Measuring Early Spanish Literacy: Factor Structure and Measurement Invariance of the *Phonological Awareness Literacy Screening for Kindergarteners* in Spanish (*PALS español K*)

Francis L. Huang

University of Missouri

Karen L. Ford, Marcia Invernizzi

University of Virginia

Xitao Fan

University of Macau

April 30, 2013

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A090015 to the University of Virginia. The opinions expressed are those of the authors and do not represent views of the Institute or the Department of Education.

Abstract

We investigated the latent factor structure of the *Phonological Awareness Literacy Screening for Kindergarteners* in Spanish (*PALS español K*). Participants included 590 Spanish-speaking, public-school kindergarteners from five states. Three theoretically-guided factor structures were measured and tested with one half of our sample using confirmatory factor analytic (CFA) methods. We then replicated the best fitting model with the second half of our sample and used multigroup CFA (MGCFA) to test for the model's configural and metric invariance. MGCFA revealed that PALS español K exhibited measurement invariance when comparing populations of boys and girls. Our results support the educational utility of PALS español K as a tool for assessing important reading constructs and informing early interventions across gender groups with Spanish-speaking students.

Keywords: measurement invariance; PALS español; hierarchical CFA; literacy assessment; Spanish

Measuring Early Spanish Literacy: Factor Structure and Measurement Invariance of the *Phonological Awareness Literacy Screening for Kindergarteners* in Spanish (*español-K*)

Objectives & Purpose

Alphabet knowledge (AK), phonological awareness (PA), invented spelling, and concept of word in text (i.e., the ability to match speech to print) are important precursors to early literacy. The current study examined the relationship among these skills by exploring the factor structure of a newly-developed early literacy assessment in Spanish, *PALS español K* (Ford & Invernizzi, 2014). To do this, we tested three competing factor models. The first model explored the assumption that all of the foundational skills in early literacy formed a single factor (e.g., Perney, Morris, & Carter, 1997). The second model was based on research that provided evidence for a correlated two-factor model defined by AK and Print/Phonological Awareness (Townsend & Konold, 2010). The third model was a hierarchical model (Marsh, 1987) in which a single second-order Early Literacy factor influenced three first-order factors: AK, PA, and contextual knowledge (consisting of invented spelling and concept of word). The second objective of the study was to evaluate the configural and measurement invariance of the best fitting model to determine whether it exhibited the same psychometric properties when used with different populations (Keith et al., 1995), in this case, boys and girls.

Theoretical Framework

PA is the ability to attend to the sounds of language without attention to meaning. A large body of research has shown the beneficial association of PA to later reading achievement, both in English and in Spanish (e.g., Kirby, Parilla, & Pfeiffer, 2003; Manis, Lindsey, & Bailey, 2004). AK refers to the ability to recognize the names and sounds associated with written letters of the alphabet. AK in young children has also been consistently shown to be one of the strongest predictors of future reading ability in English and in Spanish (Adams, 1990; Bravo, Villalón, & Orellana, 2006; Lindsey, Manis, & Bailey, 2003).

Although PA and AK are important foundational literacy skills, they are also constrained skills that tend to have a ceiling effect in kindergarten assessments. It is likely that other skills, referred to as Contextual Knowledge (CK), may be stronger predictors of later reading (Invernizzi, Justice, Landrum, & Booker, 2004). These skills include invented spelling and concept of word in text (COW).

Researchers have demonstrated that children's spelling inventions are phonetically motivated and reveal varying degrees of phonological awareness and alphabet knowledge (Read, 1986). To invent a phonetic spelling, children must analyze the speech stream and choose letters that represent the speech sounds they extract. Although their selections may not be conventionally correct, researchers have demonstrated the phonetic logic behind such spelling as HKN for *chicken* (using the letter name "aitch" to represent the /j/ sound at the beginning of that word) in English and BADO for *pato* (using phonemes with the same point of articulation to represent the /p/ and /t/ sounds) in Spanish. Research in English has shown that invented spelling is one of the best predictors of subsequent reading (Speece, Ritchey, Cooper, Roth, & Schatschneider, 2004).

