



Introduction

The development of novel educational assessment models founded on Item Response Theory (IRT), as well as software tools designed to implement these models, has contributed to the surge in Computerized Adaptive Tests (CATs) (Embretson & Reise, 2000). The distinguishing characteristic of CATs is that the sequence of items on a test progressively adapts to the performance levels of students as they are taking it. An important advantage of CATs is that they can reduce the duration of the assessment by automatically excluding in real time those items that are either too easy or too hard for a student's capabilities.

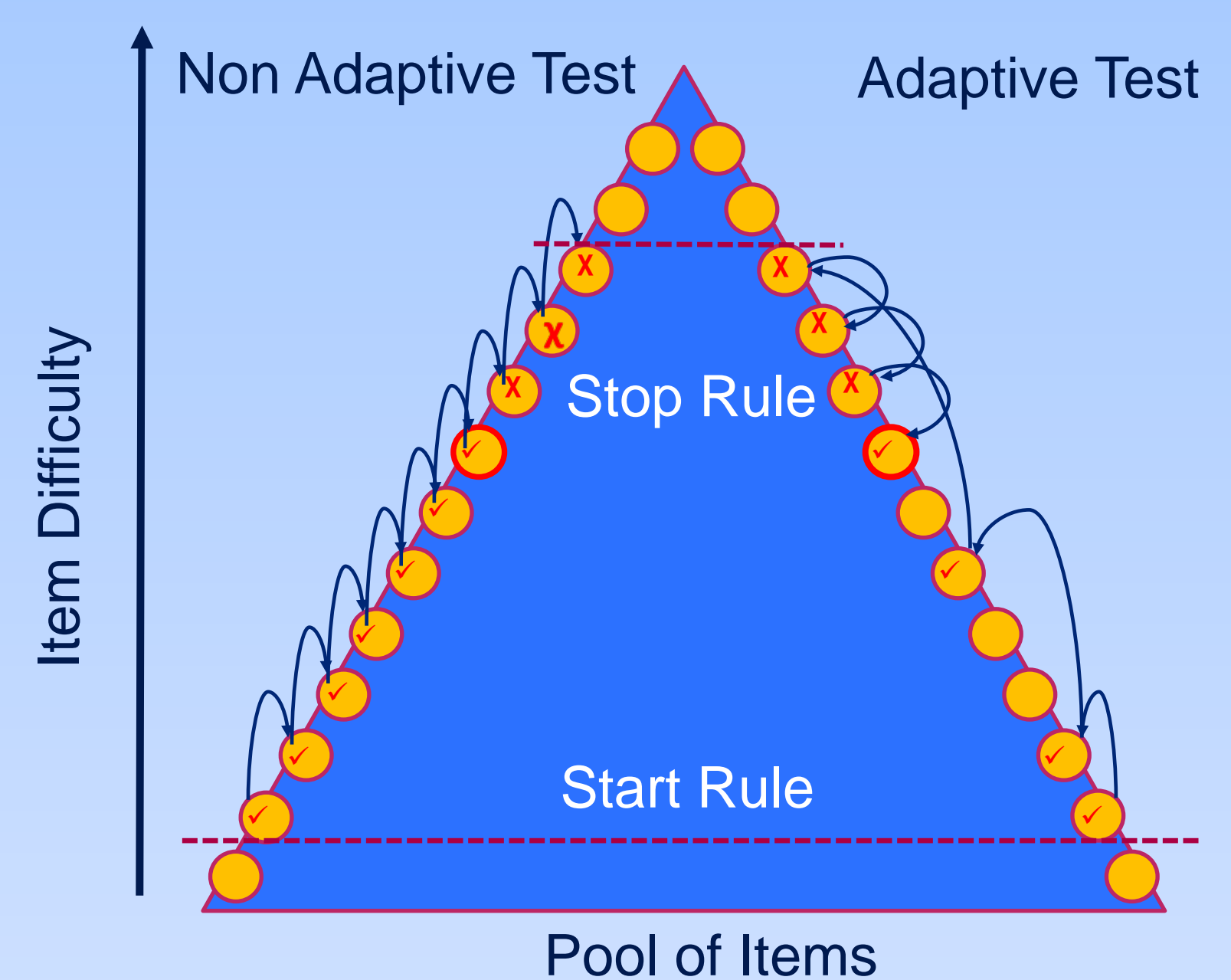


Fig. 1. Computerized Adaptive Test

More recently, dynamic CATs have emerged that include unique features (e.g., graduated prompts) to assess the *zone of proximal development* (ZPD) of the students. This allows test administrators to obtain information about the kind and level of mediation required by the students to reach their optimal performance (Shapiro, 2012).

Proposal of the present study
 This study is part of the large comprehensive research. This research aimed to develop a *computerized adaptive dynamic assessment battery of reading processes* (EDPL-BAI) and to analyze their predictive and incremental validity on reading competence. The present study aims to analyze the incremental validity on reading competence of the dynamic scores obtained from the implementation of a set of adaptive dynamic tests of morpho-syntactic processes integrated into the EDPL-BAI