Listening Comprehension Assessment: Validity Studies of Two Vertically Scaled Tests for Portuguese Students*

Evaluación de la Comprensión del Lenguaje Oral: Estudios de Validez de Dos Tests con Escalamiento Común para Alumnos Portugueses

Fernanda Leopoldina Viana **
Sandra Santos ***
Iolanda Ribeiro ****
Séli Chaves-Sousa *****
Sara Brandão
Irene Cadime
University of Minho, Portugal
José Maia ******
University of Porto, Portugal

**   Associate Professor at the Institute of Education of the University of Minho. Deputy director of the Research Centre on Child Studies (CIEC). Investigator-in-charge of a project with financial support of the Foundation for Science and Technology (FCT). E-mail: fviana@ie.uminho.pt

*** Fellow at the Research Center on Psychology (CIPsi) since May 2010. E-mail: sandra.css@gmail.com

**** Assistant Professor at the School of Psychology of the University of Minho. Member of the Research Center on Psychology (CIPsi). E-mail: iolanda@psi.uminho.pt

***** Fellows at the Research Center on Psychology (CIPsi) of the University of Minho, School of Psychology. E-mails: seli.chaves.sousa@gmail.com, sarabran@gmail.com, renecadime@psi.uminho.pt

***** Full time professor at the Faculty of Sport Sciences and Physical Education, University of Porto. E-mail: jmaia@fade.up.pt

Abstract

This study aims to gather construct- and criterion-related validity evidence for the Test of Listening Comprehension of Narrative Texts (TLC-n) and the Test of Listening Comprehension of Expository Texts (TLC-e), each composed of a vertically scaled form for Portuguese students from the first to fourth grades of primary school. A sample of 1342 students and a sample of 1168 students participated in the study of the TLC-n and TLC-e forms' dimensionality, respectively. Two hundred and eighty students participated in the criterion-related validity study of the tests. Confirmatory factorial analyses were performed for the study of the test forms' dimensionality. To collect criterion-related validity evidence, Pearson correlations with measures of listening and reading related skills were calculated. Results provide evidence of an acceptable fit for the one-factor solution, for all forms, and statistically significant correlations between the test forms and the external criteria measures. There is evidence for the validity of the test forms.
Keywords
construct validity; criterion validity; expository texts; listening comprehension assessment; narrative texts

Resumen

Los objetivos de este trabajo fueron recoger evidencia de validez de constructo y de criterio para dos tests, el *Test of Listening Comprehension of Narrative Texts* (TLC-n) y el *Test of Listening Comprehension of Expository Texts* (TLC-e), con cuatro versiones escaladas verticalmente para evaluar la comprensión oral de los alumnos portugueses de enseñanza primaria. Una muestra de 1342 alumnos y una muestra de 1168 alumnos participaron en el estudio de la dimensionalidad del TLC-n y TLC-e. Participaron 260 alumnos en el estudio de validez de criterio de los tests. La dimensionalidad de los tests se examinó mediante un procedimiento de análisis factorial confirmatorio. Para recoger datos sobre evidencia de validez de criterio, se calcularon las correlaciones de Pearson con otras pruebas de comprensión oral y de lectura. Los resultados revelaron que el modelo de un factor se ajusta a los datos. Se obtuvieron coeficientes de correlación estadísticamente significativos entre el TLC-n y TLC-e y los otros tests. Los resultados apoyan la validez de las versiones de los tests.
Palabras clave
evaluación de la comprensión del lenguaje oral; textos expositivos; textos narrativos; validez de constructo; validez de criterio

Fernanda Leopoldina Viana, Sandra Santos, Iolanda Ribeiro,
Séli Chaves-Sousa, Sara Brandão, Irene Cadime, José Maia

Listening comprehension refers to the ability to extract meaning from spoken discourse (Snowling & Hulme, 2005). This is a core component of the Simple View of Reading model (Gough & Tunmer, 1986; Hoover & Gough, 1990) according to which reading comprehension equals the product of word recognition and listening comprehension.

