - Materials development
 - Class time (cost/benefit ratio?)
 - Potential for long-term benefit?



Results: RQ1 \rightarrow change over time?

• Overall effect size: *d* = 0.66 [.62, .69]



Two possible explanations

- "a notably greater standardization of intervention frameworks has gradually emerged in the past decade" (Ardasheva et al., 2017)

- Methodological (vs. theoretical?) maturity (Plonsky & Gass, 2011; Plonsky & Oswald, 2014)

RQ 2: Effects of SI Across Learning Contexts



RQ 2: Effects of SI Across Treatment Types



RQ 2: Effects of SI Across L2 Skill Areas



- Meta-analysis of the effects of L2SI on
 - RQ1: L2 performance
 - RQ2: Other self-regulated outcomes (e.g., anxiety, self-efficacy, attitudes)
- 2008-2014 only
- Sample
 - RQ1: 39 reports (47 samples)
 - RQ2: 16 reports (17 samples)

Overall Results



Results, RQ1 (linguistic outcomes)





Additional meta-analytic evidence for strategies

- English learning (overall) (Elahi Shirvan, 2014)
- Reading (Chaury, 2015; Maeng, 2014; Taylor et al., 2006)
- Vocabulary learning strategies for EFL learners (Nematollahi et al., 2017)
- Web-based instruction (Chang & Lin, 2013)

Preliminary Implications and Discussion

SI *can* be effective in all contexts and for all skills but appears to be stronger:

(a) with non-beginners ("threshold" in Chamot, 2016; "the rich get richer"?)

(b) with metacognitive strategies

(c) over longer periods of time, and

(d) fewer target strategies (i.e., less is more)

BUT

A great deal of further research is still needed across...

- Learner demographics and contexts
- Linguistic (i.e., skills) and non-linguistic domains (e.g., anxiety)
- Individual strategies

The *HOW* (SI Methods)

Part II

We have some issues

- Design & Instrumentation (see e.g., Pawlak, 2019; Rose et al., 2018)
 - Small samples
 - Lack of delayed posttests
 - Lack of theoretical or empirical justification of strategies taught
 - Evidence of reliability (internal consistency) and validity often unknown BOTH for measures of strategies AND L2 performance!

"You can't fix with analysis what you bungled by design" (Light et al., 1990)

No analysis—however sophisticated or elegant—can make up for poor instrumentation.

At least we're not alone?

• True. These problems are pervasive throughout pretty much all of applied linguistics (and throughout the social sciences)!

Reliability evidence

 $O_{(observation)} = T_{(true \ score)} + E_{(error)}$

Reporting of reliability across domains of L2 research



What about the *amount* of (measurement) ERROR? What's typical for the field?

- Reliability generalization meta-analysis (RGM) (Plonsky & Derrick, 2016)
- *K* = 537 from 16 L2 journals
- 2,244 reliability estimates



In other domains of L2 research?

• **TBLT** (*K* = 85; Plonsky & Kim, 2016)

TABLE 8. Descriptive Statistics for Meta-Analysis of Reliability Coefficients

Μ	SD	95% CI	Min–Max
.94	.05	[.93, .95]	.80–1.00

In other domains of L2 research?

• L2 pronunciation (K = 77; Saito & Plonsky, 2019)



What about for L2 strategies?

- Or even for different categories of strategies?
- •We really don't know!

• Why does this matter?

- Unreliability \rightarrow Error \rightarrow Threat to validity
- Attenuation of effects (signal vs. noise)

Validity Evidence

O_(observation) = T_(true score) + E_(error)

What we *say* about validity

- **Chapelle (in press):** "validation should be of central importance for the credibility of research results"
- **TQ Author Guidelines:** Authors should provide a "Description of the instruments, what they are designed to measure, and a report of their validity to the extent possible, and their reliability."
- Ellis (in press): "While researchers have always recognized this issue [validity in SLA measurement], they have largely ignored it, often happy to talk about learning with no consideration of the type of data they had collected"
- Norris & Ortega (2012): "Problematic... is the tendency to assume—rather than build an empirical case for—the validity for whatever assessment method is adopted (pp. 574-575).
- Schmitt (2019): "Most vocabulary tests are not validated to any great degree."

What about strategy scales???

See seminal works by Cronbach & Meehl; Messick; Kane; Chapelle, Enright, & Jamieson

Are questionnaires to blame?

(Great examples of alternatives in Gu's plenary and Yashima & MacIntyre's symposium)

- 1. Indirect measures of the construct of interest
- <u>Suggestion</u>: triangulation (e.g., +observations; +interview)
- 2. Responses often limited to what is being asked
- <u>Suggestions</u>: piloting; open-ended items; interviews
- 3. Self-selection bias
- <u>Suggestions</u>: random or purposive sampling; missing data an alysis
- 4. Anonymity \rightarrow +/- truthfulness?
- <u>Suggestion</u>: triangulation
- 5. Response values? (3, 5, 9, 1,000?)
- Suggestions: piloting; clear instructions; scale descriptors
- 6. Quantification without consideration of numerical va SunTrust
- Suggestion: rich qualitative data; better use of stats
- 7. Ambiguous ("double-barreled") items
- Suggestion: Pilot. Leave room for comments.