battery. The analysis was conducted using a structural equation model was implemented to check the relationship between the potentially predictive variables and criterion-referenced tests. A model was built to test the main hypothesis. In this sense, we expected a dynamic scores to signify an incremental explicative factor of reading competence in relation to the static tasks of intelligence and comprehension.

Methods

Participants
 The research frame involved 1831 students (46% girls) from 13 public schools in three regions of Chile. A subsample of 378 students was selected for the present study. These students had completed the adaptive dynamic tests of morpho-syntactic processes during the second phase of the study. 54 students were removed as outliers. Remaining 324 students (46% female) were in the 3rd (26), 4th (73), 5th (118), and 6th (107) grades and aged between 8 and 12 years old ($M = 10.27$, $SD = 1.22$).

Methods

Instruments
Tests of the EDPL-BAI battery: Dynamic Morpho-syntactic Awareness Test (MS), Dynamic Syntactic Awareness Test – To sort disordered sentences (OF), and Dynamic Syntactic Awareness Test (CS).
Criteria measures: Reading Comprehension Tests, CLPT, Pretest and Posttest EDPL-BAI, Test of Raven's progressive matrices, and Teachers' assessment of reading performance.

Procedure
 The framework research in this study was developed in three phases. After the item calibration was addressed at Phase 1 of the study, the administration of the EDPL-BAI and the criteria tests were conducted at Phase 2 (Fig. 2).

