Construct Validity as a Foundation of Evidence-Based Practice: The Case of the Preschool Language Assessment Instrument

La validité conceptuelle comme fondement de la pratique axée sur l'expérience : le cas du Preschool Language Assessment Instrument

Elizabeth Skarakis-Doyle, Linda T. Miller, and Meghan Reichheld
University of Western Ontario
London, Ontario

Abstract

Evidence-based practice requires demonstration of the adequacy and appropriateness of assessment procedures, as well as treatment procedures. The purpose of this paper was to demonstrate the process of obtaining such evidence by using the example of the Preschool Language Assessment Instrument (PLAI). This test is used primarily for the purposes of profiling children's abilities to cope with the language of instruction and then to develop treatment programming based upon the profile. It is based upon a model of instructional language comprised of four distinct levels of abstraction. We approached the task of obtaining evidence for this model, and hence the test, from several perspectives. In the first analysis, the appropriateness of the four levels was assumed. Average performance scores for three age groups of children at each test level were compared. In the second set of analyses, the appropriateness of the model was not assumed and was examined based upon the grouping of items delineated by the model (i.e., the four-level assignment) via factor analysis. Evidence from all analyses led to the conclusion that a single central dimension was being evaluated in the PLAI. The four-level hierarchy of perceptual-language distance was not confirmed in the results. Instead an alternative five-factor solution was suggested. Current practices based upon this model were reexamined based upon the evidence obtained.

Abrégé

La pratique fondée sur l'expérience n'est possible que lorsque sont démontrés la pertinence et le bien-fondé des procédures d'évaluation, ainsi que des procédures de traitement. L'objectif de cette étude était justement de documenter le processus permettant d'obtenir une telle preuve à l'aide de l'exemple de l'instrument d'évaluation linguistique préscolaire. Ce test sert principalement à déterminer les aptitudes des enfants à comprendre la langue d'enseignement, puis à élaborer un programme de traitement en fonction du profil établi. Il se fonde sur un modèle de langue d'enseignement qui comprend quatre différents niveaux d'abstraction. Nous avons envisagé la tâche d'obtenir les preuves pour ce modèle, et donc pour le test, de différentes perspectives. Dans la première analyse, nous avons pris pour acquis la pertinence des quatre niveaux. Les résultats de performance moyens ont été comparés pour les trois groupes d'âge dans chaque niveau. Dans la deuxième analyse, nous n'avons pas pris pour acquis la pertinence du modèle, mais l'avons examiné en fonction du groupement d'éléments délimités par le modèle (exercice aux quatre niveaux) grâce à une analyse factorielle. Les résultats de toutes les analyses ont mené à la conclusion que l'instrument d'évaluation linguistique préscolaire ne permet de jauger qu'une seule dimension centrale. La hiérarchie à quatre niveaux de la distance linguistique perceptive n'a pas été confirmée dans les résultats. Par contre, une autre solution fondée sur cinq facteurs a été proposée. Les pratiques actuelles axées sur ce modèle ont été réévaluées en tenant compte des preuves ainsi obtenues.

Key words: preschool language assessment, contruct validity, language of instruction, language development

Evidence-based practice requires that speech-language pathologists demonstrate the adequacy and appropriateness of their assessment and treatment procedures. Contemporary discussions of evidence based practice focus primarily on obtaining and evaluating available support for treatment programs and procedures. However, demonstrating the appropriateness of our assessment procedures is equally, if not more important, given that assessment data form the foundation for treatment programs. When speech-language pathologists seek evidence for the bases of their assessment tools, they are in fact seeking to validate them. Validity is most commonly thought of as a psychometric property of standardized tests. However, validity actually entails an evaluative judgment of the meaning or interpretation associated with an assessment and any action that results from that interpretation, such as a treatment program (Messick, 1995). The interpretations made from test data are based upon the assumptions clinicians hold about the behavioral domains or constructs that are being assessed. In practice, those assumptions also serve as a foundation for further clinical actions such as treatment programming. For practice to be evidence based, a clinician's assumptions must be justified. If speech-language pathologists are to demonstrate the adequacy and appropriateness of their assessment and treatment procedures, then the constructs underlying these procedures must be made explicit and judged for their logical adequacy and their empirical strength (Anastasia, 1986; Hutchinson, 1996; Messick, 1995). Thus, establishing the construct validity of assessment procedures and the treatment programs developed from them is fundamental to evidence-based practice.

This validation process is often a challenging task, particularly in the area of child language. There is a lack of an accepted "gold standard" for assessing the young child's language abilities and no real consensus exists as to which theoretical constructs should form the bases for our assessment procedures and treatment programs. As a result, the constructs underlying our contemporary assessment tools are not usually well articulated. In the rare case in which a model of language ability is delineated for an assessment tool, an opportunity arises for developing an evidence base not only for the appropriate use of the assessment tool, but also for the model on which it rests. This, in turn, provides the evidence for the treatment decisions that result from the use of such a test.

Although it is now over 20 years old, the Preschool Language Assessment Instrument (PLAI; Blank, Rose, & Berlin, 1978a) provides an excellent illustrative case. The authors based the PLAI upon a specific model for the use of language in the classroom. The PLAI was developed as the clinical version of a scale of discourse cognitive complexity (Blank, Rose, & Berlin, 1978b). As such, the PLAI is one instantiation of that model. It was originally developed to assess the conceptual complexity of the instructional exchanges between early childhood educators and their preschool students. However, researchers use the scale in broader discourse contexts, for example, in investigations of mother-child verbal interactions (Blank & Franklin, 1980; Bradshaw, Hoffman, & Norris, 1998; Cole, Dale, & Mills, 1990; Conti-Ramsden & Friel-Patti, 1986; Lehrer & deBernard, 1987; Sorsby & Martlew, 1991; vanKleek, Gillam, Hamilton, & McGrath, 1997).