To attain a concept of word, children must have a certain degree of PA, automatic alphabet and letter sound recognition, as well as other concepts related to navigating print. Achieving a COW involves the contextual application of students' letter recognition, knowledge of letter sounds, and ability to isolate beginning consonant sounds within printed contexts. Children with a firm COW are also able to recognize words from the memorized text that are presented in isolation in random order, and this is the beginning of conventional reading and spelling (Flanigan, 2007). Like spelling, COW may be viewed as both as an important literacy outcome and a predictor of future reading ability (Morris et al., 2003).

Methods

Participants

PALS español K was administered to 590 Spanish-speaking kindergarteners (50.7% boys) from 29 public schools across five states (M_{AGE} = 5.54, *SD* = 0.38). Forty-five percent of the students were from Virginia, 25% from California, 18% from Missouri, 9% from North Carolina, and 3% from Minnesota. Ninety-five percent of the students had no identified disabilities and 76% were Hispanic, 9% were Non-Hispanic Black, 7% were Non-Hispanic White, and 8% were classified as other or unknown race/ethnicity.

Data Collection Procedures

PALS español K was administered by both teachers and trained assessors in the fall of 2011. Teachers learned how to administer the assessment by completing an online training/certification module. Item-level student scores were entered into an encrypted, password-protected database via the Internet.

Measures

PALS español K is an untimed, criterion-referenced assessment that has six subtests that can be combined to form an overall summed score which is used to identify if a student is at risk for future reading difficulty. Internal consistency reliability of the measures were found to be acceptable for the total score (Cronbach's alpha = .80) and for the six subtests (Cronbach's alpha range = .81 - .96 across subtests). Descriptions of the six subtests are provided below.

Rhyme Awareness. Children are shown a target picture next to a row of three additional pictures and are asked to identify the picture that rhymes with the target (maximum score = 10).

Beginning Sound Awareness. Similar to the Rhyme Awareness task, children are shown a target picture next to a row of three additional pictures and are asked to identify the picture that shares the same beginning sound as the target (maximum score = 10).

Alphabet and Digraph Recognition. Children are asked to name 29 lower-case Spanish letters and digraphs presented in random order (maximum score = 29).

Letter-Sound Knowledge. Children are presented with 25 upper-case letters (excluding Q and X and the letter M which is used as a practice item) and are asked to say the sound the letter/digraph makes (maximum score = 25).

Invented Spelling. Children are asked to write five short, high-frequency words (i.e., un, mi, mesa, pino, and luna). Words are scored based on the number of phonemes represented with phonetically acceptable choices, and a bonus point is awarded for complete correct spelling (maximum score = 21).

Concept of Word. Prior to administering the task, the teacher helps the child memorize a short rhyme using picture supports (i.e., pictures representing each line of the rhyme). Once the child can recite the poem from memory while looking at the picture supports, the child is shown shown the written version and asked to touch each word while reciting the rhyme. After the fingerpoint reading, the child is shown a randomly sorted list of 10 words from the rhyme and is asked to identify each one (maximum score = 10).

Data Analysis

The first phase of the study evaluated three theoretically competing structures for the *PALS español K* subtests (see Figure 1). The total sample was randomly split into two samples. CFA was employed to evaluate the overall and relative fit of the three competing models in the first subsample. Thereafter, the best fitting model was replicated in the second subsample.

Several measures of fit were employed to evaluate the model quality. The traditional χ^2 was used to evaluate model fit. In addition, four measures of fit were considered in evaluating stand-alone model quality: the Tucker-Lewis Index (TLI), the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Higher values on the first two indices (above .95) are suggestive of better fitting models (Hu & Bentler, 1999). In contrast, smaller RMSEA and SRMR values are reflective of better fitting models, with values < .08 considered reasonable and values > .10 considered poor (Brown & Kudeck, 1993).

To compare the three competing models, Akaike's information criterion (AIC) and the Bayesian information criterion (BIC), were evaluated. Smaller values are reflective of better fitting models (Schreiber, 2008).

