Dynamic computerized-adaptive assessment of reading processes: Incremental validity on text comprehension

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) (Embreton & Reise, 2000). The distinguishing characteristic of CATs is that the sequence of items on a test is determined by 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.

Proposed research study

This study is part of a large comprehensive research study. The research aims to develop a computerized adaptive 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 level integrated into the EDPL-BAI battery. The analysis was conducted using a structural equation model 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 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 first 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 (197) grades and aged between 8 and 12 years old (M = 10.27, SD = 1.22).

Instruments

Tests of the EDPL-BAI battery: Dynamic Morpho-syntactic Awareness Test (MS), Dynamic Syntactic Awareness Test – To sort disordered sentences (QS), and Dynamic Syntactic Awareness Test (CS). Criteria measures: Reading Comprehension Tests, CLPT, Pretest and Posttest EDPL-BAI, Test of Raven’s progressive matrices, and Teacher’s 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).

Results

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 months, the students 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 competence.

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 test score. The correlation analyses were 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 these variables were related to each other. Using this model as a template, two different models were explored. 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.

Discussion

Regarding the 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, the EDPL-BAI posttest, and the teacher’s assessment of reading performance. In this sense, the analysis of the regression coefficients of these variables indicates that the standardized solution indicates that during the 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.

There is a part of the variance of the criterion measures can be explained as a result of the dynamic tests that is not accounted for by the static tests. In this sense, an analysis of the elements that can explain the components of the model 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 can do during the test, what are the guidelines and what are the grades—that is, informing us of his or her learning potential (King et al. 2010; 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


