Theory-Based Parameterization of Semiotics for Measuring Pre-literacy Development

N Bezruczko
Measurement and Evaluating Consulting
1524 E. 59th Street, Chicago, USA

Email: Nbezruczko@msn.com

Abstract. A probabilistic model was applied to problem of measuring pre-literacy in young children. First, semiotic philosophy and contemporary cognition research were conceptually integrated to establish theoretical foundations for rating 14 characteristics of children's drawings and narratives ($N = 120$). Then ratings were transformed with a Rasch model, which estimated linear item parameter values that accounted for 79 percent of rater variance. Principle Components Analysis of item residual matrix confirmed variance remaining after item calibration was largely unsystematic. Validation analyses found positive correlations between semiotic measures and preschool literacy outcomes. Practical implications of a semiotics dimension for preschool practice were discussed.

1. Introduction
Semiotics is the study of symbols and signs, which is directly linked to linguistics and has implications for language development, literature, and culture [1]. Relations between linguistic signs and knowledge have been studied since antiquity, and the Greeks considered signs instrumental to mental activity [2]. Augustine and Roger Bacon [3] continued an interest in semiotics in Middle Ages.

In this research, ideas of Charles Peirce, a 19th century American Pragmatist, were conceptual foundations for developing a statistical semiotic construct that models early literacy of three and four year old children. Peirce emphasized fundamental relations between semiotics and logic, while his concepts of icons, indices, and symbols were inter-related in a process of semiosis [4]. Many of Peirce's ideas re-appear in Luquet [5], and again in Piaget and Inhelder [6], as well as Bruner [7]. Semiotics, however, is not considered an integrated body of knowledge and efforts to measure language growth and learning does not appear in its literature. Despite ancient Western foundations, semiotics does not have a prominent empirical tradition or recognized body of cumulative knowledge.

Dual Coding Theory (DCT) [8, 9] is a cognition model that complements semiotics by emphasizing two symbolic systems, language (verbal) and mental imagery (nonverbal), which influence literacy development. Moreover, DCT provides a perspective on cognitive mechanisms that may underlie semiotic development in young children, and Figure 1 presents a schematic model. A key aspect of this model is referential linking between verbal and nonverbal systems, which has implications for comprehension and memory. A central DCT hypothesis in this research is semiotic function develops in conformity with a hierarchy defined by sensorimotor, iconic, and abstract symbolic components. DCT predicts children enrolled in preschools that emphasize simultaneous verbal and nonverbal processing will show higher semiotic function, as well as higher early literacy outcomes.
In Dual Coding Theory verbal and nonverbal symbol systems are linked by referential links during cognition. Panel A presents an overall schematic, while Panel B presents simultaneous activation of both systems for the concept "royal wedding" [10].

2. Problem
Researchers have long suspected early childhood language is associated with a semiotic process [11]. A problem has been lack of an empirical construct for measuring pre-literacy on a linear scale because semiotics has a weak empirical tradition, and its conceptual foundations are notoriously fragmented.

3. Objective
The objective of this research was to measure young children's acquisition of pre-literacy with a probabilistic model that is grounded in semiotics developed by Peirce and extended by others. This goal faced two major challenges. First, disparate semiotic ideas about signs and symbols needed to be consolidated into a coherent overarching construct. Then a measurement model needed to be implemented to demonstrate linear properties of this construct for measuring pre-literacy.

4. Method
The following steps were implemented to accomplish above objectives:

- standardize a procedure for collecting drawings and narratives from young children
- identify characteristics of child samples that represent semiotic function
- develop a rubric for quantifying characteristics of child drawings and narratives
- collect and rate child samples at Time 1 (fall) and Time 2 (spring)
- parameterize ratings with a probabilistic model and conduct validation analyses

4.1 Sample
Thirty preschools were selected from Chicago Public Schools (CPS), and at least four children were randomly selected from each preschool \((N = 120)\). These preschools are demographically multi-ethnic, socio-economically disadvantaged, and many children were English language learners.