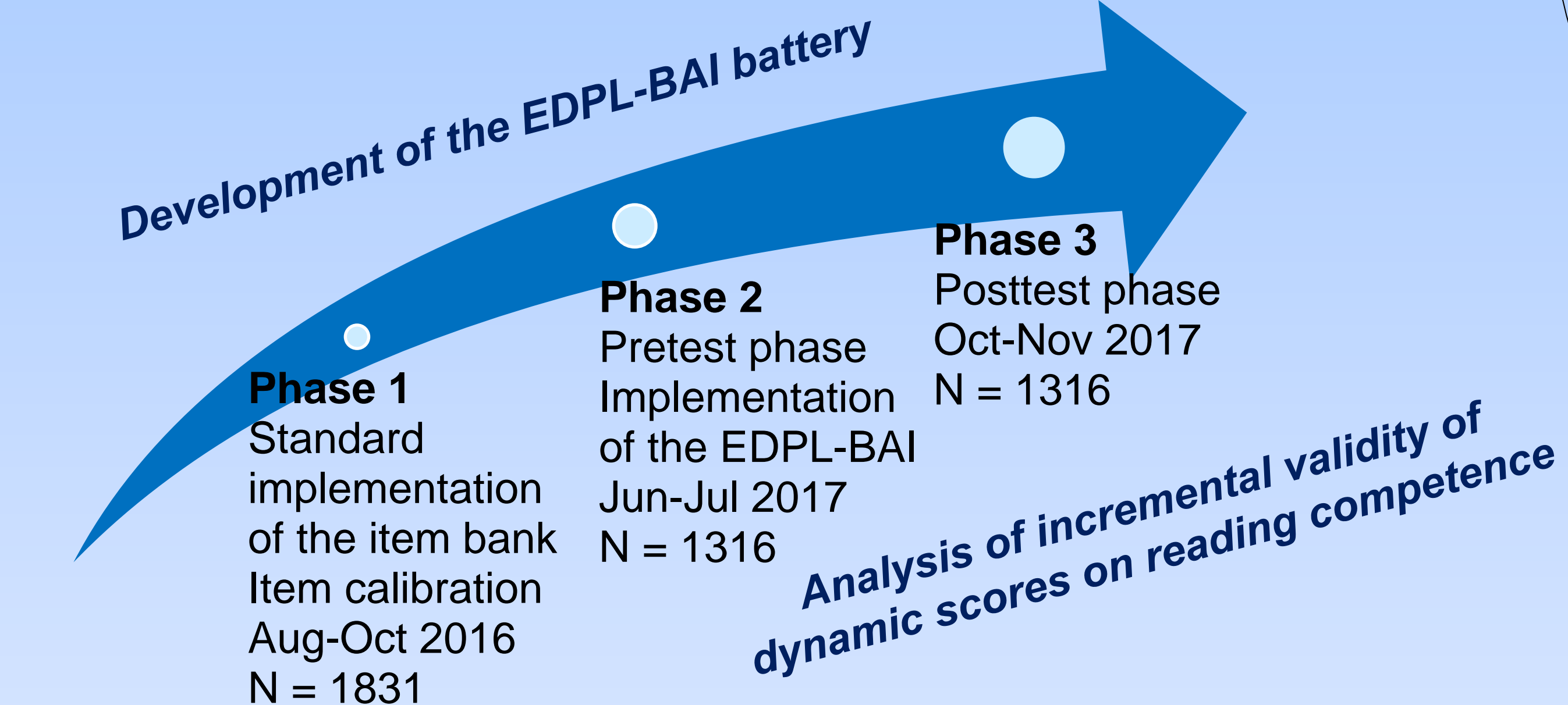


Fig. 2. Research framework

The EDPL-BAI battery was completed on a computer and was supported by the automatic evaluation web platform Siette (Conejo et al., 2016). The administration of tests was collectively carried out, in the usual educational context of the students. Each student received 8 sessions (45 to 75 min.): 1 session for the CLPT pretest, 1 session for the Raven test, 1 session for the EDPL-BAI pretest, 4 sessions for the EDPL-BAI battery, and 1 session for the EDPL-BAI posttest. After 4-5 months, each student received two tests: the CLPT posttest, and the EDPL-BAI posttest. A total of 12 teachers collaborated in the completion of the rating scales on reading performance.

Design and data analysis
 A correlational research design based on causal models was proposed. The dynamic scores were obtained from the implementation of the tests of the EDPL-BAI battery. First, the *student's knowledge level* was estimated from the previous items calibration process (Calibrated Scores). Second, the dynamic score based on the inverse of the value of the required aids to successfully solve the items performed (Inverted Dynamic Scores). The model included four potential predictive variables: a) Dynamic Assessment (DA) factor made of the dynamic scores from MS, OF and CS tests; b) Raven test; c) the EDPL-BAI pretest, and d) the CLPT pretest. All these variables are related to each other. Then the predicted variables were: a) CLPT posttest; b) EDPL-BAI posttest and; c) Teachers' assessment of reading performance (TARP). All the predicted variables were assumed to be correlated. Using this model as a template, two different models were explored. In Model 1, the DA factor was made of the calibrated scores, and Model 2 explored the dynamic scores based on the inverse of the value of the required aids. For the two models, the non-significant paths were deleted sequentially.

