SREE

2040 Sheridan Rd
Evanston, IL 60208

Society for Research on Educational Effectiveness
Advancing Education Research

www.sree.org
847.467.4001

2010 SREE Conference Abstract Template

Thank you for your interest in the Society for Research on Educational Effectiveness 2010 Annual Conference. Conference abstracts must be submitted using this template document. The template is based on the recommendations offered by Mosteller, Nave, and Miech (2004, p. 33)[*] for structured abstracts. Abstracts should follow APA Style, as specified in the Sixth Edition of the Publication Manual of the American Psychological Association.

Abstract Preparation and Submission Procedures

Save this document to your computer. Fill in each of the sections below with the relevant information about your proposed poster, paper, or symposium. Make sure to save the document again when completed. When ready, submit your abstract at http://www.sree.org/conferences/2010/submissions/

The template consists of the following sections: title page, abstract body, and appendices (references and tables and figures). Figures and tables included as part of submission should be referred to parenthetically—"(please insert figure 1 here)." The body section of your abstract should be no longer than 5 pages (single spaced, using the Times New Roman 12-point font that has been set for this document). The title page and appendices do not count toward this 5-page limit.

Insert references in appendix A of this document. Insert tables and graphics in appendix B. Do not insert them into the body of the abstract.

**For questions, or for help with abstract preparation or submission,
contact us at inquiries@sree.org, or 847-467-4001**

---

[*] Mosteller, F., Nave, B., & Miech, E. (2004). Why we need a structured abstract in education research. *Educational Researcher, 33(1),* 29–34.

Abstract Title Page
Not included in page count.

Title: Development of a Language Impairment Screener for Spanish Speaking Children - SSLIC: Phase 1 – Task Development

Author(s): M. Adelaida Restrepo, Joanna S. Gorin, Shelley Gray, Gareth P. Morgan, Nicole Barona.

Abstract Body

Background/context:

Language Impairment (LI) is a developmental disorder that affects approximately 7% of children in the U.S. (National Center for Education Statistics, 2004). Young children with LI are more likely than their peers with typical development (TD) to develop reading disabilities and to experience significant oral and written language problems (Bishop & Adams, 1990; Catts, Adlof, & Weismer, 2006; Nation, Clarke, Marshall, & Durand, 2004; Snowling, et al 2000). Although language problems are the core disability of many children with reading disabilities, their LI frequently is not identified at a young age (Catts, Adlof, & Weismer, 2006; Nation, Clarke, Marshall, & Durand, 2004), which highlights the importance of early identification and treatment of LI to prevent future academic difficulties. This is especially important in children acquiring English as a second language because LI in English Language Learners (ELLs) leads to significant academic and communication difficulties (Restrepo, 2000; Restrepo, 2003).

Current language assessment instruments are inadequate for assessing ELLs. Consequently, they contribute to poor decision making regarding the placement of ELLs in special education and gifted programs (Garcia & Pearson, 1994). Educators need a language screening instrument for Spanish-speaking children that is valid and reliable across a variety of cultural and language experiences because Spanish-speaking (SS) families in the U.S. vary in their cultural, educational, and economic backgrounds. Assessments for these children must be based on a model of LI in the Spanish-speaking population that considers their diversity.

Purpose / objective / research question / focus of study:

The main purpose of this study is to develop a Spanish language screening measure that (a) is valid and reliable for the purpose of identifying SS children at risk for LI, (b) is valid and reliable across different Spanish dialects, different socioeconomic groups, and different ethnicities, (c) uses a Spanish LI model rather than an English language model, and (d) is easy to administer and score by paraprofessionals in schools in the United States (U.S.). The screening measure is intended as a universal screening instrument in pre-kindergarten and kindergarten and as a screening tool for speech-language pathologists in first through second grade students referred by teachers, physicians, other professionals, or parents. Early and accurate identification of LI risk will lead to timely evaluation, identification and treatment of LI. This will result in increased academic success by ELLs that will positively impact academic achievement in U. S. schools.

For this presentation we report on outcomes from the first development phase of the research. Methodologically we are using Evidence Centered Design (ECD; Mislevy 1994) to develop a screening instrument that is grounded in LI theory for Spanish-speaking children. More than traditional test development approaches, ECD ties test design specifically to cognitive models and targeted inferences for the population of interest (Mislevy, 1997), in this case SS children with LI. Currently available instruments are tied to generic theoretical models or to models developed on dramatically different populations than those of SS children. These assessments are not likely to support accurate decision making for this population.