Difficulties with decoding or listening comprehension can lead to different types of reading problems: (1) children who have difficulties only with written language, due to decoding problems, but have no difficulties in comprehending the spoken language; (2) children who, despite good decoding performance, have problems with reading comprehension; and (3) children who simultaneously have a decoding problem and difficulties in comprehension of spoken language (Aarnoutse, Brand-Gruwel, & Oduber, 1997; Hulme & Snowling, 2011; Nation, 2005). The assessment of listening comprehension with valid measures has a major impact on a broader assessment of reading difficulties, facilitating the diagnosis of reading disabilities and the identification of different reading profiles (Cain, Oakhill, & Lemmon, 2005; Nation, 2005; Spooner, Baddeley, & Gathercole, 2004)"

The relationship between reading comprehension, listening comprehension and decoding changes over the stages of reading acquisition: in the early phases, decoding proficiency explains a great part of the variance in reading comprehension. When students have mastered basic decoding skills, listening comprehension becomes a better predictor of reading comprehension (Florit & Cain, 2011; Vadasy & Sanders, 2009).

Both reading and listening comprehension imply the ability to understand and evaluate meanings of the text message, but in listening comprehension, there are no demands of decoding the written text. According to the Structure Building Framework of Comprehension, the same comprehension processes take place in the comprehension of written or spoken material (Gernsbacher, Varner, & Faust, 1990). This view is also supported by the research review performed by Hedrick and Cunningham (2002) and the study of Duke and Carlisle (2011),

which suggest that both reading and listening comprehension involve common processes.

The theoretical similarities between listening and reading comprehension, along with empirical studies focusing on the factorial structure of similar tests that assess the comprehension of texts, provide empirical evidence for listening comprehension as a one-dimensional construct. Factorial analysis of the listening comprehension test items that incorporate a battery of standardized tests constructed by the Dutch National Institute for Educational Measurement (Gillijns & Verhoeven, 1992) yielded a single factor. A reading-related assessment battery in Malay (Lee, 2008) that integrates several tests, including a listening comprehension assessment, was also studied through confirmatory factorial analysis.

Listening comprehension emerged as a distinct factor. In the study of the construct validity of the Computerized Battery of Oral Language – Bilo (Joly & Dias, 2009), listening comprehension assessed through the interpretation of orally presented texts also emerged as a unique factor, which explained 38.4% of the total variance of the results. These results suggest that the interpretation of a single score for listening comprehension is reliable.

Research on listening comprehension has provided evidence for the relationships between this skill and other reading- and listening-related skills. From a review of studies in 1997, Aarnoutse, Brand-Gruwel, and Oduber concluded that the relationship between listening and reading comprehension is strong, with correlations that range between 0.5 and 0.6. According to Nation (2005), although the magnitude of the correlations may vary, the research so far indicates the existence of a strong relationship between reading and listening comprehension. Teacher ratings have also been found to be significantly correlated with listening comprehension (Gilmore & Vance, 2007).

An influence of vocabulary on reading comprehension has been systematically observed (Cain, Oakhill, & Bryant, 2004; Cain, 2010; Joshi, 2005), but this relationship is also encountered for listening comprehension (Hagtvet, 2003). A study by Ouellette (2006), with children in grade 1 and grade 6, obtained correlations of 0.302 and 0.313,

respectively. Working memory is another variable that influences listening comprehension as it is responsible for processing information during text comprehension (Daneman & Merikle, 1996; Just & Carpenter, 1992; McInnes, Humphries, Hogg-Johnson, & Tannock, 2003). If processing demands result in an overload of working memory, because of its limited resources, it is difficult to maintain the information necessary to construct a semantic representation of the text (Cain, 2006). Listening comprehension might require a greater processing load because the pace of processing is set by the speaker and the text does not remain available (Roch, Florit, & Levorato, 2012).

Taking these aspects into consideration, this brief research aims to present empirical validity evidence for the Test of Listening Comprehension of Narrative Texts (TLC-n) and the Test of Listening Comprehension of Expository Texts (TLC-e), which are aimed at assessing Portuguese students from the first to fourth grades of primary school. The specific goals of this study were to collect construct-related and criterion-related validity evidence.