Results

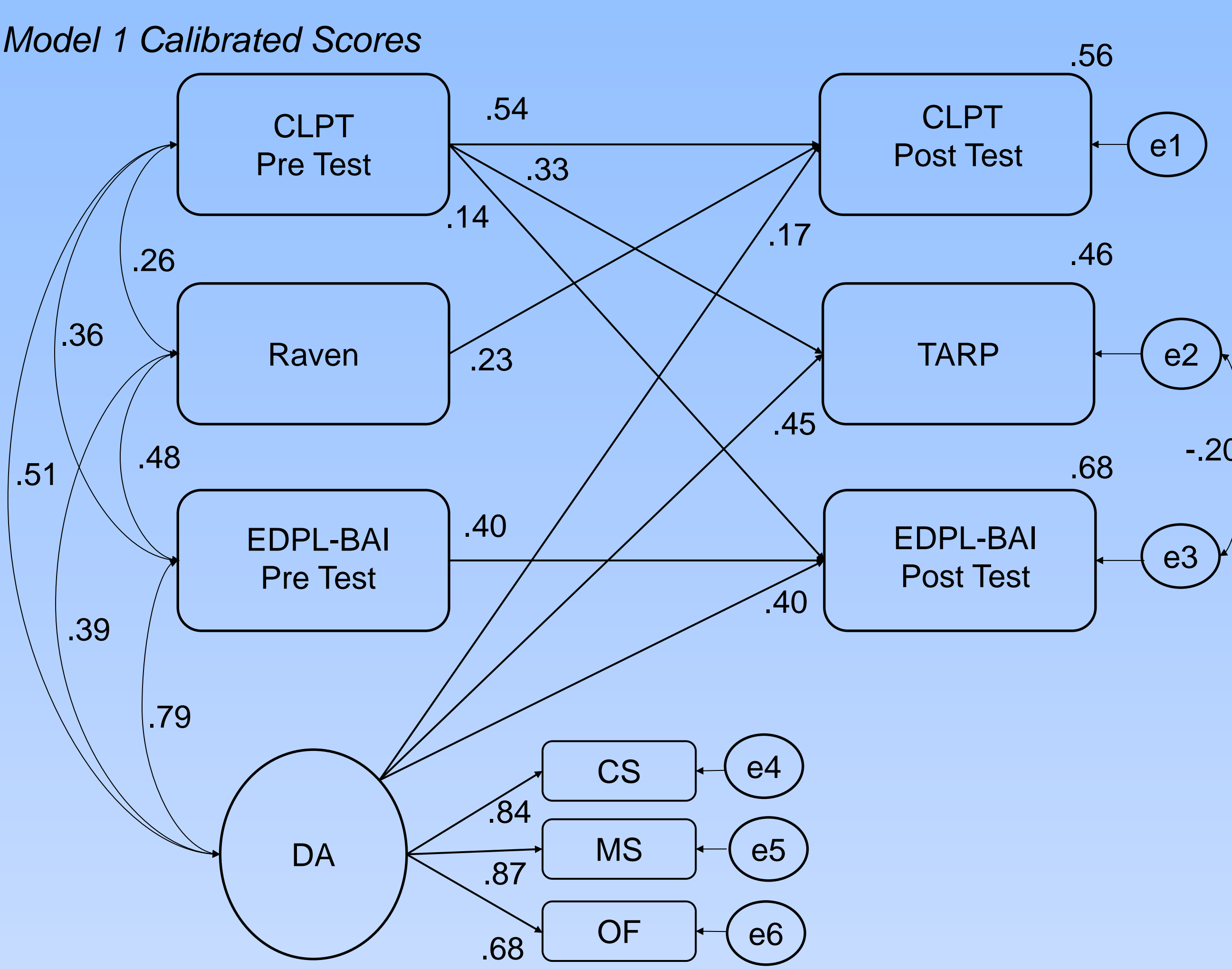


Fig. 3 Standardized coefficients Model 1: Calibrated Scores
 $\chi^2 (18) = 23.830$, $p = .161$; RMSEA = .032; CFI = .983; TLI = .966; PRATIO = .500; PCFI = .492

