Routledge
Taylor & Francis Group

Developmetrics

Evaluating listening comprehension in children using a dynamic assessment approach: An exploratory study

Sara Brandão, Irene Cadime, Iolanda Ribeiro, and Fernanda Leopoldina Viana

Universidade do Minho, Braga, Portugal

The aim of this study was to investigate the psychometric properties of the Dynamic Assessment Test of Listening Comprehension$_{4-6}$ (DATLC$_{4-6}$) for Portuguese preschoolers aged four to six years old. The test construction was based on the principles of dynamic assessment, specifically the gradual prompts model. The test includes 40 multiple-choice items, each of which includes three types of prompts. A total of 108 Portuguese preschool children participated in the study. High corrected item-total correlations were obtained. The principal component analysis results suggested that a single factor explains a high percentage of the results' variance. The value of Cronbach's alpha coefficient indicates a notably high internal consistency. The preliminary analysis of the DATLC4-6 suggests adequate psychometric properties. Limitations and directions for future studies are discussed.

Keywords: Listening comprehension; Dynamic assessment; Preschoolers; Reliability; Principal component analysis.

According to the Simple View of Reading, reading comprehension is the product of listening comprehension and decoding (Hoover & Gough, 1990). The influence of listening comprehension on reading comprehension is well documented in elementary school students (Nation, 2005). The predictive power of listening comprehension is also verified when listening comprehension is assessed in preschool students. The results obtained by Kendeou, Bohn-Gettler, White, and van den Broek (2008) indicate that listening comprehension assessed at age 4 had a significant contribution to children's text narrative comprehension at age 6. Similar results were obtained in other studies (e.g. Kendeou, van den

Broek, White, & Lynch, 2009). Research results also suggest that preschoolers are able to make inferences about events that they experience (e.g. television programmes) or stories presented orally but are less spontaneous, and explicit questioning or specific prompts are required to elicit an answer (van den Broek et al., 2005).

The principles of dynamic assessment (DA) can be particularly useful in the development of listening comprehension tests for children at these ages because they may provide the specific support and questioning that preschool children need for inference-making (Elleman, Compton, Fuchs, Fuchs, & Bouton, 2011). DA is an alternative to traditional assessment that is considered adequate to increase the predictive validity of cognitive, language and literacy tests in predicting school achievement and decrease measurement errors, given that it offers multiple opportunities for the children to succeed, unlike traditional assessment approaches, wherein responses are frequently scored right/wrong. Furthermore, DA is not so prone to floor effects (Caffrey, Fuchs, & Fuchs, 2008).

Data regarding the validity and reliability of listening comprehension measures are scarce because researchers typically use informal procedures—stories presented orally, in some cases using images or video, followed by open-ended questions or a retelling task (e.g., Kendeou et al., 2009; Paris & Paris, 2003)—without considering standardization or the measures' psychometric properties.

One exception is the work of Elleman et al. (2011), in which a DA test intended to evaluate inference-making skills in second graders was developed. The test included seven short passages that are read aloud to the children, followed by 21 open-ended questions. The children who were unable to answer a question were given a set of increasingly concrete prompts and clues to orient them to the relevant portion of the text until they could answer the question correctly. The internal consistency of the test as measured by Cronbach's alpha was .76. Statistically significant correlations were obtained with a standardized measure of reading comprehension. This study by Elleman et al. (2011) provided an innovative approach to listening comprehension assessment by implementing DA in a formal measure. However, standardized tests of listening comprehension for preschoolers within the DA framework remain unavailable.

The Dynamic Assessment Test of Listening Comprehension$_{4-6}$ (DATLC$_{4-6}$) for Portuguese preschoolers, ages 4–6 years old, was constructed to fill this gap. It is composed of a text with 1694 words divided into 20 passages, which are followed by 40 items with three types of prompts. We used a dynamic approach based on the gradual prompts model described by Campione and Brown (1987). This model assumes that a standardized support is provided through a predefined set of sequential prompts. The literature suggests that such standardization increases the predictive potential of assessment measures (Caffrey et al., 2008).

