Universidade do Minho
Escola de Psicologia

Bruna Regina da Silva Rodrigues

Reciprocal relations between listening comprehension, oral reading fluency, metacognitive skills, vocabulary and reading comprehension: A longitudinal study with students from grades 4 to 6

April, 2022

Reciprocal relations between listening comprehension, oral reading fluency, metacognitive skills, vocabulary and reading comprehension: A longitudinal study with students from grades 4 to 6
Bruna Regina da Silva Rodrigues

UMinho | 2022

Universidade do Minho
Escola de Psicologia

Bruna Regina da Silva Rodrigues

Reciprocal relations between listening comprehension, oral reading fluency, metacognitive skills, vocabulary and reading comprehension: A longitudinal study with students from grades 4 to 6

Doctoral Thesis
PhD in Applied Psychology

Work supervised by
Professor Maria Iolanda Ferreira da Silva Ribeiro
and
Doctor Irene Maria Dias Cadime

April, 2022

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho

Atribuição-Não Comercial – Sem Derivações
CC BY-NC-ND

https://creativecommons.org/licenses/by-nc-nd/4.0/

ACKNOWLEDGMENTS

To my supervisors for all their contributions to this dissertation. To Professor Iolanda Ribeiro, for the tranquility and assertiveness employed in responding to the challenges that arose along last 4 years and for the continuous positive reinforcement of my work. To Doctor Irene Cadime, for the pragmatism, critical sense and work capacity as well as for all knowledge transmitted along this journey.

To Professor Fernanda Leopoldina Viana for the constant incentive, the sharing of knowledge and the availability for collaborating in this research. To Professors Adriana Baptista, Celda Morgado and Isabel Alçada, as well as to Doctors Ana Maria Magalhães, Dulce Rodrigues, Inês Duarte and Rosário Alçada Araújo, for their collaboration through the writing of original texts for the developed tests. To Gina Voss, for your contribution to the data analysis in this research and to the expansion of my knowledge in the field of statistics. To Professor João Lopes and Doctor Sandra Santos for their encouraging words and suggestions for the improvement of this work.

To all school boards, teachers, parents/legal tutors and students for their participation in this research as well as to all psychologist colleagues who collaborated in the data collections.

To all of my research group and PhD colleagues. A special thanks to Albertina, for accepting all my help requests and cooperating with me promptly; to Cátia, my study buddy almost 12 years ago, for always listening and helping me with her generosity and serenity; to Céu, for her friendship, empathy, and sense of humor; and to Gabriela, for all the sharing moments as well as for her attention and care.

Aos meus pais, pelo exemplo de trabalho, determinação, coragem e altruísmo. À minha mãe, pelo suporte emocional e pelas mais expressivas manifestações de orgulho. Ao meu pai, pelo apoio incondicional expresso nos mais pequenos gestos. Sou uma filha cheia de sorte!

Ao Miguel, pela objetividade empregue nos momentos mais difíceis, pela confiança depositada em mim, e pelo amor demonstrado diariamente. Sei que se orgulha de mim e que vibra com as minhas conquistas. Também eu me orgulho dele e do que temos vindo a construir há quase 11 anos.

Finally, I thank the Portuguese Foundation for Science and Technology (FCT) for the Doctoral research grant (SFRH/BD/129582/2017), financed by national funds of the Ministry of Science, Technology and Higher Education (MCTES) and the European Social Fund (FSE) through the Northern Regional Operational Program, as well as the Psychology Research Centre (PSI/01662), School of Psychology, University of Minho, also financially supported by the FCT through the Portuguese State Budget (UIDB/PSI/01662/2020), that made this adventure possible.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

RELAÇÕES RECÍPROCAS ENTRE A COMPREENSÃO ORAL, A FLUÊNCIA DE LEITURA, AS COMPETÊNCIAS METACOGNITIVAS, O VOCABULÁRIO E A COMPREENSÃO DA LEITURA: UM ESTUDO LONGITUDINAL COM ALUNOS DO 4º AO 6º ANO DE ESCOLARIDADE