The theoretical framework of the screening measure is based on evidence that LI in Spanish may be characterized by deficits in processing capacity or processing speed as characterized by limited processing capacity (LPC) and by linguistic deficits, including grammar and semantics. We are developing a variety of tasks to tap processing and linguistic skills that have been shown empirically to differ in typically-developing and LI children and in SS children

with LI. We hypothesize that children might perform poorly on a particular task either because of LPC, because of linguistic deficits, or a combination of both. See figure 1.

Setting/ Population / Participants / Subjects:

Participants were drawn from schools in the Phoenix area with large percentages of ELLs who speak Spanish as their native language. Participants were 153 SS children enrolled in kindergarten and 1ˢᵗ grade. Qualification for participation in the study was based on the following criteria: (1) parent questionnaire indicated Spanish use at home, (2) teacher questionnaire indicated the child is learning English, (3) a hearing screening, (4) a standard score of 75 or higher on the Kaufman Assessment Battery for Children – Second Edition (2004) and (5) a language proficiency score of < 5 (1-5 scale) derived from a language sample analysis in English. Qualified children were classified LI or TD based on their performance on the Clinical Evaluation of Language Fundamentals - Spanish (Wiig, Secord, Semel, 2006). Those who scored below a standard score of 85 were classified as LI; children with scores equal to or greater than 85 were classified as TD.

Intervention / Program / Practice:

The current research phase focuses on developing items in three tasks : (a) rapid automatized naming, (b) sentence repetition, and (c) a morphological cloze task. A rationale for each task and a brief description are provided below. Item were generated for each task, and item maps were developed for each item to identify the type of item, scoring, and priming for each item. Due to space limitations, the cognitive maps for each task will not be described.

Rapid Automatized Naming (RAN). RAN performance has been shown to differentiate children with LI from those with typical development (Denkla & Rudel, 1976; Manis, Seidenberg, & Doi, 1999; Catts et al., 2006; Catts, Adlof, Hogan, & Weismer, 2005). Kail (1999) argued that RAN reflects global or sequential processing skills and thus performance suffers in children with LPC. The semantic difficulties in LI have been attributed to lexical retrieval difficulties (Fried-Oken, 1987). Manis et al. (1999) found that RAN differentiated children with and without language-learning disabilities and Brackenbury and Pye (2005) suggested that naming tasks be used as one way to identify semantic difficulties in children with LI. We hypothesize that RAN will identify children with LI due to LPC speed.

Our proposed RAN tasks use common objects and colors familiar to young SS children. Children are given a set of colored or black and white objects to name as fast as they can. Object names were selected to be acquired before age 3 or with high accuracy in labeling in preschoolers. Colors and objects were selected to be common names across dialects and phonology that is acquired early. Two forms of color and black and white forms were developed. Each child received two forms of either color or black and white versions.

Sentence repetition. Sentence repetition taps processing capacity through working memory and linguistic knowledge of syntax and morphology (Conti-Ramsden et al., 2001). It has a history of use in screening batteries for LI (Stokes, Wong, Fletcher, & Leonard, 2006b; Conti-Ramsden, Botting, & Faragher, 2001). More recently sentence repetition has been identified as a possible clinical marker of LI in a variety of languages including Italian (Devescovi & Caselli, 2007), Cantonese (Stokes, Wong, Fletcher, & Leonard, 2006a), Spanish (Gutierrez-Clellen, Restrepo, & Simon-Cereijido, 2006), and English (Conti-Ramsden et al., 2001). In addition, sentence repetition also identifies children with poor reading comprehension (Nation et al., 2004).

The sentence repetition task assesses the use of both morphological and syntactic grammatical skills. In morphology we examine the number words recalled per sentence and in

syntax we increase complexity by increasing number and type of sentence complements, which have been found to be reduced in SS children with LI (Bosch & Serra, 1997) In our sentence repetition task, children repeat sentences presented to them one at a time. Sentences are constructed using increasingly more complex complements following a hierarchy of difficulty: multiple adjuncts, ditransitive, causal, relative, and conditional sentences. Sentences are controlled for number of total words, number of function words in each category, number of content words, and number of syllables across sentences and categories. Each subtest items were randomly organized and the subtests were given in one of two randomly set orders. Design characteristics used to generate items included presence of relative clauses, presence of causals, presence of temporal, presence of conditionals, presence of subordinate if clauses, number of adjuncts, and phrase structure.