The aim of this study was to investigate the psychometric properties of the DATLC$_{4-6}$, specifically, by analysing the items difficulty and discrimination, as well as the test dimensionality and reliability.

METHOD

Participants

The sample consisted of 108 preschool children from four Portuguese public kindergartens and two private kindergartens located in northern Portugal that have a collaboration protocol with University of Minho. Of these children in the sample, 57.4% were from private schools and 42.6% from public schools. Available national data from 2012 indicated that approximately 53.2% of the Portuguese preschoolers attended public schools and 46.8% private schools.

The mean age of the participants was 5.33 (SD = 0.53), ranging between 4 and 6 years old. Children with global developmental delays and/or auditory deficits were not included.

Procedures

The test was administered individually by trained psychologists in the children's kindergarten in a single session of approximately 45 minutes. All of the participants responded to all of the items that comprise $DATLC_{4-6}$. Legal authorizations were obtained from school boards and parents.

Measure

The $DATLC_{4-6}$ assesses listening comprehension of text in preschoolers. Children listen to a narrative text written by a Portuguese children's literature author, which is segmented into excerpts and followed by open questions (see the supplementary file for the text and items description). The principles of DA were integrated into the test construction by including a set of three predefined sequential prompts for each item, presented in a fixed order: (i) repetition of the full text excerpt, (ii) repetition of the portion of the text excerpt that contains the information required to produce a correct response and (iii) presentation of the item using a multiple-choice format. Each time that the children provides a non-response or an incorrect response, the next prompt for that item is presented, until no prompts for that item remain. A score of 4 points is attributed to correct responses without the use of any prompt, 3 points to a correct response after the first prompt, 2 points to a correct response after the second prompt and 1 point to a correct response after the third prompt. A score of 0 is assigned to an incorrect alternative selection after the third prompt. The items assess literal (8 items) and inferential comprehension (causal-26 items; informative-6 items). Most of the items assess causal inferences—related to elements of story grammar and organization of causal relations among story events—because this type of inference seems to be the most critical for story comprehension (Makdissi & Boisclair, 2006; Van Kleeck, 2008).

The test is presented in a digital format as a game. When the children provide an incorrect response to the item, a game character appears and offers some help. Prompts are then introduced. After the last prompt, no feedback is given to the children. The test is administered individually without a time limit and is controlled by the examiner.

Data analysis

Results were analysed using IBM SPSS Statistics 20. The items' difficulty index (Anastasi & Urbina, 1997) was calculated from the proportion of children who provided a correct response to each item, regardless of the number of prompts used. The index of discrimination of the items was estimated by computing the corrected item-total correlations, in which the item score is excluded from the total test score computation.

To explore the test dimensionality, a principal components analysis (PCA) with varimax rotation was employed. Due to the absence of a previous theoretical and empirical hypothesis about the test structure, an exploratory approach was used instead of a confirmatory approach (Thompson, 2004). To determine the number of factors to retain for subsequent analyses, we used Kaiser's rule (eigenvalues higher than 1) and the scree plot analysis (Field, 2009). Internal consistency was evaluated by computing Cronbach's alpha. The Mann–Whitney test was used to analyse sex-related differences and the Kruskal–Wallis test to analyse age-related differences.

Given that the test administration was controlled by the examiner and that all participants responded to all items, no missing data were obtained.

RESULTS

Table 1 presents the item analysis results. The item difficulty ranged from .51 to .95, with a mean value of .67. The items' mean score ranged from 1.44 (item 17) to 3.37 (item 13). Corrected item-total correlations ranged from .23 to .75 with a mean value of .59, exceeding the minimum reference value of .30 (Pallant, 2005).

Regarding the PCA results, eight components with an eigenvalue higher than 1 were found. However, the scree plot analysis suggested a one-dimensional structure (see Figure 1).

A second analysis was forced to retrieve one factor, which accounted for 38.78% of the total variance. The factor loadings were low to high, ranging between .24 and .77 (see Table 1). The value of the Cronbach's alpha coefficient for the 40 items was .96, suggesting a high internal consistency.