Among clinicians, the PLAI is used primarily for purposes of identifying treatment goals and developing home and classroom treatment programs (Carotta, Carney, & Dettman, 1990; Marvin & Wright, 1997; Moeller, Osberger, & Eccarius, 1986; Nelson, 1993; Owens, 1995; Paul, 1995; Shimotakahara & Li, 1998) rather than as a standardized test (Blank, et al., 1978a,b; Klein & Moses, 1994; Skarakis-Doyle et al. 1998). The emergence of collaborative consultation service delivery models and the emphasis upon abstract and critical thinking in contemporary "standards-based" curriculum (Palinscar, Collins, Marano, & Magnusson, 2000) have promoted these treatment applications. The PLAI's model of the language of instruction provides a framework for consultation between speech-language pathologists and teachers when using dialogic mentoring techniques (Nelson, 1993; Paul, 1995). Therefore, adequacy and appropriateness of these clinical applications rest on the validity of the cognitive complexity scale's conceptualization of instructional language and its realization in the PLAI.

Blank et al.'s (1978 b) model of the language of instruction was intended to capture the manner in which the preschool teacher facilitates intellectual development via learning dialogues. It was intended to organize the child's experience through language at "different hierarchical levels" (Blank & Franklin, 1980, p. 129). Moffett's (1968) continuum of abstraction was adapted to reflect the specific aims of early childhood education instruction. The amount of abstraction entailed in a learning dialogue was conceptualized in a continuum of perceptual-language distance or the distance in time and space of a referent from the discussion of it. The least abstract instructional language occurs when the material being discussed (i.e., the perception) is very close to the language used to discuss it. This is represented in Level I "Matching Perception" on the PLAI. At this level, language focuses on the "here and now." According to the authors, responding to requests to name objects, to imitate simple sentences, or to identify pictured objects or incidental information from memory are typical of the discourse formulations at this level. Such demands can often be adequately met with a nonverbal or single word response.

The other anchor of the continuum, the most abstract instructional language, occurs when the perception is removed in time and space and the child must hypothesize about what might happen under a set of particular conditions. This type of formulation is realized in the PLAI's Level IV "Reasoning about Perception." On these tasks, the child must go beyond what he or she can perceive directly, and both discern the relationships among objects and events and formulate explicitly the reasons and logic responsible for these relationships" (Blank et al., 1978b, p.32). Requests to predict, to justify predictions or decisions, to identify causes of an event or to formulate a solution comprise the discourse formulations at this level.

In addition to the two anchors, Blank et al. (1978b) propose two other distinct points along the perceptual-language continuum that lie between the concrete and abstract language described above, Level II "Selective Analysis of Perception", and Level III "Re-ordering Perception." These intermediate points purportedly require the child to respond in an increasingly selective manner. In Level II, language is still closely tied to the perceptual experience found in Level I, but the child is to focus on only one aspect of the material. Scanning for an object defined by its function, describing a scene, recalling items from a statement, sentence completion, and attending to multiple characteristics of functions or identifying differences comprise the demands at this level. In Level III, the child is required to reorganize the material according to specific constraints imposed by the teacher's language. Following sets of directions, assuming the role of another person, describing what would happen next, telling a brief story, identifying similarities, defining words, and class exclusion problems purportedly measure the essential characteristic of this level. Displacement of the language in time is an important distinction between the lower two levels of the model and the two advanced levels.

Although asserting that the four levels of the model are somewhat fluid, Blank et al. (1978b) nonetheless are very clear in maintaining that there is an essential distinguishing core element to each of the four levels of formulations that must be mastered in order for a child to succeed in the classroom. "... the levels are deemed to reflect the 'qualitatively' different demands for abstraction that are placed on the child..." (p. 17). Their intent was to be able to categorize any discourse formulation in a learning dialogue or on the test into one of the four levels (Blank & Franklin, 1980). The manner in which the authors advocate using the model reinforces their assertion that each level is discreet with a distinctive essential characteristic. Blank, et al.(1978b) state:

> Any set of discourse demands that receives less than a fully adequate response in either the test or teaching situation can be taken to represent an area of discourse that the children have not yet mastered. Accordingly, it also represents an area that can, and eventually should be taught (p. 90-91).

Blank et al. (1978b) designed a remedial teaching sequence whereby teachers would adjust learning dialogues when necessary along the four-level hierarchy of their model. Thus, the PLAI is used to assess a child's relative strengths and weaknesses across these four distinct levels. Based on the resulting profile, speech-language pathologists demonstrate to teachers how to scaffold the child's response by first simplifying their instructional question to the level at which the child currently demonstrates skill. Once the child successfully responds, the teacher is coached to "up the ante" by following up with a question at the next highest level on the hierarchy (Blank, et al., 1978b; Paul, 1995). The assessment profile and the treatment sequence are based upon the logical evidence for a four-level hierarchy presented in the authors' model of the language of instruction and assumes its accuracy and appropriateness. If the characteristic feature of each level has not been established, then there is little basis for the proposed sequence and presumably scaffolding is not actually being achieved. Fortunately, with such a well-specified model of the language of instruction and clearly stated purpose, an empirical examination of the construct validity of the PLAI is possible. If the data verify the four-level model, then the clinical actions resulting from the test's scores will be justified.