Morphological Cloze Task. Young children with LI often demonstrate grammatical errors (Leonard, 1998). In preschool and kindergarten morphology seems to be a vulnerable to error across languages, and this may persist well into later school years, therefore the utility of morphological skills as identifiers of LI across languages, including Spanish, is strong (e.g., Restrepo, 1998; Restrepo & Gutierrez-Clellen, 200; Gutierrez-Clellen et al., 2006; Gutierrez-Clellen, Restrepo, & Simon-Cereijido, 2004). Restrepo (1998) found that the number of grammatical errors per sentence was a significant predictor of LI in SS children in the U.S. Gutierrez-Clellen, Restrepo, and Simon-Cerreido (2006) found that Spanish sentence repetition and cloze tasks examining clitics, articles, and subjunctives classified SS children with and without LI with accuracies above 85%.

A cloze task with picture support is used to examine children's use of articles, clitic pronouns, subjunctive verbs, prepositions, and derivational morphemes. The cloze task uses a sentence completion activity that elicits a target word. Based on prior research, we hypothesized that each areas can differentiate children by ability group. Design characteristics used to differentiate items included number, gender, verb tense, semantic category, and word ending.

Research Design:

SSLIC development and validation procedures are aligned with Evidence Centered Design (ECD; Mislevy, 1994). ECD frames an assessment in terms of an evidentiary argument, connecting what we observe of individuals' behavior to what we know about their abilities (Mislevy, Steinberg, & Almond, 2003). The ECD approach was designed to address substantive validity questions including the following: (1) What combination of knowledge, skills, and abilities should be assessed?; (2) What behaviors or performances should reveal those constructs?; and (3) "What tasks or situations should elicit those behaviors? (Mislevy, 1994). The SSLIC design begins with a strong theoretical model of the critical constructs associated with LI in SS children. Next, tasks with processing and linguistic characteristics relevant to the construct were designed based on the literature. Finally, statistical analysis is conducted to identify the tasks and items that provide the strongest evidence for inferences about SS children with LI. The current study will present Phase I of III of the project.

The purpose of Phase I is ***to generate and validate tasks*** to assess individual theoretical components of SS children with LI. Specifically we address these questions with the first three tasks of six, (a) What are the critical theoretical components of LI for SS children?, (b) What types of tasks generate evidence of the critical theoretical components?, and (c) For each task, what characteristics of individual items maximize score differentiation for SS LI?

Data Collection and Analysis:

Children are first screened through parent report, teacher report, and a language

proficiency rating scale to determine whether their English proficiency is sufficient to disqualify them from participation. Qualified participants are then screened for hearing, language proficiency, and nonverbal IQ through the K-ABC II. Qualifying children are then administered the CELF-S to determine classification as LI or TD for comparison on the experimental measures. Finally, all remaining children are administered the three experimental measures in random order. Data collection on these measures is still underway; however, preliminary results are presented. The current analysis included 153 children in Preschool through 3rd grade. Given the small sample sizes for preschool, second and third grades, results are disaggregated for kindergarten and first grade only. The final presentation will provide results disaggregated by grade, gender, age, and language status (LI versus TD).

Descriptive statistics were computed on all items and subscales for LI and TD students separately. Analysis independent samples t-tests of means on individual items (for Morphology and Sentence Repetition), and subscale scores for all measures were examined. For the morphology task, subscale scores were calculated as percent correct scores on items of each type (e.g., subjunctives, clitics). For the RAN task, total time and total number of errors were examined separately for two parallel Black and White forms and two parallel Color forms. For Sentence Repetition, separate analyses were conducted for dichotomous item scoring (i.e., correct/incorrect) and for the number of words repeated correctly. These scores were averaged across items to produce two total scores, Number of Correct Sentences and Number of Words Correct.