Potential threats to validity

How likely are you to continue to do business with SunTrust because you

5

Likely

10

value SunTrust? Not at all Likely

2



"There is perhaps an unwritten agreement that readers will accept measures used in an SLA study at face value without asking about their reliability and validity for the task at hand." (Cohen & Macaro, 2013, p. 133; see Bachman & Cohen, 1998).

- Is this true in general?
- And for strategies research?
- Do you ever see validity evidence?



- How could we address this Q?
- Collect a representative sample of studies...
- Synthetic approach
 - Very time-consuming
 - Subject to high inference judgments
- Corpus-based approach
 - Fast and objective
 - Valid?

- Second Language Research Corpus (L2RC; Plonsky, n.d.)
 - 22 journals
 - 22,363 articles (1946-2018)
 - 147,293,764 words
- Searched for occurrences of:
- [predictive, discriminant, divergent, construct, face, convergent, concurrent] + validity
- validity argument

AL, ALL, AP, BLC, CMLR, ELTJ, FLA, IJAL, IRAL, JSLW, LAQ, LA, LL, LL&T, LTeaching, LTR, LTesting, MLJ, SLR, SSLA, System, TQ



(What about false positives? False negatives?)

It's not all bad!

- Nakatani (2006): Scale development/validation
 - Development of the Oral Communication Strategy Inventory (OCSI)
 - Stage 1: Open-ended questionnaire (N=80)
 - Stage 2: Piloted with 400 → (exploratory) factor analysis (item structure) → 8 categories for speaking and 7 for listening strategies
 - Stage 3: Compared with data from SILL (N=62)
- Mizumoto & Takeuchi (2012): Scale validation
 - Self-regulating Capacity in Vocabulary Learning Scale (SRCvoc)
 - Study 1: N = 443 \rightarrow item-analysis: ITC of >.4; alpha for subscales
 - EFA to examine factor structure
 - Study 2: N=914 \rightarrow alpha for subscales; CFA
- Ardasheva & Tretter (2013) : Validation of modified version of SILL
 - Revision of items and piloting
 - Administered to 1057 child learners of ESL
 - CFA \rightarrow 6 factor solution

It's not all bad!

- See also
 - Tragant et al. (2013)
 - Ardasheva (2016)
 - Teng & Zhang (2016)

Summary for Part II

• What we need is a non-casual, rigorous, and systematic agenda focused on **measurement** as it pertains to L2 strategies and strategy instruction.



Looking ahead \rightarrow L2SI research wish list

(see Sudina & Plonsky, in press)

What / substance

- SI across all skill areas. Esp: writing, listening, pronunciation, test-taking
- Aptitude-treatment interactions with L2SI (e.g., with beliefs, working memory; see Yashima, Nishida, & Mizumoto, 2017)
- SI for specific learning contexts: SA, CALL/MALL, EMI/CLIL
- Teacher training
 - Studies of teacher beliefs regarding SI and
 - Effectiveness of teacher training interventions for SI
- The role of strategic transfer (L1 \rightarrow L2; L2 \rightarrow Ln)

Looking ahead → L2SI research wish list (see Sudina & Plonsky, in press)

How / Method: Designs

- Non-"WEIRD" samples: e.g., SL, pre-adolescent, advanced learners
- Validity evidence/arguments for the utility of individual strategies ← essential justification for L2SI studies but RARELY present
- A clearer understanding of the long-term effects of SI
- "Bigger" and more longitudinal designs—at the curricular level

Looking ahead → L2SI research wish list (see Sudina & Plonsky, in press)

How / Method: Measurement

• Validity arguments for measures of both

(a) strategy usage (Takeuchi, 2019; Tseng et al., 2006) and

- Situated, qualitative, and mixed methods (Pawlak & Oxford, 2018; Rose et al., 2018)
- Macro+micro perspective (Pawlak, in press)
- Scenario-based scales \rightarrow +contextualization (see Teimouri, 2018)

- Studies of the predictive validity of individual strategies and L2 performance (as a pre-requisite for SI)

(b) L2 performance

... all to them be made available on the IRIS database iris-database.org \rightarrow +consistency across studies!!

Looking ahead → L2SI research wish list (see Sudina & Plonsky, in press)

How / Method: Data report and analyses

- More thorough reporting of
 - Sample characteristics (e.g., proficiency)
 - Treatments (e.g., length/intensity, materials)
 - Data (ESs, CIs, visuals, reliability coefficients)
- More informed use of quantitative analyses (Mizumoto & Plonsky, 2015; Nix, 2018; Takeuchi, 2019)
 - E.g., Rasch analysis; Multivariate models; corrections for attenuation due to measurement error

Looking ahead → L2SI research wish list (see Sudina & Plonsky, in press)

How / Method: Beyond individual studies

- Replication studies!
- Additional meta-analyses of SI focused on individual strategies or skills (e.g., vocab, speaking)
- Systematic review and meta-analysis of reliability coefficients (what's normal?)

Thank you!

Luke Plonsky lukeplonsky@gmail.com lukeplonsky.wordpress.com