Method

Participants

Two samples were used in this study of construct validity. A sample of 1342 students (309 first graders, 239 second graders, 360 third graders and 344 fourth graders) participated in the study of the TLC-n dimensionality. In each grade, the sample exhibited similar sex distribution: male students represented 54.4% (n = 168) of the first grade sample, 52% (n = 124) of the second grade sample, 48.9% (n = 176) of the third grade sample and 51.5% (n = 177) of the fourth grade sample. In the study of the dimensionality of the TLC-e, 1168 students participated (285 first graders, 279 second graders, 287 third graders and 317 fourth graders). Male students represented 56.1% (n = 160) of the first grade sample, 49.1% (n = 137) of the second grade sample, 56.1% (n = 161) of the third grade sample and 51.4% (n = 163) of the fourth grade.

A sample of 62 students from the first grade, 66 from the second grade, 68 from the third grade and 84 from the fourth grade participated in the criterion-related validity study of the TLC-n and the TLC-e. Male students represented 58.1% (n = 36) of the first graders, 48.5% (n = 32) of the second graders, 58.8% (n = 40) of the third graders and 50% (n = 42) of the fourth graders. All participants attended Portuguese public schools in the north of Portugal. None spoke Portuguese as a second language or had any recognized special educational needs. Children were between 6 and 12 years of age.

Instruments

Test of Listening Comprehension of Narrative Texts (TLC-n) and the Test of Listening Comprehension of Expository Texts (TLC-e). These tests were used for the dimensionality study. They measure the ability to comprehend narrative and expository texts that are orally presented to students in short passages using a previously recorded file dictated by a professional. Each test is composed of four vertically scaled forms, each one designed to assess students from the first, second, third and fourth grades, respectively: TLC-n-1, TLC-n-2, TLC-n-3, TLC-n-4 and TLC-e-1, TLC-e-2, TLC-e-3, TLC-e-4.

Each test form is composed of 30 items, of which 10 are anchor items. Items are multiple-choice questions, with each one assessing literal comprehension (LC), inferential comprehension (IC), critical comprehension (CC) or reorganization (R). All tests' forms show high reliability coefficients (Person Separation Reliability – PSR, Item Separation Reliability – ISR and the Kuder-Richardson formula 20 – KR20), ranging from 0.7 and 0.98 for the TLC-n forms and from 0.72 to 0.95 for the TLC-e forms. In addition to these instruments, others were used to gather criterion-related validity evidence for the TLC-n and the TLC-e.

The Test of Reading Comprehension of Narrative Texts (TRC-n) and the Test of Reading Comprehension of Expository Texts (TRC-e). The goals of these tests are to assess, respectively, reading comprehension of narrative and expository texts. Each test is

composed of three vertically scaled forms, with each one having been developed to assess students from the second, third and fourth grades of primary school. The items are multiple choice and assess LC, IC, R and CC.

The forms of the TRC-n are composed of 27 items, from which six are anchor items, and the forms of the TRC-e comprise 33 items, eight of which are anchor items. All of the tests' forms show high reliability coefficients (Person Separation Reliability – PSR, Item Separation Reliability – ISR and the Kuder-Richardson formula 20 – KR20), ranging from 0.7 and 0.96 for the TRC-n forms and from 0.72 to 0.95 for the TRC-e forms.

Vocabulary (Wechsler, 2003). This test is a subtest of the Wechsler Intelligence Scale for Children (WISC-III), is individually administered and demands that students provide the definition of a given word.

Digit Span (Wechsler, 2003). This test is a subtest of the WISC-III, is individually administered, and is intended to assess working memory.

Teachers' ratings of students' reading and listening comprehension skills. Teachers rated their students on a scale ranging from 1 (poor) to 5 (excellent) based on their performance in tasks that they were provided during the school year.

