Use of the ROWPVT with Small-town Canadian Kindergarten Children

Application du ROWPVT aux enfants de la maternelle d'une petite ville canadienne

Peter Flipsen Jr.
School District #81
Fort Nelson, BC

Key words: ROWPVT, standardized testing, kindergarten, receptive language, Canadian children, vocabulary

Abstract

Some concern has been raised by clinicians about assessing Canadian children with tests whose standards were developed in the United States. It has been generally assumed that Canadian and American children are similar enough for the norms to be valid and thus for the tests to be appropriate to use with both groups. This study was an attempt to test this assumption by determining if the ROWPVT is appropriate in this respect. Testing was carried out on 176 small-town Canadian kindergarten children. Results indicated that the obtained standard scores were not normally distributed as would be expected in a sample of this size. Far fewer subjects than expected achieved standard scores that were more than 1 S.D. below the test mean. The test thus appeared to lack sensitivity, making its use as a screening tool questionable with the target population. Results also indicated that the recommended starting point for testing had been set too high. Basal scores were established at much lower values than might be expected. Consequently, if the ROWPVT is used, testing should begin at a lower level than recommended.

Résumé

Les clinicien(ne)s s'interrogent sur l'évaluation des enfants canadiens au moyen de tests dont les normes ont été élaborées aux États-Unis. On suppose généralement que les enfants canadiens et américains sont assez semblables pour que les mêmes normes s'appliquent dans les deux pays, donc que les épreuves conviennent aux deux groupes. La présente étude devait vérifier cette hypothèse en déterminant si le ROWPVT donne de bons résultats. Cent soixante-seize enfants de maternelle d'une petite ville canadienne ont subi le test. Les résultats révèlent que la cote normale obtenue à l'issue du test n'est pas distribuée comme on pourrait s'y attendre pour un échantillon de cette sorte. Beaucoup moins d'enfants ont obtenu la note normale qui se trouvait par ailleurs plus de 1 E.-T. sous la moyenne. L'épreuve semble donc manquer de sensibilité, ce qui amène les auteurs à s'interroger sur son utilité pour le dépistage dans la population visée. Les résultats montrent également que le point de départ recommandé est trop élevé. La note de base est beaucoup plus faible que prévu. Si l'on se sert du ROWPVT, les tests devraient donc débuter à un niveau plus bas que celui recommandé.

Introduction

Despite some limitations, standardized tests remain an integral part of speech-language pathology practice (McCauley & Swisher, 1984). A significant number of the standardized tests currently used in Canada were developed in the United States and thus have normative data based on American children. Some clinicians in Canada have commented that the experience base of Canadian children may not be the same as that of American children and therefore question the appropriateness of using these tests with Canadian children.

This study was an attempt to address that issue. It was intended to determine whether the *Receptive One-Word Picture Vocabulary Test (ROWPVT)* (Gardner, 1985) is an appropriate tool for measuring receptive vocabulary skills in small-town Canadian kindergarten children. This was attempted by assessing whether the standard scores achieved by such children are normally distributed.

The norms for the ROWPVT were established using 1128 children aged 2;0 - 11;11. At 5 and 6 years of age, the children in the current study were in the centre of the age range for the normative group. In addition, the norms for the ROWPVT were established using children in the San Francisco area of California. The current study was then a comparison between small-town Canadian children and those from a large American city. This latter contrast is an important one since a recent study found significant differences between the performances of rural and suburban children on the *Preschool Language Scale* (Hilton & Mumma, 1991). A

secondary objective of the current study was to test an observation based on the author's previous experience with the ROWPVT which suggested that the recommended starting point for test administration had been set too high.

Method

Subjects

The current study was conducted as part of a kindergarten speech and language screening project, in a small resource-based town (population: 5000) in northern British Columbia. All kindergarten students in the school district in the 1991-92 and 1992-93 school years were included in the project.

Attendance at kindergarten is mandatory in British Columbia (although entry at age 5 can be delayed until age 6). The screening then effectively evaluated the entire cohort of kindergarten students in the community.