The second phase of analysis involved examining multigroup invariance between boys and girls in the context of the best fitting model identified earlier. Subsample two was employed for this purpose. Multigroup CFA (MGCFA) to was used to compare the resulting factor structure between groups through examination of both configural invariance (CI), where the same pattern of fixed and free factor loadings is specified for each group, and metric invariance (MI). Our primary interest was in evaluating whether the factor loadings linking the subtests to the respective factors were statistically indistinguishable between groups (i.e., boys and girls), and if supported would indicate that the factor scores are comparable between groups in the sense that they are measured in the same way with the same degree of accuracy (Milfont & Fischer, 2010).

To test for MI we used the χ^2 difference test to measures changes between the CI and MI models (Aderson & Gerbling, 1988). We also evaluated CFI changes (ΔCFI) between the CI and

MI models. Non-significant changes in chi-square ($\Delta\chi^2$) and ΔCFI < .01 can be taken as evidence to support MI (Chen, 2007; Cheung & Rensvold, 2002).

All analyses were conducted with Mplus 6.1 (Muthén & Muthén, 2010) using a maximum likelihood mean-adjusted estimator that is robust to non-normality (Finney & DiStefano, 2006) and used Satorra-Bentler (2001) scaled χ^2 values.

Results

Means, standard deviations, and correlations between the *PALS español K* subtests are presented in Table 1. All of the subtests were moderately to strongly correlated. Fit statistics for the three competing models are presented at the top of Table 2. Results failed to provide strong support for the one factor solution comparison (e.g., RMSEA = .11) or the two-correlated factor model. The hierarchical model was found to demonstrate both good stand-alone fit (TLI = .99, CFI = .995, SRMR = .01, RMSEA = .04), and provided comparatively better fit (AIC = 9,519, BIC = 9,597) when contrasted to the other models. The hierarchical model's χ^2 was also not significant, $\chi^2(6) = 9.44$, $p = .15$, indicating that the data fit the model well. Replication of this model in the second subsample revealed an excellent fit using the traditional χ^2 measure and the alternative fit indices.

Multigroup invariance between boys and girls was investigated in the context of the best fitting hierarchical model. Results of the configural invariance test revealed good model fit (e.g., RMSEA = .00). Between group equality constraints imposed on the factor loadings (metric invariance) resulted in good model fit (TLI = 1.0, CFI = 1.0, SRMR = .05, RMSEA = .00). Further, both the χ^2 difference test, using the modified SB χ^2 difference testing procedure and the ΔCFI failed to demonstrate a statistically significant decline in fit with the additional equality constraints imposed on the factor loadings ($\chi^2_\Delta(\Delta df = 5) = 7.41$, $p = .19$ and ΔCFI < .001). As a result, factor loadings among the factors and tasks were not statistically different between boys and girls.

Discussion and Significance of Study

The result of the hierarchical CFA using the *PALS español K* showed a good fit with the data. All of the first-order factor loadings suggest that six tasks are all good measures of their corresponding factors. Empirical evidence supporting the use of the six tasks is important since when scoring the *PALS español K*, scores from all of the tasks are summed together to create an overall score.

Figure 2 shows that the three corresponding latent factors are strongly affected by the second-order Early Literacy factor. Our findings support that one general factor, Early Literacy, gives effects to the other first-order factors. Theoretically, early literacy development in Spanish appears to be highly associated with contextual knowledge in addition to the more constrained skills of AK and PA.

Our finding of measurement invariance between boys and girls is relevant and provides further evidence that the construct of Early Literacy is being measured the same way between groups. As a result, the same dimensions of PA, AK, and Contextual Knowledge can be applied to both student populations.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Bravo Valdivieso, L., Villalón, M., & Orellana, E. (2006). Diferencias en la predictividad de la lectura entre primer año y cuarto año básicos [Differences in the predictability of reading between first grade and fourth grade]. *Psykhe, 15*.

Flanigan, K. (2007). A concept of word in text: A pivotal event in early reading acquisition. *Journal of Literacy Research, 39*, 37-70.