To examine the potential for a short version of each task type to be part of the screener, a subscale score was calculated for each task using the "best six items". To select these, the individual item results were examined for the Morphology and Sentence Repetition tasks to identify items with the largest between-group effect sizes given the LI and TD mean differences. To account for possible changes in item properties across the two grades, this process was conducted twice – once to identify the best items for the Kindergarten students and then for the 1st grade students. First, the design characteristics of these items were examined to identify promising features for future item development. Next, scores for each grade on the six best items were calculated and means were compared. For each Kindergarten, 1st grade, and the entire sample independent samples t-tests were conducted: 12 tests for morphology (six subscales tested with the best six items from Kindergarten and 1st grade), four tests for RAN (Error Scores and Time Scores for each of the two Black and White and two Color forms), and four tests for Sentence Repetition (Sentences Correct Scores and Words Correct Scores based on the best six items from Kindergarten and 1st grade). This yielded a total of 60 independent sample t-tests at the scale or subscale levels. Given that the current purpose of the analysis is to identify promising items and tasks, not to interpret or generalize mean differences to formally test a hypothesis about population differences, Type I error rates were not adjusted.

Findings / Results:

A report of item analysis, including item descriptive statistics, item total correlations, inter-item correlations, subscale descriptive statistics, and results of all inferential tests will be reported in the full paper. Further, the relationship between item design characteristics for the Morphology and Sentence Repetition tasks and item quality will be discussed. Due to space limitations, highlights of subscale level findings are summarized. The most general conclusion based on the descriptive statistics, regarding all items and subscales per grade, is that mean scores were always higher for the TD students than for the LI students. Regarding the Morphology task, using all initially developed items, all subscales yielded significant mean differences between the

LI and TD samples in both Kindergarten and 1st grade. The effect sizes ranged across subscales, with the largest eta-squared observed for the Clitics in K and 1st grade ($\eta^2 = .35$ and $\eta^2 = .21$, respectively) and the smallest effect size observed for the Articles in Kindergarten ($\eta^2 = .16$) and Subjunctives in 1st grade ($\eta^2 = .21$). These effect sizes grew when subscale scores were computed based only on the "best six items" with the largest eta-squared observed for the Preposition Subscale for the Kindergarten Students ($\eta^2 = .45$).

Significant differences between TD and LI mean Error Scores were observed on both parallel Black and White forms of the RAN, $F(1,70)=12.19$, $p<.01$ and $F(1,70)=4.03$, $p=.05$. Differences in Mean Completion Times were only significant on one of the two Black and White forms $F(1,69)=4.12$, $p=.05$. No significant differences in performance on either Color form (Completion Times or Error Rates) were observed. In general the effect sizes were quite small ($\eta^2 >.10$) with the exception of the Error Rates on Black and White Form A, which yielded an effect size of $\eta^2 = .15$.

For Sentence Repetition, all independent sample t-tests for the Kindergarten children yielded significant observed statistics. Only one comparison for the 1st grade children, the Number of Words Correct based on the best six items, was significant, $F(1,29)=8.58$, $p=.01$. Effect sizes were largest for tests of mean differences between LI and TD children on the Number of Words Correct scores as opposed to the Number of Sentences Correct.. The largest effect size was observed for the Kindergarten children's Number of Words Correct based on the best six items, $\eta^2 = .40$.

Conclusions:

(a) What are the critical theoretical components of LI for SS children?, Results indicate that the three measures are differentiating LI vs. TD groups at each age, as hypothesized: Sentence repetition, RAM and Morphology. (b) What types of tasks generate evidence of the critical theoretical components?, and (c) For each task, what characteristics of individual items maximize score differentiation for SS LI? The item level results indicate variability in the quality of items such that a subset of items show strong differentiation between LI and TD student responses. In some cases a relationship with task characteristics were observed. In most cases, the item characteristics were independent of the quality of the item. For example, in the morphological task, results indicated that all subtests differentiated groups. In the clitic pronouns, articles, and subjunctives there was no apparent pattern in design characteristic for the best six items. In prepositions, however, the indirect object preposition differentiated between groups using four of the six best items. In sentence repetition scores varied by sentence type; however, sentences longer than 14 words had poor differentiation. In the RAN task, differentiation was poor for the color forms, due to the errors in color names and the slow retrieval of forms across both groups. The black and white forms appear to be differentiating across groups.

In summary, the three tasks designed to differentiate groups are working as hypothesized, leading to significant differences between TD and LI groups. Further, specific items in each task are also differentiating the groups, and the selection for the best six items reflect large effect sizes in each subtask. Of the three tasks, morphological subtests for clitics and prepositions in kindergarten and preposition and articles in first grade show the greatest potential, along with sentence repetition. Using the best six items characteristics, we will generate more items for the second part of item generation. For the RAN task, only one form differentiated groups, and thus a new form with its characteristics will be developed.