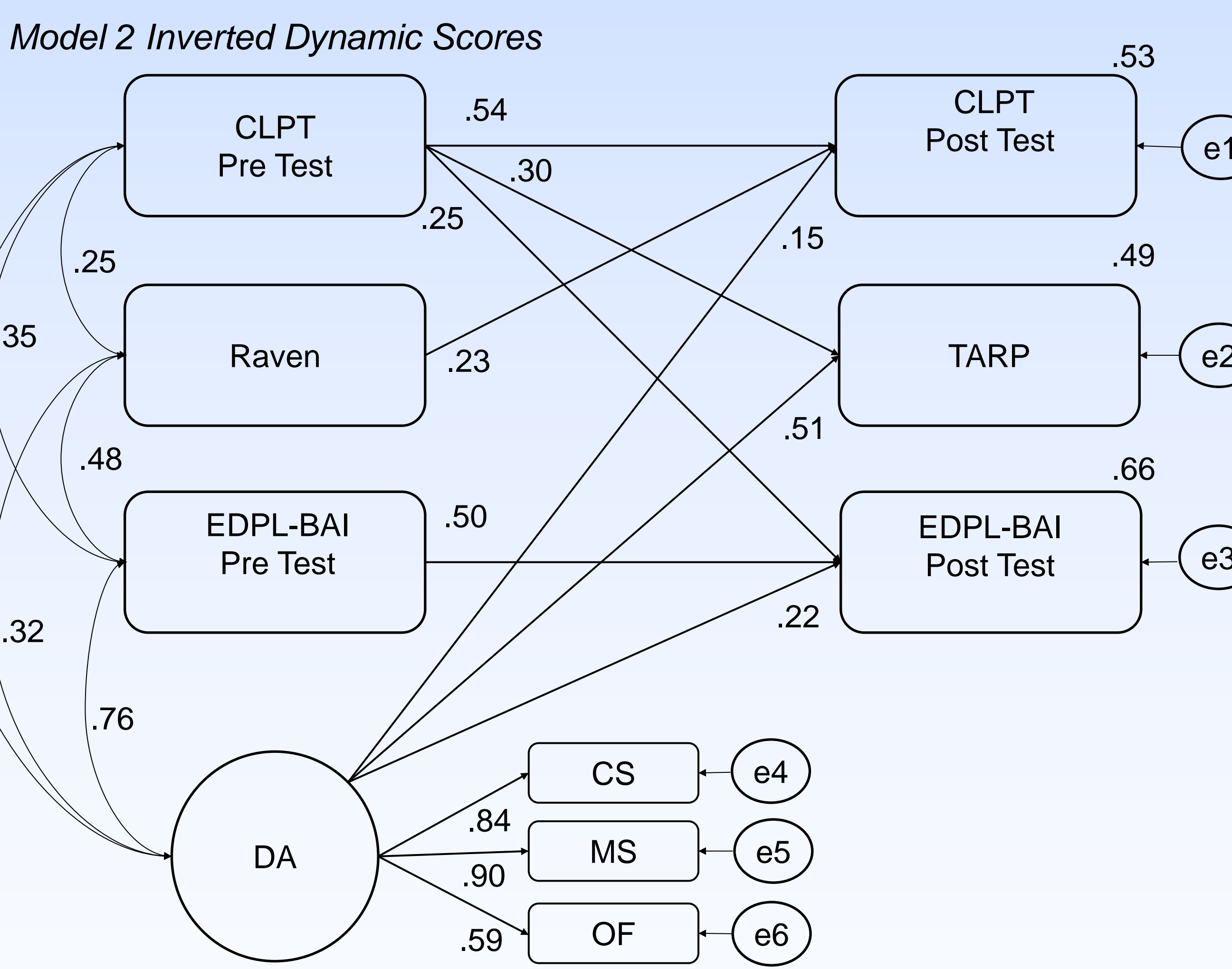


Fig. 4 Standardized coefficients Model 2: Inverted Dynamic Scores
 $\chi^2 (19) = 28.562$, $p = .073$; RMSEA = .039; CFI = .972; TLI = .948; PRATIO = .528; PCFI = .513

Note: CFI = Comparative Fit Index; TLI = Tucker-Lewis coefficient; RMSEA = Root Mean Square Error of Approximation; PRATIO = Parsimony Ratio; PCFI = Parsimony fit to the CFI.

The results included in this poster are part of the FONDEF project ID15110067 "Design and initial validation of a Computerized Adaptive Battery for the assessment of processes involved in reading and comprehension of texts", funded by the National Commission of Scientific and Technological Research, CONICYT (Chile).

Results

The results achieved in the structural equation modeling indicate a good global fit. Individual relationships show significant effects of estimated knowledge level on reading competence, as well as dynamic scores based on the assigned value of graduated prompts required by the students. The dynamic version showed incremental validity over the non-dynamic version. Furthermore, graduated prompts predicted reading comprehension performance.

Model 1 (Fig. 3) explains 56% of the variance in reading comprehension, (CLPT Post Test), 46% of the variance in reading performance measured by the teacher's assessment, and 68% of the variance of EDPL-BAI Posttest. Model 2 (Fig. 4) explains 53% of the variance in the posttest of CLPT, 49% of the variance in reading performance measured with the teacher's assessment, and 66% of the variance in the EDPL-BAI posttest.

Discussion

Regarding our hypothesis, the results show that the dynamic scores obtained from the application of the EDPL-BAI battery further explain the variability in reading competence as measured with the CLPT test, the EDPL-BAI posttest, and the teacher's assessment of reading performance. In this sense, the analysis of the regression coefficients of the model's standardized solution indicates that the dynamic application of the tests maintains a significant and incremental effect on the three measures of reading competence once the rest of the predictor variables are controlled for. This was observed for both the estimated student's knowledge level and the dynamic score obtained from the inverse of the value of the required aids.