The authors provide limited empirical evidence for the validity of the PLAI. Content validity, or that aspect of construct validity that addresses the relevance and representativeness of test items (Messick, 1995), was examined by Blank et al. (1978a). Five psychologists and special education teachers were asked to assign test items to the four levels defined in the model following the assumption that each of the items fit appropriately into one of the levels. Raters agreed on the level assignment for 75% of the test items. Expected performance differences over time also were revealed; performance scores improved with age providing additional evidence for construct validity. Further, within each age group (3-, 4- and 5-year-olds), scores decrease from Level I to Level IV, demonstrating increasing difficulty of items. Although these results support aspects of the model, they do not directly assess its fundamental construct, the four-level abstraction hierarchy. Both of these efforts at validation assume the validity of the perceptual-language distance model when, in fact, the validity of the four levels of abstraction as defined by Blank et al. (1978b) is yet to be established.

As stated previously, Messick (1995) argues that validity must be defined not only as the evidence for a particular interpretation of score meaning, but also as the actual and potential consequences of that interpretation. If clinical actions, such as profiling children's language skills necessary for school success or the development of intervention programs are going to be based upon the four-level perceptual-language distance model, convincing empirical evidence must be added to the test authors' argument supporting those actions. Thus, the PLAI provides an excellent example of how an evidence base for a clinical assessment and subsequent treatment programming should be established. In this spirit, we attempted to directly evaluate the validity of the PLAI's four-level perceptual-language distance model using several different analytic approaches.

Method

Participants

One hundred and fifty-two children between the ages of three and five participated in this study. There were at least 50 children in each age group (three-year-olds, four-year-olds, and five-year-olds). They attended either preschool or kindergarten in London, Ontario or Elliot Lake, Ontario, or were enrolled in a speech and hearing clinic in London, Ontario. All the children who participated in this study spoke English as their primary language at home. Further, the sample was comprised of both children without and with communication impairments[1] in order to reflect the prevalence of communication impairments within the general preschool population. Thus, 11% of this sample (17/152) consisted of children with impaired communication, coinciding with the prevalence rate found by Beitchman, Nair, Clegg, and Patel (1986). Specific characteristics of this sample are shown in Table 1.

Table 1. Summary of Participant Demographics.

Group	Total Sample n = 152	Three-year-olds n = 50	Four-year-olds n = 50	Five-year-olds n = 52
Boys	85 (55.9%)	30 (60.0%)	22 (44.0%)	33 (63.5%)
Girls	67 (44.1%)	20 (40.0%)	28 (56.0%)	19 (36.5%)
SES 1	36 (23.7%)	5 (10.0%)	15 (30.0%)	16 (30.8%)
SES 2	51 (33.6%)	20 (40.0%)	13 (26.0%)	18 (34.6%)
SES 3	65 (42.8%)	25 (50.0%)	22 (44.0%)	18 (34.6%)
Normal Communication	135 (88.8%)	48 (96.0%)	43 (86.0%)	44 (84.6%)
Impaired Communication	17 (11.3%)	2 (4.0%)	7 (14.0%)	8 (15.6%)

Note. SES 1 indicates elementary or high school education of parents; SES 2 indicates one to three years of postsecondary education of parents, SES 3 indicates four or more years of postsecondary education of parents.

The socioeconomic status (SES) of each child was determined by the highest reported level of education achieved by the child's parents. In the event that parents differed with respect to their level of education, the highest level reported by the pair was taken.

Procedure

Data collection. Each child's hearing was screened via pure tones at 20 dB HL for frequencies of 500, 1000, 2000, and 4000 Hz.[2] Subsequent to the hearing screening, the PLAI was administered as outlined in the test manual (Blank et al., 1978a). The PLAI consists of 60 items, either questions or directives; children respond verbally to some items and nonverbally to others. Reinforcement for attending and short breaks were provided for the children throughout testing as necessary. Scoring was completed using the 4-point quantitative scale (0 - 3) provided in the test manual. A score of 3 represents a fully adequate and appropriate response, or one that fully meets the demands of the task; 2 represents an acceptable response that may have some extraneous or imprecise information; 1 is given to responses that are considered ambiguous or one for which it is impossible to determine the adequacy; and 0 is assigned to incorrect responses, "I don't know", and no response. The assignment of scores is based on a rating procedure for which criteria guidelines and examples are provided for each individual item. These guidelines outline the essential components of adequate responses and illustrate the characteristics that define inadequate responses. As well as scoring each individual item, the mean of all items within a level was obtained, as directed in the test manual. As described by the test's authors, this is done because it is the child's ability to meet the core demands of the test level that is of primary interest and, hence, performance on the entire group of items within a level is more important than on any single item. The level mean scores are used in creating a profile of discourse skill along the perceptual-language continuum.

Agreement. PLAI test protocols were scored by five graduate students in speech-language pathology. To determine interrater agreement of the test scores, 10% of the test protocols were re-scored by an additional speech-language pathology graduate student who was not involved in the initial protocol scoring. Point by point agreement was 93%.

Results

Analysis of Age and Level Differences

To begin evaluation of the PLAI's construct validity, the first analyses conducted assumed that the four-level model is indeed appropriate for describing the demands of instructional discourse as presented by the authors. A core assumption of the model is that the ability to cope with increasingly complex cognitive demands of language should be sensitive to age differences during the preschool years. Thus, the model would predict that scores for each level should increase with increasing age, reflecting greater ability in older children, and that within each age group, scores should decrease as the test levels increase, reflecting increasing difficulty with each successive level.

Mean scores for each age group at each test level were calculated and are shown in Table 2. To test for age and test level differences, a 3 (age) x 4 (test level) analysis of variance (ANOVA) was conducted, with test level as a repeated factor and age as a between-groups factor.

Table 2. Mean Scores and Standard Deviations at Each of the Four Test Levels for Three Age Groups.