Scale scores were computed by taking the total sum of the items. No gender- (male: $M = 107.96$, SD $= 42.53$; female: $M = 108.36$, SD $= 45.70$; $U = 1421$, $p = .82$) or age-related (4 years old: $M = 50.00$, SD $= 40.04$; 5 years old:

TABLE 1
Items difficulty, corrected item-total correlations, descriptive statistics for the items scores and factor loadings ($n = 108$)

| | | | | | Score | | | | | | | | | | |
| | | | | | 4 | | 3 | | 2 | | 1 | | 0 | | |
Item	D_i	r	M	SD	n	%	n	%	n	%	n	%	n	%	Loading
1 (CI)	.84	.45	2.69	1.57	55	50.9	14	13	7	6.5	15	13.9	17	15.7	.47
2 (II)	.79	.53	2.29	1.63	42	38.9	13	12	10	9.3	20	18.5	23	21.3	.56
3 (LC)	.95	.48	3.25	1.23	71	65.7	15	13.9	5	4.6	12	11.1	5	4.6	.51
4 (II)	.74	.62	1.98	1.71	41	38	4	3.7	3	2.8	32	29.6	28	25.9	.65
5 (LC)	.85	.67	2.64	1.50	48	44.4	17	15.7	15	13.9	12	11.1	16	14.8	.70
6 (CI)	.90	.54	2.55	1.51	50	46.3	7	4	14	6.5	26	24.1	11	10.2	.56
7 (CI)	.82	.49	2.80	1.66	66	61.1	7	6.5	1	.9	15	13.9	19	17.6	.53
8 (II)	.87	.59	2.83	1.56	63	58.3	9	8.3	5	4.6	17	15.7	14	13	.62
9 (II)	.77	.43	2.26	1.62	40	37	13	12	15	13.9	15	13.9	25	23.1	.46
10 (CI)	.87	.62	2.72	1.57	60	55.6	5	4.6	10	9.3	19	17.6	14	13	.65
11 (CI)	.91	.51	3.24	1.41	81	75	3	2.8	3	2.8	11	10.2	10	9.3	.54
12 (LC)	.87	.69	3.13	1.48	76	70.4	5	4.6	6	5.6	7	6.5	14	13	.71
13 (IC)	.91	.59	3.37	1.32	84	77.8	6	5.6	2	1.9	6	5.6	10	9.3	.62
14 (LC)	.77	.51	2.38	1.67	46	42.6	14	13	8	7.4	15	13.9	25	23.1	.53
15 (LC)	.90	.57	2.94	1.48	67	62	5	4.6	10	9.3	15	13.9	11	10.2	.61
16 (LC)	.81	.53	2.10	1.58	36	33.3	11	10.2	9	8.3	32	29.6	20	18.5	.55
17 (CI)	.59	.23	1.44	1.57	23	21.3	5	4.6	12	11.1	24	22.2	44	40.7	.24
18 (LC)	.76	.63	2.16	1.65	40	37	9	8.3	13	12	20	18.5	26	24.1	.66
19 (CI)	.51	.60	1.63	1.82	36	33	3	2.8	7	6.5	9	8.3	53	49.1	.63
20 (CI)	.83	.57	2.85	1.66	70	64.8	4	3.7	16	14.8	0	0	18	16.7	.60
21 (LC)	.89	.59	2.87	1.44	59	54.6	9	8.3	19	17.6	9	8.3	59	54.6	.62
22 (CI)	.76	.72	2.76	1.73	67	62	6	5.6	3	2.8	6	5.6	26	24.1	.75
23 (CI)	.84	.67	2.34	1.62	49	45.4	2	1.9	11	10.2	29	26.9	17	15.7	.70
24 (CI)	.74	.58	2.30	1.76	51	47.2	6	5.6	3	2.8	20	18.5	28	25.9	.61
25 (CI)	.84	.64	2.77	1.61	64	59.3	3	2.8	10	9.3	14	13	17	15.7	.68
26 (CI)	.62	.42	1.89	1.76	39	36.1	3	2.8	14	13	11	10.2	41	38	.44
27 (II)	.87	.65	3.22	1.48	82	76	3	2.8	2	1.9	7	6.5	14	13	.68
28 (CI)	.74	.54	2.54	1.78	61	56.5	3	2.8	5	4.6	11	10.2	28	25.9	.55
29 (CI)	.80	.73	2.81	1.69	68	63	6	5.6	2	1.9	10	9.3	22	20.4	.75
30 (CI)	.78	.49	2.74	1.67	62	57.4	8	7.4	10	9.3	4	3.7	24	22.2	.52
31 (CI)	.60	.65	1.55	1.58	21	19.4	14	13	11	10.2	19	17.6	43	39.8	.67
32 (CI)	.56	.68	1.60	1.69	26	24.1	14	13	7	6.5	13	12	48	44.4	.70
33 (II)	.73	.66	2.00	1.59	30	27.8	17	15.7	13	12	19	17.6	29	26.9	.69
34 (CI)	.73	.62	2.39	1.80	57	52.8	2	1.9	4	3.7	16	14.8	29	26.9	.64
35 (CI)	.69	.59	2.09	1.79	44	40.7	10	9.3	20	18.5	0	0	34	31.5	.61
36 (CI)	.74	.75	2.37	1.73	50	46.3	9	8.3	8	7.4	13	12	28	25.9	.77
37 (CI)	.59	.61	1.72	1.76	36	33.3	2	1.9	10	9.3	16	14.8	44	40.7	.63
38 (CI)	.62	.66	2.04	1.86	47	43.5	5	4.6	2	1.9	13	12	41	38	.69
39 (CI)	.71	.73	2.15	1.79	49	45.4	3	2.8	2	1.9	23	21.3	31	28.7	.75
40 (CI)	.72	.72	2.30	1.79	53	49.1	4	3.7	3	2.8	18	16.7	30	27.8	.74