As shown in Table 1, the study included 176 subjects with a mean age of 66.9 months (range 61-79 months). The children were largely of white European background, with approximately 5% being native North Americans.

Procedures

The ROWPVT was chosen for the screening as a measure of receptive language skills. A standard score of 80 was set as the lower limit for passing the screen. Any subject scoring below this value (i.e. below the 10th percentile) would be referred for additional assessment.

Expressive skills (speech, language, voice and fluency) were screened by eliciting a spontaneous language sample which was tape-recorded. Each subject was asked to outline a story from pictures in a wordless storybook. Evaluation of the language sample was made based on a combination of on-line notes and tape transcription.

In all cases, the ROWPVT was administered first, followed by language sampling. All speech and language screenings and data analyses were conducted by the author.

Following data collection, the main research question was evaluated by an examination of overall performance trends. The standard score distribution which was arrived at was compared with that expected on the basis of a normal distribution.

The assessment of the starting point for testing was made by examining the median item number at which basal score was established. A lower than expected basal score would suggest that the starting point for testing had been set too high. Such a situation would necessitate frequent reversals of direction during test administration.

Results

Standard Scores

Each child's raw score was converted to a standard score using the normative tables provided with the test. As shown in Table 1, the standard score distribution obtained (mean of 100.0 and S.D. of 13.1) was very similar to that reported in the normative information (mean of 100 and S.D. of 15). The scores ranged from 60-145. The mean score for females (99.0) was only marginally lower than for males (100.9) but appeared in a slightly narrower range (75-139 for females versus 60-145 for males).

Table 1. Subject Ages and ROWPVT Standard Score

	n	Age (in months)			Std. Score		
		Mean	S.D.	Range	Mean	S.D.	Range
Males	86	67.3	3.7	61-79	100.9	14.8	60-145
Females	90	66.6	3.5	61-75	99.0	11.2	75-139
Total	176	66.9	3.6	61-79	100.0	13.1	60-145

If one were to use the obtained mean and standard deviation data as the sole criterion, it would appear that the scores were distributed normally. It would then be logical to assume that the test is an appropriate one to use with the target population. However, only 4/176 scores (2.3%; 3 males) fell below the cutoff of 80, and thus represented screening failures. Since a standard score of 80 represents the 10th percentile, one would expect a normal distribution to yield 15-18 failures (i.e. 9-10% of 176).

To resolve this discrepancy, the distribution of the obtained standard scores was compared to the normal distribution using standard deviation categories (see Table 2).

Table 2. Obtained vs. Expected Distribution of Scores*

Ranges	Males	Females	Overall	EXPECTED**
Below 70	3	0	3	4
70-84	2	1	3	24
85-100	45	56	101	60
100-115	23	27	50	60
116-130	9	5	14	24
131+	4	1	5	4
TOTAL	86	90	176	176

* expressed as number of subjects
** from areas under the normal curve (Triola, 1980)

The expected values shown in Table 2 were derived using the areas under the normal curve as follows (Triola, 1980):

Mean to +1 S.D.	(100-115)	includes	34.13%
Mean to -1 S.D.	(85-100)	"	34.13%
+1 S.D. to +2 S.D.	(116-130)	"	13.59%
-1 S.D. to -2 S.D.	(70-84)	"	13.59%
Area beyond +2 S.D.	(131+)	"	2.28%
Area beyond -2 S.D.	(below 70)	"	2.28%
			100.00%

The two distributions were compared using a chi-square (X^2) one-sample test. The resulting X^2 value of 52.74 (5 df, n=176) was statistically significant ($p<0.001$). The test scores obtained in the current study were therefore not normally distributed.

As can be seen from Table 2, there appears to be a mismatch between the obtained and the expected distributions in at least 3 of the 6 ranges. It should be noted that there were far fewer scores (3) than expected (24) in the 70-84 range (where the bulk of the screening failures should have occurred). This contrasts with a far greater number of scores (101) than expected (60) in the 85-100 range.