Procedure

The administration of tests was authorized by the Portuguese Ministry of Education and the school boards of the educational establishments. Informed consent was obtained from children's parents or legal tutors. Listening and reading comprehension tests were administered collectively, during the regularly scheduled class time, in different sessions. Individually administered tests took place in a school room. No time limit was set for several tests, whose application was performed by trained psychologists.

Participants of the study of the dimensionality of the TLC-n and the TLC-e performed, respectively, the TLC-n or the TLC-e forms that were specific to their academic grade in a single session. Participants in the criterion-related validity study

were administered every instrument in four collective sessions and in one individual session.

Data analyses

Confirmatory factor analysis (CFA) was used to test the one factor measurement model of the TLC-n and the TLC-e forms' results. Software M*plus* 6.1 (Muthén & Muthén, 2010) was used. The chi-square test of model fit index (χ^2) was calculated to assess how well the model reproduced the covariance matrix (Byrne, 2011) (Byrne, 2011. A good fit is suggested when the probability associated with the chi-square statistic is nonsignificant. However, with large samples, this index is usually significant (Harrington, 2009). Therefore, four descriptive fit indices that are less vulnerable to sample size were then determined: (a) the Comparative Fit Index (CFI), (b) the Tucker-Lewis Index (TLI), (c) the Root-Mean-Square Error of Approximation (RMSEA), and (d) the Weighted Root Mean Square Residual (WRMR). CFI and TLI values of 0.9 or greater are indicators of a good model fit (Byrne, 2011). With RMSEA, a value below 0.05 indicates a good fit (Schreiber, Nora, Stage, Barlow, & King, 2006). The WRMR is an appropriate index for binary data (Brown, 2006). For a good fit, a cut-off value lower than 1.0 is expected (Yu, 2002).

Pearson correlations were calculated with IBM SPSS Statistics 20 for the collection of criterion-related validity evidence. The goal was to analyze the relationships between the TLC-n and TLC-e and the measures used as external criteria of reading comprehension, vocabulary, memory and teachers' ratings of students' reading and listening comprehension skills.

Results

Construct-related validity evidence

A one-factor model was tested for each test form of the TLC-n and the TLC-e. Results are presented in Table 1. Chi-square statistics values were statistically significant for all of the forms of the TLC-n and the TLC-e. CFI and TLI indices were approxi-

mately 0.9, except for some models whose values were below the cutoff value. This was the case for the CFI=0.824 and TLI=0.811 of the TLC-n-1 model, and the CFI=0.848 and TLI=0.837 of the TLC-e-1 model. In all of the TLC-n and the TLC-e forms, the RMSEA was lower than 0.05. WRMR values were less than 1.0 in the TLC-n-2, the TLC-e-2 and the TLC-e-3 models; the other test forms' values were slightly greater than 1.0.

Criterion-related validity evidence

Table 2 presents the correlations between the TLC-n and the TLC-e and the external criteria. Correlations between the forms of the TLC-n and TLC-e and the forms of the TRC-n and the TRC-e were positive and statistically significant, ranging from low (0.25) to moderate correlations (0.64). Low to moderate correlations were found between the TLC-n and TLC-e forms and the teachers' evaluation concerning listening and reading comprehension skills. Correlations with listening comprehension skills ranged from 0.28 to 0.54 and correlations with reading comprehension ranged between 0.28 and 0.63. Correlations between the TLC-n and TLC-e forms and vocabulary were low, ranging from 0.26 to 0.44. Low correlations were also found between the second grade forms of the TLC-n and the TLC-e and the digit span results. Correlations of these tests for the remaining academic grades were not statistically significant.

Discussion

In this study the construct- and criterion-related validity of two tests of listening comprehension were analyzed, one with narrative texts (TLC-n) and the other with expository texts (TLC-e), each composed of four vertically scaled forms for the assessment of students from the first to fourth grades of primary school.