Age		Level 1	Level 2	Level 3	Level 4
Three yrs	M	2.06	1.51	0.93	0.71
	SD	0.46	0 .49	0.49	0.46
Four yrs	M	2.44	2.11	1.71	1.48
	SD	0.29	0.42	0.52	0.51
Five yrs	M	2.60	2.36	2.16	2.05
	SD	0.38	0.47	0.68	0.72

Results of the ANOVA revealed a significant effect for ages (F (2, 149) = 75.49, $p < .01$), a significant effect across levels (F (3,447) = 237.85, $p < .01$), and a significant age by level interaction (F (6,447) = 15.04, $p < .01$). These results indicate that, although scores decreased across the levels and increased across age groups as predicted, the rate of score change across the age groups was not consistent across all four levels. Tukey post hoc comparisons revealed that the three age groups differ significantly from one another for Levels II, III and IV, showing the expected pattern of change. However, at Level I, 4- and 5-year-olds did not differ significantly, indicating that their scores on this lowest level are indistinguishable from one another. This may simply reflect the ease of Level I items for children in these age groups.

To examine further the extent to which the four level scores were able to distinguish among the three age groups, discriminant function analysis (DFA) was performed. The purpose of the DFA was to use the four level scores to classify children into their age group, and to determine which levels' scores were essential to the distinction among age groups. The DFA revealed that rather than any level score being particularly essential, a single composite variable combining the four test levels accounted for 95% of the variance among age groups and maximally differentiated the three age groups of children. Thus, when each age group's mean score for each test level were analyzed, the most powerful discriminator of age was not the four discrete test levels but a single weighted composite variable combining all four levels in a manner to maximally differentiate age groups. Levels III and IV were most heavily weighted in the composite variable, indicating that discrimination among age groups is more highly influenced by Level III and IV scores than by Level I and II scores. These results indicate that none of the four, level scores on their own are able to differentiate accurately among age groups and further suggest that a model comprised of four distinguishable levels might not be sustainable on the basis of the performance data. To evaluate the viability of the four-level model, a series of item analyses were performed. These analyses were performed using each participant's score on each of the 60 items included in the test rather than the participants' mean scores on the four levels.

Internal Consistency of Level Items

The first item analysis conducted was an evaluation of the internal consistency of items comprising each test level. Internal consistency is a function of the intercorrelations among individual items and reflects the extent to which the items reliably measure a common construct (Anastasia, 1986). In the case of the PLAI, such a construct would be the purported "essential feature" of each level. The internal consistency was calculated separately for the items within each level for each of the three age groups using Cronbach's Alpha (Cronbach, 1951). Alpha coefficients within each test level for each group are shown in Table 3.

Table 3. Internal Consistency of Each Test Level by Age Group.

	Three-year-olds	Four-year-olds	Five-year-olds	All age groups
Level 1	.66	.47	.77	.73
Level 2	.61	.62	.71	.77
Level 3	.61	.63	.84	.85
Level 4	.68	.64	.85	.87

As can be seen in this table, the items within the four levels demonstrated moderate internal consistency for all of the age groups. The highest alpha coefficients were obtained for 5-year-olds, demonstrating that in general, level scores are most reliable for 5-year-olds and less so for 3- and 4-year-olds. However, all alpha coefficients are less than .90, the recommended standard for test reliability (APA, 1985), although Levels III and IV approach this value for 5-year-olds.

Factor Analyses

Factor analysis is an analytic approach that allows the examination of the relational structure that underlies the intercorrelations among test items. Intercorrelations among items indicate that the correlated items share a common underlying construct. Items that are correlated can be assumed to measure similar constructs. Factor analysis is used to determine the common aspects shared by a set of items, thereby allowing identification of the construct(s) underlying them. Two approaches to factor analysis exist. The first is the confirmatory approach that aims to confirm an hypothesized model imposed on the data. The second is the exploratory approach, that aims to reveal the model that best accounts for the observed pattern of item correlations. We again began by assuming the accuracy of the authors' four-level model and examined how well the children's data actually fit those distinct levels.

Confirmatory factor analysis. Confirmatory factor analysis (CFA) permits a researcher to test how well a proposed model actually corresponds to the relationships and patterns observed among data. This is done by forcing the items to conform to the proposed model and evaluating the extent to which the imposed model is able to be account statistically for the observed relationships among the items. Confirmatory factor analysis of the PLAI was conducted using LISREL 8.03. Analysis was performed on a four-factor model with the factors corresponding to the four levels of the PLAI and items modeled onto their respective factors. Resulting factor loadings can be found in Table 4.

The confirmatory model allowed for correlations among the factors. The model was not found to provide an acceptable fit to the data ($\chi^2(1704) = 2557.62, p < .001$; $GFI = .65$; $CFI = .70$; $NNFI = .69$). Loadings of the items onto their respective factors indicated that many items, particularly those in Levels I (e.g., #5, 25, 26, 51) and II (e.g., #18, 52) did not correlate strongly with the factors designated by the perceptual-language distance model. The loadings suggested that the model provided a better fit for Levels III and IV than for Levels I and II. Further, the correlations among the factors were sufficiently high as to suggest that distinction among the four factors may not be warranted as can be seen in Table 5.

The high correlations among the factors, along with the

Table 4. Factor Loadings for the Four Levels of the PLAI - Confirmatory Analysis.