Resumo

O estudo da direccionalidade das relações entre competências de leitura, linguísticas e metacognitivas constitui um objetivo importante para a investigação e a intervenção em leitura uma vez que há evidências empíricas de que essas relações mudam ao longo do tempo. Deste modo, nesta tese procurou-se explorar as inter-relações longitudinais entre a compreensão da leitura, a compreensão oral, a fluência de leitura, o vocabulário e as competências metacognitivas entre o 4.º e o 6.º ano. Para tal, foram definidos os seguintes objetivos específicos: (1) a construção e a validação do Teste de Compreensão de Textos na Modalidade Oral e de Leitura para alunos do 5.º e do 6.º ano (TCTMO-$n_{5/6}$ e TCTML-$n_{5/6}$); (2) a construção e a validação do Teste de Avaliação da Fluência para alunos do 1.º ao 6.º ano (TAF_{1-6}); (3) a análise da existência de relações recíprocas entre a compreensão da leitura, a compreensão oral, a fluência de leitura, o vocabulário e as competências metacognitivas do 4.º ao 6.º ano. Os resultados dos estudos de construção e de validação do TCTMO-$n_{5/6}$ e do TCTML-$n_{5/6}$ forneceram evidências para a unidimensionalidade e independência local dos itens, bem como para a fidelidade e a validade dos referidos testes. Em relação ao desenvolvimento do TAF_{1-6}, os resultados mostraram que os procedimentos de equalização asseguraram a equivalência das pontuações entre as formas de teste dentro de cada ano. Como esperado, a média das pontuações equalizadas aumentou do 1.º para o 6.º ano. O TAF_{1-6} apresentou indicadores de fidelidade e de validade adequados. As relações recíprocas entre a compreensão da leitura, a compreensão oral, a fluência de leitura, o vocabulário e as competências metacognitivas foram testadas usando um modelo longitudinal cruzado. Os resultados deste estudo indicaram que a fluência de leitura é um preditor significativo da compreensão da leitura, que existe uma relação recíproca entre a compreensão da leitura e a compreensão oral assim como entre a compreensão da leitura e o vocabulário. Os resultados sugerem também que o desempenho em compreensão da leitura é melhor explicado pela fluência de leitura e pelo vocabulário do que pela fluência de leitura e pela compreensão oral ou pela fluência de leitura e pelo uso de estratégias de leitura. Os resultados desta tese são integrados numa reflexão final, onde são apresentados os principais contributos para a investigação, a avaliação e a intervenção na área da leitura, bem como orientações para estudos futuros.

Palavras-chave: avaliação, compreensão da leitura, compreensão oral, fluência de leitura, relações recíprocas.

RECIPROCAL RELATIONS BETWEEN LISTENING COMPREHENSION, ORAL READING FLUENCY, METACOGNITIVE SKILLS, VOCABULARY AND READING COMPREHENSION: A LONGITUDINAL STUDY WITH STUDENTS FROM GRADES 4 TO 6

Abstract

The study of the directionality of the relations among reading, linguistic, and metacognitive skills is an important goal for research and intervention in reading since there is empirical evidence that these relations change over time. Thus, this thesis aimed to explore the longitudinal interrelations between reading comprehension, listening comprehension, oral reading fluency, vocabulary, and metacognitive skills between fourth and sixth grades. For this purpose, the following specific goals were defined: (1) the construction and validation of the Test of Listening Comprehension and the Test of Reading Comprehension for fifth and sixth grade students (TLC-$n_{5/6}$ and TRC-$n_{5/6}$); (2) the construction and validation of the Test of Fluency Assessment for students from first to sixth grades (TAF_{1-6}); (3) the analysis of the existence of reciprocal relations between reading comprehension, listening comprehension, oral reading fluency, vocabulary, and metacognitive skills from fourth to sixth grades. The results of the studies focused on the construction and validation of the TLC-$n_{5/6}$ and the TRC-$n_{5/6}$ provided evidence for the unidimensionality and local independence of the items, as well as for the reliability and validity of these tests. Regarding the development of the TAF_{1-6}, the results showed that equating procedures guaranteed the equivalence of scores between test forms within each grade level. As expected, mean of the equated scores increased from first to sixth grades. The TAF_{1-6} showed adequate indicators of reliability and validity. The reciprocal relations between reading comprehension, listening comprehension, oral reading fluency, vocabulary, and metacognitive skills were tested using a longitudinal cross-lagged design. The results of this study indicated that oral reading fluency is a significant predictor of reading comprehension, that there is a reciprocal relation between reading comprehension and listening comprehension as well as between reading comprehension and vocabulary. The results also indicate that reading comprehension performance is better explained by oral reading fluency and vocabulary than by oral reading fluency and listening comprehension or by oral reading fluency and reading strategy use. The results of this thesis are integrated in a final conclusion, where the main contributions of this work to research, assessment, and intervention in the field of reading are discussed and guidelines for future studies are presented.