Regarding construct validity, a one-factor model was tested using CFA for each test form. Given the sample sizes, chi-square values were statistically significant for all TLC-n and TLC-e test forms. As it is well known that this statistic is sample sensitive, and it should not be considered as the only indicator of global model fit (Harrington, 2009). CFI and TLI indices found in the one-factor solution to all test forms were approximately 0.9, except for the TLC-n-1 and TLC-e-1 models that were below that recommended cut-off. On the other hand, in all of the TLC-n and TLC-e forms, the RMSEA values and the WRMR values indicated a good fit to the data. Taken together, these fit indices support an acceptable fit for the one-factor solution of all of the test forms.

Data concerning criterion-related validity indicated that the TLC-n and TLC-e forms' results were positively correlated with the results of the TRC-n and TRC-e forms, ranging from 0.25 to 0.64.

The literature supports the view that the relationship between listening and reading comprehen-

Table 1
Global Fit Indices

Test Form	χ^2	df	CFI	TLI	RMSEA	WRMR
TLC-n-1	506.17*	405	0.824	0.811	0.028	1.027
TLC-n-2	451.51*	405	0.951	0.948	0.019	0.925
TLC-n-3	517.92*	405	0.898	0.891	0.028	1.02
TLC-n-4	506.53*	405	0.893	0.885	0.027	1.019
TLC-e-1	569.36*	405	0.848	0.837	0.036	1.093
TLC-e-2	497.66*	405	0.9	0.893	0.029	0.999
TLC-e-3	472.33*	405	0.941	0.936	0.024	0.947
TLC-e-4	514.58*	405	0.871	0.862	0.029	1.023

Note. *p<0.05; df = Degrees of freedom; TLI = Tucker–Lewis Index; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; WRMR=Weighted Root Mean Square Residual

Source: own work

Fernanda Leopoldina Viana, Sandra Santos, Iolanda Ribeiro,
Séli Chaves-Sousa, Sara Brandão, Irene Cadime, José Maia

sion increases with advancing school grade, as a result of the mastering of decoding skills (Hudson, Pullen, Lane, & Orgesen, 2009; Rasinski, Homan, & Biggs, 2009). Additionally, according to the model of the Simple View of Reading and the research review of Aarnoutse, Brand-Gruwel and Oduber (1997), it would be expected that correlations between the TLC-n and TLC-e forms and reading comprehension assessments would be stronger. However, these expectations were not found by our results. This could be interpreted in line with the findings of Carver (1998), who suggests that there is no evidence to state that listening comprehension becomes more important after the students have mastered the basic decoding skills. Ouellette (2006) also found lower correlations between listening and reading comprehension in grade 1 ($r = 0.288$) and grade 6 ($r = 0.275$).

Teachers' rating of listening and reading comprehension skills presented low to moderate correlations with the TLC-n and TLC-e forms. The magnitude of correlations of the TLC-n and TLC-e forms with listening comprehension ratings were slightly lower than the correlations with reading comprehension ratings. It is possible that teachers based their evaluation of listening comprehension on similar paper-and-pencil reading tests (Laitusis, 2012). Listening comprehension and vocabulary correlations were low, ranging from 0.26 to 0.44. These values are in accordance with the results obtained by Ouellette (2006).

Correlations between the TLC-n and TLC-e forms and the digit span results were statistically significant but low for the second grade forms of the TLC-n and TLC-e. No statistically significant correlations with digit span were found in the other test forms. In future studies correlations between the two tests of listening comprehension and working memory will provide further evidence of the relationships between these two variables. It is important to highlight that similar results were found both for the TLC-n and TLC-e, which support the findings of other studies (e.g., Kintsch & Young, 1984; Roller & Schreiner, 1985) in which no differences were found in memory between different text genres.

Although this study provides new data regarding listening comprehension assessment and its relationship with variables more frequently associated with reading comprehension assessments, some limitations should be mentioned. To collect criterion-related validity evidence, there are no other validated measures of the same construct for Portuguese students to which our results can be compared.

The TLC-n and TLC-e forms emerge as an important contribution for the Portuguese assessment context. To date, listening comprehension of texts has been assessed with informal measures, and students' performance has been scored according to the percentage of correct answers.