4.2 Data
Drawings, narratives, and standardized preschool outcomes were collected of children, while classrooms and teachers were evaluated with standardized observation and interview protocols.
4.2.1 Child samples and standardized assessments. A standard procedure was developed to collect authentic samples of child drawings and narratives. More specifically, teachers led a thematic storybook presentation, which emphasized "sharing" followed by whole group discussion. Instructions then were given for an "imagery" exercise, which requested children to close their eyes and imagine how they might experience sharing at home, in their neighborhoods, or at school. Finally, children formed small groups to talk about sharing and were given materials to draw about it.

4.2.2 Data collection. Authentic drawings and narrations were collected in preschools at Time 1 (fall) and Time 2 (spring). An adult transcribed child narratives. CPS provided standardized preschool outcomes.

4.2.3 Scoring rubric for drawings and narratives. Fourteen rating scale items with six categories (0, 1, 2, 3, 4, & 5) were derived from Peirce’s triadic model (icons, indices, and symbols), as well as cognitive and developmental theory. Each child sample was rated by at least three trained evaluators.

4.2.4 Parameterization, analysis, and validation. Child ratings were transformed with a Rasch model for rating scales, which is summarized below and was implemented with Winsteps software [12].

\[
\Pi_{nix} = \frac{\exp \sum_{j=0}^{x} [\beta_n - (\delta_i + \tau_j)]}{\sum_{k=0}^{m} \sum_{j=0}^{k} \exp \sum_{j=0}^{x} [\beta_n - (\delta_i + \tau_j)]}
\]

where \(\beta\) = observations, \(\delta\) = item difficulties, and \(\tau\) = rating scale thresholds. \(\Pi_{nix}\) is probability any item \(\delta_i\) will be rated X by participant \(\beta_n\) where X takes a value from a fixed range (j = 1, 2, 3, 4, & 5), m = number of steps for an item, and k = ith step.

5. Results
5.1 Principle Components Analysis
Thirty evaluators provided ratings and their agreement was high (> .90). Figure 2 presents examples of child drawings and narratives. Preliminary Principle Components Analysis (PCA) of raw data identified a single prominent factor (eigen value = 7.0), which accounted for 47 percent of raw data. Aggregation of three largest remaining factors only accounted for ~20 percent of variance.

5.2 Model fit, measurement properties, and residual analysis
With exception of two items, model fit was excellent. Likewise, all rating scale category thresholds were ordered, and conventional psychometric reliability was 0.84 (alpha). Some drawings, however, were unusual because children included both sophisticated and primitive characteristics such as Scribbling, Stick figures, and Tadpoles, which led to inconsistent ratings. In order to clarify dimensionality, PCA was conducted of item residuals. Results indicated calibrated dimension accounted for 79 percent of overall ratings, while only 3.2 percent of residual variation was systematic hence associated with unexpected factors that could threaten dimensionality. Figure 3 presents a plot of parameterized semiotic dimension with model residuals in eigen value units. These results support construct dimensionality. Validity was further investigated by decomposing item difficulties into sensorimotor, iconic, and abstract symbolic processing components, which showed semiotic components were associated with 70 percent of item difficulty variance (\(R^2 = 0.71, F = 13.32, p<.001\)).
Figure 2. A hierarchy of 14 characteristics was identified in child samples collected in fall and spring, then parameterized with a Rasch model. Oral narration was generally easier than drawing, while spatial relations, human figure details, and coherent graphic expression were more difficult.

Figure 3. Residuals falling outside the measurement zone (factor loadings > +/- 0.3) indicate violation of local independence, which may threaten unidimensionality. PCA, however, indicated only 3.2 percent of residual variance was associated with nonrandom residual structures. Most item estimates follow an expected vector trajectory and their residuals are random. However, this residual cluster presents positive inter-correlations. Items are:

2. Are figures and objects easy to identify?
4. Does drawing show human figures?
5. Does child emphasize details?

While their disturbance is minor, these items should be monitored in future research.
5.3 Semiotic construct map

Figure 4 presents a calibrated construct map of items and children after ratings were transformed to linear units. Items on right of hierarchy are ordered from low to high difficulty, while linear measures of child ability appear on left side in common units. Drawings with stick figures, tadpoles, primitives, and scribbling on this hierarchy established lowest level of pre-literacy performance. Slightly higher were drawings with human figures and details, as well as those that demonstrated narrative clarity and consistency between narrative and drawing. Drawings at highest level emphasized spatial organization and tended to include letters and names. This progression from low to high is consistent with Peirce's triadic hierarchy of signs, icons, and symbols. Of particular significance, children between fall and spring tended to progress sequentially through this probabilistic structure. Item parameter invariance was confirmed across several criteria: gender, economic status, and home language background.