Notes: D_i, item difficulty; r, corrected item-total correlations; M, mean; SD, standard deviation; LC, literal comprehension; CI, causal inference; II, informative inference. The description of the items is available in the supplementary file.
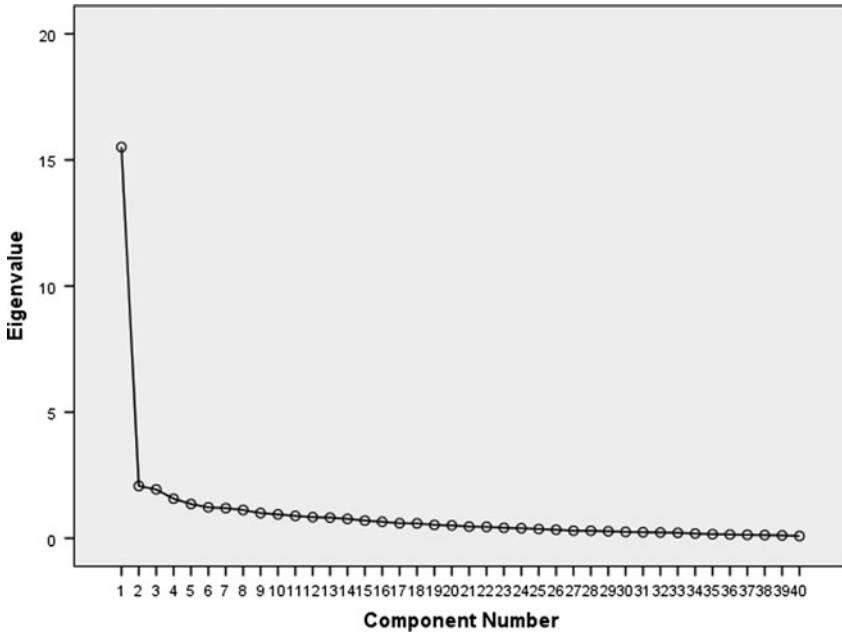
Figure 1. Scree plot.

$M = 105.62$, $SD = 45.97$; 6 years old: $M = 115.08$, $SD = 39.27$; $\chi^2(2) = 5.13$, $p = .08$) differences were found.