TABLE 2

Correlation Matrix for the TLC-n and the TLC-e Forms' Results and Other External Criteria

Test form	TRC-n	TRC-e	LC assessed by teachers	RC assessed by teachers	Vocabulary	Digit Span
TLC-n-1	n.a.	n.a.	0.32*	0.31*	0.4**	0.08
TLC-n-2	0.5***	0.64***	0.52***	0.47***	0.44***	0.36**
TLC-n-3	0.42***	0.25*	0.3*	0.35**	0.36**	0.02
TLC-n-4	0.42***	0.48***	0.31*	0.33**	0.26*	0.11
TLC-e-1	n.a.	n.a.	0.28*	0.35**	0.43**	0.09
TLC-e-2	0.46***	0.59***	0.47***	0.55***	0.29*	0.36**
TLC-e-3	0.41**	0.32**	0.3*	0.28*	0.35**	0.11
TLC-e-4	0.47***	0.48***	0.54***	0.63***	0.36**	0.11

Notes: LC=Listening comprehension; RC=Reading comprehension; * $p<0.05$; ** $p< 0.01$; *** $p< 0.001$; n.a. = not applicable (first graders' reading comprehension was not assessed)
Source: own work

Another major contribution of these tests is the possibility of assessing listening comprehension performance separately for narrative and expository texts, which may be important for clarifying academic difficulties in specific areas more related to one or the other text genre. Furthermore, each test is composed of four vertically scaled forms, enabling the tracking of students' performance in each test throughout primary school.

The TLC-n and the TLC-e forms could also have a major impact on the assessment of reading disabilities as they contribute to the identification of the reading skill profiles of students. These profiles could lead to the implementation, by teachers, of more effective strategies for prevention and intervention in cases where there are listening comprehension problems in the classrooms, improving students' results and their academic success (González, Pérez, Díaz, & Rodríguez, 2002).

This study provides validity evidence for the use of the TLC-n and TLC-e forms as valid measures of listening comprehension. The results of each form can be converted to percentiles and standardized scores. This enables psychologists to use a total score in each test form to assess students' performance and to compare performances among students (Smith, 2002). This study provides evidence that the TLC-n and the TLC-e forms are effective measures of listening comprehension.

References

Aarnoutse, C., Brand-Gruwel, S., & Oduber, R. (1997). Improving reading comprehension strategies through listening. *Educational Studies, 23*(2), 209–227.

Brown, T. (2006). *Confirmatory factor analysis for applied research.* New York: The Guilford Press.

Byrne, B. (2011). *Structural equation modeling with Mplus: Basic concepts, applications and programming.* New York: Routledge Academic.

Cain, K. (2006). Children's reading comprehension: The role of working memory in normal and impaired development. In S. J. Pickering (Ed.). *Working memory and education* (pp. 62–91). San Diego, CA: Academic Press.

Cain, K. (2010). *Reading development and difficulties.* Chichester, UK: BPS Blackwell.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96*(1), 31–42.

Cain, K., Oakhill, J. V., & Lemmon, K. (2005). The relation between children's reading comprehension level and their comprehension of idioms. *Journal of Experimental Child Psychology, 90,* 65–87.

Carver, R. P. (1998). Predicting reading level in grades 1 to 6 from listening level and decoding level: Testing theory relevant to the simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 10,* 121–154.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422–433.

Duke, N. K., & Carlisle, J. (2011). The development of comprehension. In M. Kamil, P. D. Pearson, E. Moje, & P. Afflerbach (Eds.). *Handbook of reading research: Volume IV* (pp. 199–228). New York: Routledge.

Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review, 23*(4), 553–576.

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 16*(3), 430–445.

Gillijns, P., & Verhoeven, L. (1992). Het CITO leerlingvolgsysteem: Met het oog op de praktijk [The CITO pupil monitoring system: Focus on practice]. *Pedagogische Studiën, 69,* 291–296.

Gilmore, J., & Vance, M. (2007). Teacher ratings of children's listening difficulties. *Child Language Teaching and Therapy, 23*(2), 133–156.