Item	Level 1	Level 2	Level 3	Level 4
1	.416			
2	.385			
5	.139			
19	.701			
20	.658			
21	.598			
25	.066			
26	.013			
29	.385			
30	.566			
37	.352			
51	.195			
55	.287			
59	.330			
60	.586			
6		.397		
7		.516		
8		.438		
9		.449		
10		.415		
17		.657		
18		.139		
31		.512		
38		.345		
39		.461		
40		.488		
52		.240		
53		.320		
54		.511		
57		.542		
3			.452	
4			.491	
13			.590	
14			.547	
15			.655	
22			.569	
32			.297	
33			.513	
34			.641	
35			.626	
42			.477	
43			.315	
44			.541	
45			.554	
48			.588	
11				.414
12				.623
16				.499
23				.531
24				.744
27				.648
28				.519
36				.670
41				.649
46				.611
47				.496
49				.574
50				.537
56				.493
58				.390

Table 5. Intercorrelation of Test Levels.			
	Level 1	Level 2	Level 3
Level 2	.88		
Level 3	.84	.98	
Level 4	.77	.92	.94

poor fit of the model to the data, may indicate the presence of a higher-order factor, that is, a common cause contributing to the apparent overlap among the four factors. Alternatively, it may also suggest the need to explore a different model.

To test the hypothesis that the factor structure of the PLAI is defined by four first-order factors and one second-order factor, a hierarchical confirmatory factor analysis was conducted. All four first-order factors were defined as in the initial confirmatory factor analysis. The four first-order factors were all modeled onto a single second-order factor. The inclusion of the second-order factor did not improve the fit of the model (χ^2 (1706) = 2560.63, $p < .001$; $GFI = .65$; $CFI = .70$; $NNFI = .69$). Confirmatory factor analysis suggested that the data collected on the 60 PLAI items do not adequately fit the four-level model proposed by Blank et al. (1978b).

Exploratory factor analysis. Given that the CFA demonstrated that the four-level model could not account for the performance data of our participants, we explored whether some alternative model could better explain the data using the second approach to factor analysis. Exploratory factor analysis (EFA), which creates a model based upon the data in contrast to CFA which attempts to fit data to an existing model, was conducted next. The extraction of a five-factor model was supported by both a scree plot and the Kaiser Rule (i.e., eigenvalues of greater than 1.0). These are decision rules that guide the determination of the mathematically based factors that are present in the intercorrelations among all test items. Varimax rotation was used to rotate the extracted factors and clarify the pattern of items that load onto each factor.

A five-factor solution, accounting for 40% of the variance among the items was extracted. It should be noted that 60% of the variance among the items was left unaccounted for by the five factors extracted in the analysis. This indicates that the items are measuring substantial variance that cannot be accounted for by common constructs, and that is likely due to specific variance associated with the individual test items.

Exploratory factor analysis investigates the relationships among individual variables and uncovers the dimensions that underlie them (Gardner, 1997). Thus, evidence that would support the continuum of perceptual-language abstraction would be a dimension with which concrete test items (i.e., those with little abstract content) would have a weak relationship, and items that are increasingly abstract would relate with increasing strength. Examination of the first unrotated factor revealed that all but one test item (#26: "Show me your shoes.") loaded positively on this factor. The remaining loadings ranged between 0.02 and 0.71, ($M = 0.47$, $SD = 0.16$.) with four items loading at less than .3. The predominance of moderate, positive loadings on the first unrotated factor provides evidence for a general factor. It accounted for 24% of the variance, which is a large portion of the total variance explained by the five-factor solution (40%). However, the expected pattern for the continuum that was described above did not emerge. Level I sentence repetition items were weighted as high in magnitude as many of the Level IV "what if" items. The similarity in the magnitude of weightings suggests that those Level I items share a similar feature with the Level IV items as identified by the general factor. The presence of a general factor indicates that there is a single construct underlying almost all the items in the PLAI.

Table 6 displays the rotated factor matrix. Items loading meaningfully on Factor 1 had loadings, or correlation with the factors, ranging between .27 and .70. This factor loaded 13 items from Level IV, 7 items from Level III, and 5 items from Level II. These items were examined to determine what they share in common and in doing so to identify the construct underlying them. The items loading onto this factor require analogic or analytic reasoning. For example, items that loaded onto this factor require that the child solve problems, explain why, or compare similar and different features of objects. Further, the items that loaded onto this factor are those for which pictures support the problem to be solved. The hypothetical nature of the problems, and the mental manipulations required to solve these items suggest that Factor 1 could be characterized as a higher order mental operation - visually supported.

The second and third factors extracted in the analysis utilize language as both the vehicle for presenting the stimulus and as the stimulus itself. Using the terminology of the model, language was both the perception and the vehicle used to discuss it. The second rotated factor was comprised of items with loadings between 0.28 and 0.71. Level III items split nearly equally across this factor and Factor 1. The Level III items that load onto this factor involve analytic and analogic reasoning, but differ from those loading on the first factor in that linguistic, rather than pictured stimuli elements comprised the perceived material. Examples of the kinds of

Table 6. Rotated Factor Loadings - Exploratory Factor Analysis. (Part 1)

Item	Assigned Level	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
12	IV	.70				
24	IV	.66				
13	III	.65				
46	IV	.62				
17	II	.56				
27	IV	.55				
36	IV	.55				
41	IV	.54				
45	III	.53				
28	IV	.52				
33	III	.52				
35	III	.51				
49	IV	.51				
56	IV	.50				
47	IV	.49				
11	IV	.49				
57	II	.47				
58	IV	.47				
6	II	.47				
44	III	.42				
48	III	.44				
22	III	.41				
23	IV	.39				
18	II	.33				
38	II	.27				
14	III		.71			
15	III		.60			
34	III		.52			
32	III		.50			
30	I		.48			
8	II		.48			
2	I		.46			
7	II		.44			
31	II		.42			

Table 6. Rotated Factor Loadings - Exploratory Factor Analysis. (Part 2)

Item	Assigned Level	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
16	IV		.38			
1	I		.36			
42	III		.35			
3	III		.35			
4	III		.31			
5	I		.29			
43	III		.28			
21	I			.82		
20	I			.81		
19	I			.68		
9	II			.57		
10	II			.45		
50	IV			.42		
59	I				.62	
53	II				.58	
29	I				.56	
40	II				.47	
54	II				.44	
37	I				.38	
60	I				.37	
39	II				.36	
55	I				.30	
25	I					.70
26	I					.67
51	I					.54
52	II					.30

items that loaded onto this factor include hypothetical conversations (e.g., What do you think the man said…) and definitions. Due to the central nature of language in the problems to be solved, this factor could be characterized as a higher order mental operation-metalinguistic.