Appendix A. References

Aram, D. M., Ekelman, B. L., & Nation, J. E. (1984). Preschoolers with language disorders: 10 years later. *Journal of Speech and Hearing Research, 27,* 232-244.

Bishop, D. V. M. & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry, 31,* 1027-1050.

Bosch, L. & Serra, M. (1997). Grammatical morphology deficits of Spanish-speaking children with specific language impairment. *Amsterdam Series in Child Language Development, 6.*

Brackenbury, T. & Pye, C. (2005). Semantic deficits in children with language impairments: issues for clinical assessment. *Language, Speech, and Hearing Services in Schools, 36,* 5-16.

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language Deficits in Poor Comprehenders: A Case for the Simple View of Reading. *Journal of Speech, Language, and Hearing Research, 49,* 278-293.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment. *Journal of Child Psychology & Psychiatry, 42,* 741-748.

Denkla, M. B. & Rudel, R. G. (2007). Rapid Automatize Naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14,* 471-479.

Devescovi, A. & Caselli, M. C. (2007). Sentence repetition as a measure of early grammatical development in Italian. *International Journal of Language & Communication*

García, G.E., & Pearson, P.D. (1994). Assessment and diversity. *Review of Research in Education*, 20, 337-391.

Gorin, J. S. (2007). Test Construction and Diagnostic Testing. In J. P. Leighton & M. J. Gierl, Eds. Cognitive Diagnostic Assessment in Education: Theory and Practice. Cambridge University Press.

Gorin, J. S. (2006). Item Design with cognition in mind. *Educational Measurement: Issues and Practice*, 25, 21-35.

Gorin, J. S. (2005). Manipulation of processing difficulty on reading comprehension test questions: The feasibility of verbal item generation. *Journal of Educational Measurement*, 42, 351-373.

Gutierrez-Clellen, V. F., Restrepo, M. A., & Simon-Cereijido, G. (2006). Evaluating the discriminant accuracy of a grammatical measure with Spanish-speaking children. *Journal of Speech and Hearing Research. 49,* 1209-1223.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children:* Circle Pines, MN: American Guidance Service.

Manis, F. R., Seidenberg, M. S., & Doi, L. M. (1999). See Dick RAN: Rapid naming and the longitudinal prediction of reading subskills in first and second graders. *Scientific Studies of Reading, 3,* 129-157.

Mislevy, R. J., Sheehan, K. M., & Wingersky, M. S. (1993). How to equate tests with little or no data. *Journal of Educational Measurement, 30,* 55-78.

Mislevy, R. J. (1995). Probability-based inference in Cognitive Diagnosis. In P. D. Nichols, S. F. Chipman, and R. L. Brennan (Eds.) *Cognitively Diagnostic Assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika, 59,* 439-483.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3-62.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). On the roles of task model variables in assessment design. In S.H. Irvine, & P.C. Kyllonen (Eds.) Item generation for test development (pp. 97-128). Mahwah, NJ: Lawrence Erlbaum.

Morgan, G., Restrepo, M. A., & Auza, A. (in press) Variability in the grammatical profile(s) of Spanish-speaking children with specific language impairment. In Grinstead, J. (Ed.). *Hispanic child languages: typical and impaired development.* John Benjamins.

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden Language Impairments in Children: Parallels Between Poor Reading Comprehension and Specific Language Impairment? *Journal of Speech, Language, and Hearing Research, 47,* 199-211.

National Center for Education Statistics (2004). *Higher Percentages of Fourth- and Eight-Graders Performed at or Above Proficient in 2003 Compared to 1992* Washington DC.

Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1398-1411.

Restrepo, M. A. (2003). Spanish language skills in bilingual children with specific language impairment. In S.Montrul & F.Ordoñez (Eds.), *Linguistic Theory and Language Development in Hispanic Languages. Papers from the 5th Hispanic Linguistics Symposium and the 4th Conference on the Acquisition of Spanish and Portuguese* (pp. 365-374). Summerville, MA: Cascadilla Press.

Restrepo, M. A. & Gutierrez-Clellen, V. (2001). Article use in Spanish-speaking children with specific language impairment. *Journal of Child Language, 28,* 433-452.