González, L., Pérez, J., Díaz, F., & Rodríguez, N. (2002). Elaboración y evaluación de un programa de mejora de la comprensión oral. [Construction and assessment of a listening comprehension improvement program] *Psicothema, 14*(2), 293-299.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6–10.

Hagtvet, B. E. (2003). Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic. *Reading and Writing: An Interdisciplinary Journal, 16,* 505–539.

Harrington, D. (2009). *Confirmatory factor analysis.* Oxford, UK: Oxford University Press, Inc.

Hedrick, W. B., & Cunningham, J. W. (2002). Investigating the effect of wide reading on listening comprehension of written language. *Reading Psychology, 23*(2), 107–126.

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

Hudson, R. F., Pullen, P. C., Lane, H. B., & Orgesen, J. K. (2009). The complex nature of reading fluency: A multidimensional view. *Reading & Writing Quarterly, 25,* 4–32.

Hulme, C., & Snowling, M. J. (2011). Children's reading comprehension difficulties: Nature, causes, and treatments. *Current Directions in Psychological Science, 20*(3), 139–142.

Joly, M. C., & Dias, A. S. (2009). Evidências de validade de uma prova informatizada de linguagem oral – Bilo [Evidences of validity of a computerized instrument of oral language - Bilo]. *Psicologia: Teoria e Prática, 11*(2), 50–68.

Joshi, R. M. (2005). Vocabulary: A critical component of comprehension. *Reading & Writing Quarterly, 21*(3), 209–219.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122–149.

Kintsch, W., & Young, S. R. (1984). Selective recall of decision-relevant information from texts. *Memory and Cognition, 12*(2), 112–117.

Laitusis, C. C. (2012). Examining the impact of audio presentation on tests of reading comprehension. *Applied Measurement in Education, 23*(2), 153–167.

Lee, L. W. (2008). Development and validation of a reading-related assessment battery in Malay for the purpose of dyslexia assessment. *Annals of Dyslexia, 58*(1), 37–57.

McInnes, A., Humphries, T., Hogg-Johnson, S., & Tannock, R. (2003). Listening comprehension and working memory are impaired in attention-deficit hyperactivity disorder irrespective of language impairment. *Journal of Abnormal Child Psychology, 31*(4), 427–443.

Muthén, B. O., & Muthén, L. (2010). Mplus Version 6.1 [Software]. Los Angeles, CA: Muthén&Muthén.

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

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*(3), 554–566.

Rasinski, T., Homan, S., & Biggs, M. (2009). Teaching reading fluency to struggling readers: Method, materials, and evidence. *Reading & Writing Quarterly, 25*(2-3), 192–204.

Roch, M., Florit, E., & Levorato, M. C. (2012). The advantage of reading over listening text comprehension in Down syndrome: What is the role of verbal memory? *Research in Developmental Disabilities, 33*(3), 890–899.

Roller, C. M., & Schreiner, R. (1985). The effects of narrative and expository organizational instruction on sixth-grade children's comprehension of expository and narrative prose. *Reading Psychology: An International Quarterly, 6,* 27–42.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99,* 323–338.

Smith, E. V. J. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*(2), 205–231.

Snowling, M. J., & Hulme, C. (2005). *The science of reading: A handbook.* Oxford, UK: Blackwell Publishing Ltd.

Spooner, A. L. R., Baddeley, A. D., & Gathercole, S. E. (2004). Can reading accuracy and comprehension be separated in the Neale Analysis of Reading

Ability? *British Journal of Educational Psychology, 74,* 187–204.

Vadasy, P. F., & Sanders, E. a. (2009). Supplemental fluency intervention and determinants of reading outcomes. *Scientific Studies of Reading, 13*(5), 383–425.

Wechsler, D. (2003). *Escala de Inteligência de Wechsler para Crianças (WISC-III).* (3rd ed.). Lisboa: Cegoc-Tea Edições.

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes.* Los Angeles, CA: University of California.