The third rotated factor loaded Level I and Level II items moderately high to very high (0.45 to 0.81). One Level IV item loaded onto this factor (i.e., item # 50). However, simple declarative sentence repetitions, and sentence completion items characterize this factor, with the former having the highest loading. These items require sentence structure knowledge and memory, but not sophisticated mental operations; thus, this factor could be characterized as one of basic linguistic operations. The fourth rotated factor had moderate to high loadings of between .30 and .62. The items loading on this factor consisted of visual matching, with either visual or verbal recall and may comprise a memory factor.

The fifth rotated factor is the last linguistically based factor. Three of the four items have moderately high to high loadings ranging from 0.54 to 0.70 that come exclusively from Level I items. These items require the child to engage in single word naming or identification. It is also important to note that these items did not load on the general factor. That is, these items are not measuring the same construct measured by the rest of the items on the test. This factor may be characterized as fundamental verbal identification.

Finally, in order to determine whether the 5 factor alternative solution generated from our data was also sensitive to expected age related performance differences and factor difficulty, we once again tested for age and factor differences. A 3 (age) x 5 (factors) ANOVA with the factors generated by our data as the repeated measure and age as a between-groups factor was conducted. Mean scores for each age group for each factor are shown Table 7.

Table 7. Mean Scores For Each of the Five Factors at Three Age Groups.

Age		Level 1	Level 2	Level 3	Level 4
3 yrs	M	2.06	1.51	0.93	0.71
	SD	0.46	0 .49	0.49	0.46
4 yrs	M	2.44	2.11	1.71	1.48
	SD	0.29	0.42	0.52	0.51
5 yrs	M	2.60	2.36	2.16	2.05
	SD	0.38	0.47	0.68	0.72

Results revealed a significant effect for age (F (2, 149) = 57.655, p< .001), a significant effect across factors (F (4,596) = 185.09, p<.001), and a significant age by factor interaction (F (8,596) = 11.949, p<.000) just as in the initial analysis of the four-level model. Tukey's post hoc analyses revealed age differences for every factor except Factor 5. Additionally, within every age level, factors decreased in difficulty from Factor 1 to 5. Reorganizing test items into five factors based upon children's actual responses produced an alternative hierarchy of discourse formulations that was also sensitive to age differences and difficulty. That is, the same items grouped differently yielded an alternative hierarchy of difficulty that was developmentally sensitive.

Discussion

Through this study, we attempted to demonstrate the importance of obtaining evidence of the adequacy and appropriateness of assessment tools and the treatment planning based upon them. The example of the PLAI, with its model of the language of instruction, was used to illustrate this point. The PLAI is an assessment tool that predominantly serves as the basis for sequencing treatment programs. The model upon which the PLAI is based and which underlies both assessment and treatment applications is clearly articulated and hence, testable (Carotta, Carney, & Dettman, 1990; Marvin & Wright, 1997; Moeller, Osberger, & Eccarius, 1986; Nelson, 1993; Owens, 1995; Paul, 1995). Blank, et al.'s (1978b) model of the language of instruction proposes four distinct levels of abstraction. The clinical actions most frequently associated with the PLAI depend upon whether the essential characteristic associated with each level and its place in the hierarchy are accurate. In this study, we sought empirical evidence that would support those actions.

We approached evaluation of these clinical actions by analyzing the model. Comparisons of mean level scores by children's age, and item level analyses were conducted. We began by assuming the accuracy (i.e., the validity) of the four-level model as presented by the authors. We found that performance improved across the three age groups and that difficulty increased across the levels within any of the age groups. This supported the authors' preliminary investigations of construct validity (Blank et al., 1978a.) The confirmation of a progression of difficulty within and across age groups of children could be construed as supporting the clinical action of sequencing treatment according to the authors' four-level hierarchical model. To the contrary, although a progression of difficulty was confirmed for the four-level model, this does not mean that the items assigned to a given level actually reflected a similar essential characteristic that was distinct from the items comprising one of the other four levels. Additionally, it does not mean that another arrangement of the items would not also show a similar progression. The present study also showed that the age related performance differences were not distinguished by the four distinct levels of the model, but rather by a single composite variable combining all four. This was the first indication that the discourse characteristics embodied in the four-level hierarchy might not be empirically sustainable, although the PLAI and its model did capture preschool children's developing ability to cope with the abstraction in the language of instruction. Given that the four levels form the sequence of scaffolding suggested for treatment programming (i.e., dictate what type of instructional question is easier or more difficulty) further examination of model was pursued.

In the second set of analyses, the groupings of item scores as delineated by the model (i.e., the four-level assignment) and an alternative grouping were examined via factor analytic procedures. The first analysis, CFA, demonstrated that the items did not fit adequately the groupings suggested by the four-level model. That is, the results suggested that the type of items identified as reflecting a particular level did not necessarily relate to one another uniquely as would be expected if the levels were distinct and had an essential defining characteristic. Rather, results indicated that *all* four levels and hence, *all* the items that comprised them, were highly related as shown on Table 5 (i.e., r = .77 to .98). The significant amount of overlap indicated by the high intercorrelations is evidence that counters the assumption of four distinct levels of abstraction. Blank et al. (1978b) suggest that there might be some overlap between boundaries of the four levels; however, these data suggest near total overlap rather than the gradual blending of one level into the next. Given that the four distinct levels proposed in the model did not account for how items on the test were related, we then explored whether an alternative organization of items would better describe the data.