Ford, K. & Invernizzi, M. (2014). *Phonological Awareness Literacy Screening for Kindergarteners* in Spanish (*PALS español K*). Charlottesville, VA: University of Virginia.

Invernizzi, M., Justice, L., Landrum, T., & Booker, K. (2004). Early literacy screening in kindergarten: Widespread implementation in Virginia. *Journal of Literacy Research, 36*, 479-500.

Keith, T. Z., Fugate, M. H., DeGraff, M., Dimond, C. M., Shadrach, E. A., & Stevens, M. L. (1995). Using multi-sample confirmatory factor analysis to test for construct bias: An example using the K-ABC. *Journal of Psychoeducational Assessment, 13*, 347-364.

Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology, 95*, 453-464.

Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first-grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology, 95*, 482-494.

Manis, F. R., Lindsey, K. A., & Bailey, C. E. (2004). Development of reading in grades K-2 in Spanish-speaking English-language learners. *Learning Disabilities Research & Practice, 19*, 214-224.

Marsh, H. (1987). The hierarchical structure of self-concept and the application of hierarchical confirmatory factor analysis. *Journal of Educational Measurement, 24,* 17-39.

Morris, D., Bloodgood, J., & Perney, J. (2003). Kindergarten predictors of first- and second-grade reading achievement. *The Elementary School Journal, 104,* 93-109.

Perney, J., Morris, D., & Carter, S. (1997). Factorial and predictive validity of first graders' scores on the Early Reading Screening instrument. *Psychological Reports, 81,* 207-210.

Read, C. (1986). *Children's creative spelling*. London: Routledge & Kegan Paul.

Speece, D. L., Ritchey, K., Cooper, D. H., Roth, F. P., & Schatschneider, C. (2004). Growth in early reading skills from kindergarten to third grade. *Contemporary Educational Psychology, 29*, 312-332.

Townsend, M., & Konold, T. (2010). Measuring early literacy skills: A latent variable investigation of the Phonological Awareness Literacy Screening for Preschool. *Journal of Psychoeducational Assessment, 28,* 115-128.

Table 1

PALS español K Means, Standard Deviations, and Correlations for Confirmatory (Upper Diagonal) and Exploratory (Lower Diagonal) Samples

Variable	RA	BS	ABC	LS	SP	COW
Rhyme Awareness (RA)	1.00	.57	.44	.42	.49	.33
Beginning Sounds (BS)	.52	1.00	.53	.54	.60	.34
Alphabet Recognition (ABC)	.34	.50	1.00	.67	.59	.38
Letter Sound Knowledge (LS)	.40	.49	.60	1.00	.67	.39
Invented Spelling (SP)	.43	.55	.57	.64	1.00	.48
Concept of Word (COW)	.37	.36	.46	.48	.60	1.00
Exploratory sample (n = 296)						
M	5.99	5.87	9.43	11.51	5.56	1.03
SD	2.90	2.96	9.29	7.04	5.27	2.03
Confirmatory sample (n = 294)						
M	5.87	5.85	10.37	11.80	5.61	1.02
SD	2.96	3.06	9.50	7.27	5.40	1.86

Note. PALS español K = Phonological Awareness Literacy Screening for Kindergarteners in Spanish.

All correlations are statistically significant, all *ps* < .001.