A part of the variance of the criterion measures can be explained as a result of the information that is derived from the application of the dynamic tests. In this sense, an analysis of the elements that can explain the changes could offer valuable information about the functioning of the subject. In particular, in the context of dynamic assessment, this analysis of change is aimed at establishing what the subject is capable of performing when offered guidelines and graduated prompts—that is, informing us of his or her learning potential (King et al. 2015; Poehner et al. 2015).

The implementation of DA tests would have provided valuable information regarding the process followed by the students during the task resolution. This information would contain, in our case, data on the aids that were most effective in successfully resolving the different items, which might be useful in terms of understanding the difficulties and the ways of intervening to resolve them.

References

Conejo, R., Guzmán, E., & Trella, M. (2016). The SIETTE Automatic Assessment Environment. *International Journal of Artificial Intelligence in Education*, 26(1), 270-292.
 Embretson, S., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
 King, M., Binger, C., & Kent-Walsh, J. (2015). Using Dynamic Assessment to evaluate the expressive syntax of children who use augmentative and alternative communication. *Augmentative and Alternative Communication*, 31(1), 1-14.
 Poehner, M.E., Zhang, J., & Lu, X. (2015). Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation. *Language testing*, 32(3), 337-357.
 Shapiro, E.D. (2012). *New thinking in response to intervention: A comparison of computer-adaptive tests and curriculum-based measurement within RTI*. Wisconsin Rapids: Renaissance Learning, Inc.