The second factor analysis, EFA, revealed a general factor underlying almost all test items, supporting the notion of a composite variable found in the DFA and the high intercorrelation among levels that was previously described. Thus, evidence from all approaches to analysis of the data converged on the finding that a single central dimension, as opposed to four distinct levels of abstraction, was being evaluated by the PLAI. Items involving the highest level language (i.e., those coming primarily from Levels III and IV) were most strongly related to the composite variable revealed by the DFA and also loaded the highest on the general factor revealed in the EFA. Given that these items typically involved analogic or analytic reasoning, it may be suggested that the underlying factor measured by the PLAI is abstraction or intellect (Sternberg & Detterman, 1986) as proposed by Blank, et al. (1978b).

The EFA also revealed that a structure with five factors better described the interrelationships among items. As shown in Table 6, items from the original four levels were interspersed through the five factors. Clustering of the test items into the alternative five factor structure seemed to depend not only upon the complexity of the mental operation, but also on mode of presentation of the stimulus materials, and importantly, on the role of language as an actual property of the materials being discussed. The first two of the five factors included items requiring a higher order mental operation such as problem solving. However, these factors differed in how distinct the medium of the materials was from the vehicle of discussion (i.e., language). The materials that supported Factor 1 items were visually based (i.e., presented in picture format) and higher level language was used to describe or set up the problem the child was required to solve. In Factor 2, linguistic stimuli were the actual material presented (e.g., a short narrative, or a word requiring a definition), and typically no pictorial support was provided. These required metalinguistic or metapragmatic ability. The items loading on our Factors 3 and 4 had notable memory components, but did not require abstract manipulation. They were distinguishable on a metalinguistic basis as well. Factor 3 was comprised of items that required the child to engage in verbal manipulation of verbal stimuli (e.g., providing completion to or repetition of a simple declarative sentence.) Factor 4, to the contrary, was characterized by items requiring children to match like pictures, then recall them once they were removed. Once again the material was separated from the medium of discussion, as in the first factor. Factor 5 included items that involved simple verbal identification from Levels I and II, and these items loaded weakly on the general factor, suggesting they were not measuring a level of abstraction.

In a factor analysis, items are clustered together because they are strongly related. Thus, it could be argued that the items comprising each of our alternative five factors do have an empirically supported essential characteristic; whereas, the level assignments for items that were proposed in the authors' model do not have empirically supported essential characteristics. Evidence for an essential characteristic distinguishing each grouping of items from one another is necessary to support treatment sequencing along the continuum of abstraction. However, we would also need to demonstrate that our five-factor structure would be developmentally sensitive and show a progression of difficulty for it to be valid. Our final analysis comparing age and factor groupings did exactly this. The only factor that did not distinguish the four age groups was Factor 5. This factor was comprised of very basic verbal identification task such as "Show me your nose." Most 18 month olds typically master such a verbal skill so it is not unexpected that 3-, 4- and 5–year-olds would all respond similarly. The factors progressed in difficulty from Factor 5, the easiest, to Factor 1, the most difficult.

To summarize, all the analyses conducted in the present study supported the assumption that the PLAI's model of the language of instruction evaluated a preschool child's ability to cope with abstract language of varying degrees. We did not find support for the four distinct levels proposed by Blank, et al. (1978b) and thus, could not find evidence for the clinical practices that rest upon the use of those four levels. If the construct validity of the model is threatened, then the validity of actions based upon the four-level model are similarly threatened. The clinical practices that utilize the four-level model to profile a child's strengths and weaknesses in coping with the language of instruction and then to develop treatment procedures for teachers so that they can prompt the child to more abstract levels of responding (Paul, 1995; Nelson, 1993) must be questioned. The current study suggests that there is no evidence for the clustering of discourse formulations into the four levels suggested by the model. Thus, sequencing instructional question according to that model cannot be supported (Blank et al., 1978b; Blank & White, 1986; Paul, 1995.)

In evidence-based practice, if the validity of our assumptions is challenged by empirical data, then the clinical actions derived from those assumptions must be reexamined. When we could not verify the four distinct levels, we pursued alternative grouping of items and sought empirical verification. Our five-factor structure is based upon items that are all strongly related within any one factor, all reflect some level of abstraction, and are sensitive to expected age and difficulty progression. Thus, we have provided evidence for alternative profiling and treatment sequencing patterns that could be used in collaborative consultation with teachers.

Validation plays a fundamental role in establishing the evidentiary base of clinical practice. The evidence obtained from empirical studies such as the present one is intended to influence the rationale decision-making driving clinical action. The present study provides an example of how clinically relevant evidence for an assessment practice and the treatment programming resulting from it may be obtained and utilized. Evidence based practice, like validation, is an ongoing process; one we have just begun in the area of language assessment and intervention.

Author Notes

Funding for this study was provided by the London Chapter of the Barbershoppers' Harmonize for Speech Fund. The authors wish to thank Wm. Yovetich, PhD for his assistance. In addition, the authors wish to acknowledge the efforts of Linda Levy Fisk, Kristen Strauss, Andrea Storie and Donna Torrie in data collection. Finally, we thank the children and staff of the North Shore Board of Education, Montessori House of Children, Acorn, Oak Park, North Park and Jigsaw Preschools, and the University of Western Ontario Speech Clinic who generously gave their time and cooperation.

Meghan Reichheld is currently affiliated with the Niagara Peninsula Children's Centre, St. Catharines, Ontario. Please address all correspondence to Elizabeth Skarakis-Doyle, School of Communication Sciences and Disorders, University of Western Ontario, London, Ontario, Canada N6G 1H1. E-mail: eskaraki@uwo.ca.