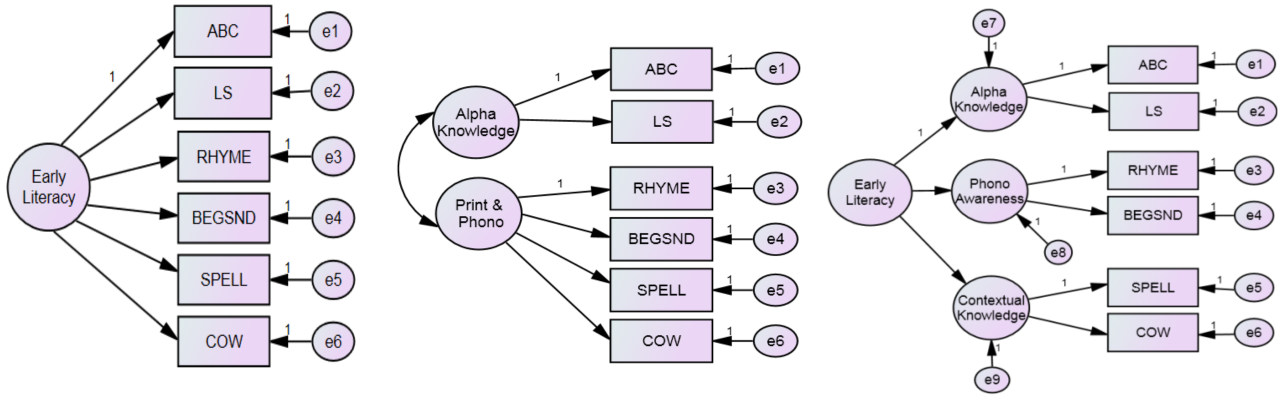
Table 2

Model Fit Indices

Model	Model description	χ^2	df	SRMR	TLI	CFI	RMSEA	AIC	BIC
	Competing PALS español K Models (Subsample One *n* = 296)								
A	One factor	40.98*	9	.039	.922	.953	.110	9546	9612
B	Two correlated factors	36.30*	8	.036	.922	.958	.109	9541	9611
C	Hierarchical factor	9.44	6	.016	.987	.995	.044	9519	9597
	Replication and Invariance Testing (Subsample Two *n* = 294)								
D	Replicated hierarchical model	7.44	6	.014	.995	.998	.029	9438	9515
E	Configural invariance	9.59	12	.016	1.008	1.000	.000	9457	9612
F	Measurement invariance	16.92	17	.045	1.000	1.000	.000	9453	9590

Note. All χ^2s are Satorra-Bentler (2001) scaled χ^2 which account for non-normal data. SRMR = standardized room mean square residual. TLI = Tucker-Lewis index. CFI = comparative fit index. RMSEA = root mean square error of approximation. AIC = Akaike's information criterion. BIC = Bayesian information criterion.

* $p < .001$.



Model A:
One-factor

Model B:
Two-correlated factors

Model C:
Hierarchical factor

Figure 1. Three factor models to be tested.

Note. ABC = alphabet and digraph recognition. LS = letter sound knowledge. RHYME = rhyme awareness. BEGSND = beginning sounds awareness. SPELL = spelling. COW = Concept of Word. Alpha = alphabet. Phono = phonological.

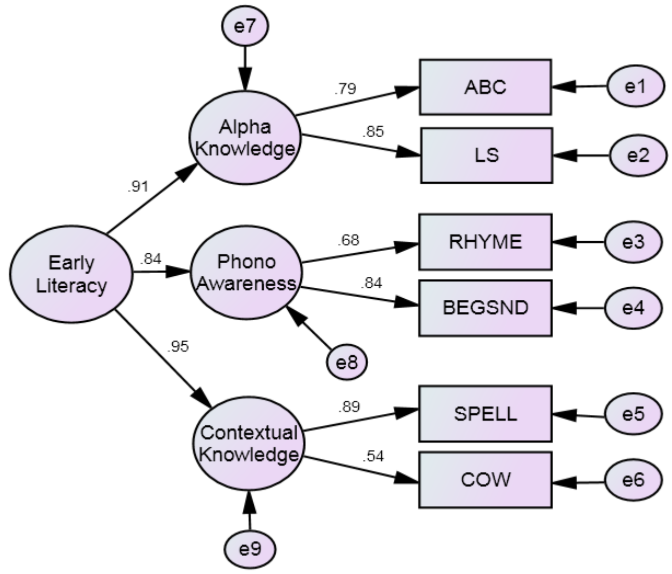


Figure 2. Standardized factor loadings for model D (n = 294).

Note. ABC = alphabet and digraph recognition. LS = letter sound knowledge. RHYME = rhyme awareness. BEGSND = beginning sounds awareness. SPELL = spelling. COW = Concept of Word. Alpha = alphabet. Phono = phonological.