DISCUSSION

The aim of this study was to investigate the psychometric properties of the DATLC$_{4-6}$ regarding items difficulty and discrimination, and test dimensionality and reliability. The results indicate that the DATLC$_{4-6}$ has adequate psychometric properties. The obtained values for the items difficulty and discrimination indexes were adequate. The lack of items with severely low difficulty indexes (i.e. very difficult items) is congruent with the DA characteristics: providing prompts gives children opportunities to learn and improve their performance in the test (Caffrey et al., 2008).

Evidence for a one factor structure was obtained. A similar factor structure has been reported previously for reading comprehension tests (e.g. Cadime et al., 2013), suggesting that listening and reading text comprehension are both one-dimensional constructs.

A high Cronbach's alpha coefficient was obtained for the DATLC$_{4-6}$. The value obtained for our test is higher than that obtained by Elleman et al. (2011), but the differences may be related to the number of items in each test.

The main limitation of this study is related to the sample used. The proportions of students from private and public schools were not completely representative of the Portuguese preschooler population, and the sample size is relatively small. Future research should replicate this study with a larger number of participants and a more representative sample. This increase in sample size will allow the use of item response theory analyses to develop a shorter version of the test.

It is also necessary to collect criterion-related evidence of validity for the $DATLC_{4-6}$ before it can be used in developmental research and in clinical and educational settings to assess children's listening comprehension competences. Future studies should also test for measure invariance when investigating group differences (e.g. gender or age-related).

The results, with or without prompts, that are the best predictors of achievement in elementary education in reading and listening comprehension must also be addressed in future research.

The $DATLC_{4-6}$ is one of the few standardized tests for the assessment of listening comprehension and is the first to be specifically constructed for Portuguese preschoolers. The major potentiality of this test is the early identification of kindergarten children with low scores in listening comprehension, who may be at risk for future difficulties in reading acquisition. Specific intervention programmes can be designed for these children to promote this competence, thereby preventing future difficulties.

Supplementary material

Supplementary material is available via the "Supplementary" tab on the article's online page (http://dx.doi.org/10.1080/17405629.2014.938628).

REFERENCES

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Cadime, I., Ribeiro, I., Viana, F. L., Santos, S., Prieto, G., & Maia, J. (2013). Validity of a reading comprehension test for Portuguese students. *Psicothema*, *25*, 384–389. doi:10.7334/psicothema2012.288

Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The predictive validity of dynamic assessment: A review. *The Journal of Special Education*, *41*, 254–269. doi:10.1177/0022466907310366

Campione, J. C., & Brown, A. L. (1987). Linking dynamic assessment with school achievement. In S. C. Lidz (Ed.), *Dynamic testing* (pp. 82–115). New York, NY: Guilford.

Elleman, A. M., Compton, D. L., Fuchs, D., Fuchs, L. S., & Bouton, B. (2011). Exploring dynamic assessment as a means of identifying children at risk of developing comprehension difficulties. *Journal of Learning Disabilities*, *44*, 348–357. doi:10.1177/0022219411407865

Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*, 127–160.

Kendeou, P., Bohn-Gettler, C., White, M. J., & van den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*, *31*, 259–272. doi:10.1111/j.1467-9817.2008.00370.x

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, *101*, 765–778. doi:10.1037/a0015956

Makdissi, H., & Boisclair, A. (2006). Interactive reading: A context for expanding the expression of causal relations in preschoolers. *Written Language & Literacy*, *9*, 177–211.

Nation, K. (2005). Children's reading comprehension difficulties. In M. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 248–265). Oxford: Blackwell.

Pallant, J. (2005). *SPSS survival manual* (2nd ed.). New York, NY: McGraw-Hill.

Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, *38*, 36–76. doi:10.1598/RRQ.38.1.3

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Van den Broek, P., Kendeou, P., Kremer, K., Lynch, J. S., Butler, J., White, M. J., & Lorch, E. P. (2005). Assessment of comprehension abilities in young children. In S. Stahl & S. Paris (Eds.), *Children's reading comprehension and assessment* (pp. 107–130). Mahwah, NJ: Erlbaum.

Van Kleeck, A. (2008). Providing preschool foundations for later reading comprehension: The importance of and ideas for targeting inferencing in storybook-sharing interventions. *Psychology in the Schools*, *45*, 627–643. doi:10.1002/pits.20314