Footnotes

[1] Participants with impaired communication included five children with identified or suspected articulation or phonological impairments, 10 children with language impairments and two children with fluency disorders. The children with identified articulation, language, and fluency problems had been evaluated by a licensed speech-language pathologist and were enrolled in treatment programs in a university-based clinic. The three children suspected of having articulation/phonology problems were identified by their parents and were awaiting complete evaluations.

[2] Only nine of the 152 participants failed the hearing screening. Eight of these children went on to score at or above the mean performance for the children of their age group on the PLAI. Only the remaining child scored below the mean for his or her age on the PLAI.

Manuscript received May 31, 2000
Accepted December 9, 2000

References

American Psychological Association. (1985). *Standards for Educational & Psychological Testing*. Washington, D.C.: American Psychological Association.

Anastasi, A. (1986). Evolving concepts of test validation. *Annual Review of Psychology,37,* 1-15.

Beitchman, J.H., Nair, R., Clegg, M., & Patel, P.G. (1986) Prevalence of speech and language disorders in 5-year-old kindergarten children in the Ottawa-Carleton Region. *Journal of Speech and Hearing Disorders*, *51*, 98-110.

Blank, M., & Franklin, E. (1980). Dialogue with preschoolers: A cognitively-based system of assessment. *Applied Psycholinguistics*, *1*(2), 127-150.

Blank, M., Rose, S., & Berlin, L. J. (1978a). *Preschool Language Assessment Instrument: The language of learning in practice*. San Antonio TX.: The Psychological Corporation, Harcourt Brace Jovanovich, Inc.

Blank, M., Rose, S., & Berlin, L. J. (1978b). *The language of learning: The preschool years*. New York: Grune & Stratton.

Blank, M., & White, S., (1986). Questions: A powerful but misused form of classroom exchange. *Topics in Language Disorders, 6,* 2-12.

Bradshaw, M., Hoffman, P., & Norris, J. (1998). Efficacy of expansions and cloze procedures in the development of interpretations by preschool children exhibiting delayed language development. *Language, Speech and Hearing Services in the Schools, 29*, 85-95.

Carotta, C., Carney, A. E., & Dettman, D. (1990). Assessment and analysis of speech production in hearing-impaired children. *ASHA*, 32, p.59 (Abstract). Cited in *Introduction to Audiologic Rehabilitation* (3rd ed.), (p. 349) by R. Schow & M.A. Nerbonne (1996). Boston, MA: Allyn and Bacon.

Cole, K., Dale, P., & Mills, P. (1990). Defining language delay in young children by cognitive referencing: Are we saying more than we know? *Applied Psycholinguistics, 11,* 291-302.

Conti-Ramsden, G., & Friel-Patti, S. (1986). Mother-child dialogues: Considerations of cognitive complexity for young language learning children. *British Journal of Disorders of Communication, 24,* 245-255.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

Gardner, R., (1997). *Psychological statistics using SPSS for windows*. Unpublished manuscript, University of Western Ontario, London, Ontario.

Hutchinson, T. (1996). What to look for in the technical manual: Twenty questions for users. *Language Speech and Hearing Services in the Schools, 27*, 109-121.

Klein, H., & Moses, N. (1994). *Intervention planning for children with communication disorders: A guide for clinical practicum and professional practice.* Englewood Cliffs, N.J.: Prentice Hall.

Lehrer, R., & deBernard, A. (1987). Language of learning and language of computing: The perceptual-language distance model. *Journal of Educational Psychology, 79*(1), 41-48.

Marvin, C. A., & Wright, D. (1997). Literacy socialization in the homes of preschool children. *Language Speech and Hearing Services in the Schools, 28*, 154-163.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741-749.

Moeller, M. P., Osberger, M. B., & Eccarius, M. (1986). Cognitively based strategies for use with hearing-impaired students with comprehension deficits. *Topics in Language Disorders, 6,* 37-50.

Moffett, J. (1968). *Teaching the universe of discourse.* Boston: Houghton Mifflin.

Nelson, N.W. (1993). *Childhood language disorders in context: Infancy through adolescence.* New York: Macmillan.

Owens, R. (1995). *Language disorders: A functional approach to assessment and intervention* (2nd ed.). Boston, MA: Allyn Bacon.

Palinscar, A. S., Collins, K., Marano, N., & Magnusson, S. (2000). Investigating the engagement and learning of students with learning disabilities in guided inquiry science teaching. *Language Speech and Hearing Services in the Schools, 31*, 240-251.

Paul, R. (1995). *Language disorders from infancy through adolescence: Assessment and intervention.* St.Louis: Mosby.

Sorsby, A., & Martlew, M. (1991). Representational demands in mother's talk to preschool children in two contexts picture book reading a modeling task. *Journal of Child Language, 18,* 373-395.

Shimotakahara, J.,& Li, J. (1998). *The kindergarten language centre program: Year two report 1995-96.* (Report # 95-96-12). Scarborough, ON: Scarborough Board of Education.

Skarakis-Doyle, E., Yovetich, W., Strauss, K., Storie, A., Levy Fisk, L., & Torrie, D., (1998). A Canadian normative sample for the *Preschool Language Assessment Instrument. Journal of Speech Language Pathology and Audiology, 22*, 126-132.

Sternberg, R. J., & Detterman, D. K. (Eds.). (1986). *What is intelligence? Contemporary viewpoints on its nature and definition.* Norwood, NJ: Ablex.

vanKleek, A., Gillam, R. B., Hamilton, L., & McGrath, C. (1997). The relationship between middle class parent's book sharing discussion and their preschoolers abstract language development. *Journal of Speech Language and Hearing Research, 40,* 1261-1271.