Also of note is that 151/176 (86%) scores fell in the 85-115 range. Given this pattern, it would appear that the test lacks sensitivity when used with this particular population. When used as a screening tool (as many standardized tests are), it is the lack of sensitivity at the lower end that is most critical. Children with possible receptive vocabulary problems may not be identified by an insensitive test.

Basal Score

Table 3 lists the recommended starting item numbers given in the test manual along with the predicted basal items. These predictions are based on the assumption that the recommended starting point is a reasonable one. A "normal" subject would then be expected to establish a basal at least eight items above this point.

For the ROWPVT, a basal is established by identifying eight consecutive correct items. Failure on any item among the first eight requires the examiner to reverse direction in testing until the subject does identify eight consecutive items correctly. The examiner then returns to the forward direction to establish a ceiling.

Table 3. Predicted Basal Items

Age Ranges	Starting Items*	Predicted Basal Items**
3;0 - 3;5	11	19 +
3;6 - 3;11	20	28 +
4;0 - 4;11	30	38 +
5;0 - 5;11	40	48 +

* as per test manual

** highest of eight consecutive correct

The median item number at which basal was established was calculated for the 176 subjects and results are reported in Table 4.

Table 4. Obtained Basal Data

	Males	Females	Total
Median *	32	37	32.5
% below Item 48	92	93	93
% below Item 38	63	56	59
% below Item 28	22	23	23
% below Item 19	10	12	11

* item at which basal score established (highest of eight consecutive correct)

As can be seen by comparing Tables 3 & 4, the subjects achieved basal at a much lower level than predicted. In fact, fully 93% of the subjects achieved a basal below the predicted item 48 level. To establish a basal without changing the direction of testing for at least 75% of the subjects, an examiner would have needed to start testing at least two age ranges below that recommended by the test designer. The patterns obtained for males and females were very similar.

Discussion

With this sample size, one would expect the standard scores of normal subjects to be normally distributed. Given that the ROWPVT standard score distribution was significantly different from normal, it would appear that this particular test may not be appropriate to use to screen receptive vocabulary skill in small-town Canadian kindergarten children.

It is particularly disturbing that the number of failures was so much lower than would be expected. This suggests that the test is not sufficiently sensitive to identify those with poor receptive vocabulary skills (the specific objective of a screening tool).

From the evaluation of basal data it would also appear that if this test is to be used, administration should begin at a point lower than that recommended in the test manual by at least two age categories. This would reduce the frequency of having to change direction during testing (a practice which may be disconcerting or at least confusing to some young children).

It would of course be premature to extend the current results to other age groups or urban settings. Additional evaluation of this test should be carried out before any broad conclusions are made about the appropriateness of using this test with other groups of Canadian children.

Acknowledgements

The author very much appreciates the assistance of the kindergarten teachers during the screening project as well as the support and encouragement of Mr Claude Ruggieri and Mrs Anne Cooper during data collection and manuscript preparation. Finally, the extremely insightful comments of one anonymous reviewer are most gratefully acknowledged.

Address all correspondence to: Peter Flipsen Jr., School and Community Speech-Language Pathologist, School District #81 (Fort Nelson), P.O. Box 87, Fort Nelson, B.C. V0C 1R0.

References

Gardner, M.F. (1985). *Receptive One-Word Picture Vocabulary Test*. Novato, California: Academic Therapy Publications.

Hilton, L.W. & Mumma, K. (1991). Screening rural and suburban children with the preschool language scale. *Journal of Communication Disorders. 24*: 111-122.

McCauley, R.J. & Swisher, L. (1984). Use and misuse of norm-referenced tests in clinical assessment: a hypothetical case. *Journal of Speech and Hearing Disorders. 49*: 338-348.

Triola, M.F. (1980). *Elementary Statistics*. Menlo Park, CA: Benjamin/ Cummings Publishing Company.