5.4 Differences between Time 1 and Time 2

Semiotic functioning of most children in this sample was sensorimotor when they enrolled in preschool, and, in general, most children advanced to iconic processing by spring assessment. Very few children reached abstract symbolic functioning. Overall fall mean was -1.04 logits (SD = 2.21, SE = 0.79), while spring mean was -.93 logits (SD = 1.42, SE = 0.60). Parameterized semiotic measures were significantly correlated with early literacy outcomes such as Beginning Sounds and Prints, which supports construct validity. Standardized semiotic gains were largest for younger children in more disadvantaged neighborhoods, mainly because of very low performance at fall enrollment. Surprisingly, three and four year old children showed greater semiotic gain than five year olds, and early literacy outcomes did not significantly differ between older and younger children.

Figure 4. Transformed ratings present items and children in common logits.
6. Discussion
Semiotic hierarchy presented in this research suggests early literacy in preschool involves at least three progressive operations, namely, incoherent scribbling or sensorimotor functioning, physical integration of iconic signs in drawings, and advancement to abstract symbolic expression, which is largely consistent with Peirce's model of semiosis. Speculation suggests several cognitive mechanisms underlie child performance on this construct, that is, a) iconic matching, b) recognizing arbitrary relations between meaning and symbols, and, c) grasping relations between signs and language. Validation analyses supported relations of semiotic performance with kindergarten readiness, as well as DCT predictions that verbal and non-verbal language processing in preschool is related to both semiotic development and higher preschool outcomes. Measurement properties of the parameterized semiotic construct are generally adequate for most practical preschool applications. Further scale development should emphasize addition of more items at highest level of abstract symbol processing, which would extend validity and improve overall practical usefulness.

6.1 Limitations
Five year olds were already at top of parameterized construct at fall enrollment, a formulation that could not capture their gains. Restriction of range was also problematic because this sample was drawn from lower performing urban preschools in economically disadvantaged Chicago neighborhoods. More research is needed with a broader range of pre-literacy development and socio-economic background. Longitudinal studies are also needed to investigate long term effects of semiotic development, as well as influence on child attitude and motivation for learning.

6.2 Implications
These results support convergence of loosely organized semiotic ideas and cognitive science. Practical implication is preschools should exercise caution about emphasizing exclusively verbal learning.

7. Conclusion
Measurement results in this research provide theoretical semiotic foundations for measuring pre-literacy. This construct should be useful for identifying children who have reached abstract symbolic functioning related to early literacy and ready for advancement, as well as performance level of other children. This capacity to identify children who differ in semiotic functioning should have implications for diversifying early literacy pedagogical practices to accommodate developmental needs.

8. References
[1] Caesar M 1999 Umberto Eco: Philosophy, Semiotics (Hoboken, NJ: Wiley-Blackwell)
[2] Romeo L 1977 Semiosis 6 37
[3] Maran T 2003 Sign Systems 31 192
[4] Peirce C S 1931-35 Collected Papers (Cambridge, MA: Harvard University Press)
[5] Luquet G H 1927/2001 Le Dessin Enfantin (Trans. by Alan Costall) (London: Free Association)
[6] Piaget J and Inhelder R 1956 Child's Conception of Space (London: Routledge)
[7] Bruner J 1964 American Psychologist 19 1
[8] Paivio A 1986 Mental Representations: A Dual Coding Approach (NY: Oxford University Press)
[9] Paivio A 1989 A dual coding perspective on imagery and the brain Neuropsychology of Visual Perception ed J.W. Brown (Hillsdale, NJ: Erlbaum) pp 203-216
[10] Sardoski M and Paivio A 2001 Imagery and Text (Mahwah, NJ: Erlbaum)
[11] Dyson A H 1986 Research in the Teaching of English 20 379
[12] Linacre M J 2006 Rasch Measurement Software and Manual (Chicago: MESA Press)

Acknowledgments
I am indebted to Ava Haji and coaches at Doolittle School, as well as Barbara Bowman and Paula Cottone in Office of Early Childhood Education, Chicago Public Schools for supporting this research.