Restrepo, M. A. & Gutierrez-Clellen, V. F. (2004). Grammatical Impairment in Spanish-English Bilingual Children. In Brian A. Goldstein (Ed.), *Bilingual Language Development and Disorders in Spanish-English Speakers* (pp. 213-234). Baltimore: Brooks Publishing.

Restrepo, M. A. & Kruth, K. (2000). Grammatical characteristics of a Spanish-English bilingual child with specific language impairment. *Communication Disorders Quarterly, 21,* 66-76.

Stokes, S. F., Wong, A. M. Y., Fletcher, P., & Leonard, L. B. (2006). Nonword Repetition and Sentence Repetition as Clinical Markers of Specific Language Impairment: The Case of Cantonese. *Journal of Speech, Language, and Hearing Research, 49,* 219-236.

Snowling, M., Bishop, D. V.M., & Stothard, S.E. (2000). Is Preschool language impairment a risk factor for dyslexia in adolescence? *Journal of Child Psychology & Psychiatry & Allied Disciplines, 41,* 587-601.

Wiig, E. H., Secord, W. A., & Semel, E. (2006). *Clinical Evaluation of Language Fundamentals -- Fourth Edition Spanish*. San Antonio, TX: Harcourt PsychCorp

Appendix B. Tables and Figures
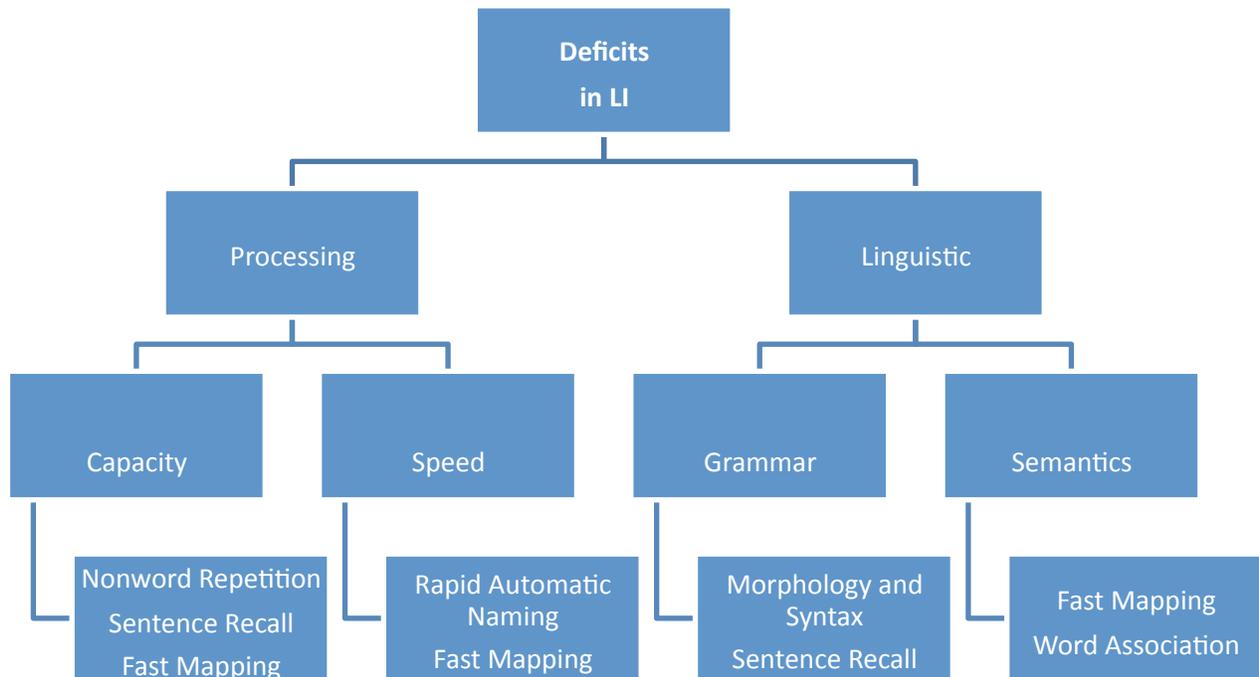
Figure 1. Theoretical Framework of the SSLIC

Figure 2. Graphic Summary of the Student, Evidence, and Task Models