Keywords: assessment, listening comprehension, oral reading fluency, reading comprehension, reciprocal relations.

TABLE OF CONTENTS

Introduction **2**

Reading Assessment in the 2nd Cycle of Basic Education 5

The Development of Reading Comprehension throughout the 2nd Cycle of Basic Education 7

Thesis Aims 8

Thesis Structure 9

References 11

Developing and Validating Tests of Reading and Listening Comprehension for Fifth and Sixth Grade Students in Portugal **19**

Introduction 19

The Present Study 21

Study 1 22

Materials and Methods 22

Participants 22

Study Design and Measures 23

Procedure 24

Data Analyses 24

Results 26

Study 2 33

Method 33

Participants 33

Measures 33

Procedure 35

Data Analyses 35

Results 35

Discussion 37

References 41

Assessing Oral Reading Fluency Within and Across Grade Levels: Development of Equated Test Forms **46**

Introduction 46

The Present Study 47

Study 1 48
Method 48
Participants 48
Instrument 49
Design and Procedures 49
Statistical Analyses 50
Results 51
Study 2 56
Method 56
Participants 56
Measures 56
Procedures 57
Statistical Analyses 58
Results 58
Discussion and Conclusions 61
References 64
Reading, Linguistic, and Metacognitive Skills: Are the Relations Reciprocal Past the First School Years? 70
Introduction 70
Interrelations between Reading Skills 70
Interrelations between Reading Comprehension and Linguistic Skills 71
Interrelations between Reading Comprehension and Metacognitive Skills 73
The Present Study 73
Method 74
Participants 74
Measures 75
Procedure 76
Statistical Analyses 76
Results 77
Discussion 81
References 85
Conclusion 92

The Development of Reading Tests 92
The Interrelations between Reading Comprehension, Listening Comprehension, Oral Reading Fluency, Vocabulary, and Metacognitive Skills 95
Social Impact 98
Research 98
Assessment and Intervention 98
Final Considerations 99
References 101
APPENDIXES 107
Appendix A. Notification of the Ethical Approval from the Portuguese General Direction of Education 108
Appendix B. Ethical Approval from the Ethics' Committee for the Research in Human and Social Sciences of the University of Minho 110

ABBREVIATIONS

A – Anchor items

AIC – Akaike Information Criterion

AR – Abstract Reasoning

BAL – Bateria de Avaliação da Leitura

BIC – Bayesian Information Criterion

BP – The bird and the pine

CA – The camping

CC – Critical Comprehension

CFI – Comparative Fit Index

CO – A composition

DB – The Dentuça Bigodaça

DEC – Decoding

df – Degrees of freedom

DIBELS – Dynamic Indicators of Basic Early Literacy Skills

DIF – Differential Item Functioning

DS – Digit Span

EV – Variance Explained by the secondary dimensions

f1 – Test form 1

f2 – Test form 2

f3 – Test form 3

G_4 – Grade 4

G_5 – Grade 5

G_6 – Grade 6

GORT – Gray Oral Reading Test

GPA – Grade Point Average

HV – Holidays in the Village

IC –Inferential Comprehension

IEA – International Association for the Evaluation of Educational Achievement

IR – Items' Residuals

ISR – Item Separation Reliability

KR20 – Kuder-Richardson 20

LB – The lost bread

LC – Literal Comprehension

LT – Loose thoughts

MC – A mysterious chest

n.a. – Not applicable.

NEAT – Non-equivalent groups with anchor test

OECD – Organization for Economic Cooperation and Development

ORF – Oral Reading Fluency

PCA – Principal Component Analyses

PIRLS – Progress in International Reading Literacy Study

PISA – Programme for International Student Assessment

POR – Classification in the subject "Portuguese"

PSR – Person Separation Reliability

R – Reorganization

RC – Reading Comprehension

RMSEA – Root Mean Square Error of Approximation

ROC – Receiver Operating Characteristic

RSU – Reading Strategy Use

RV – Raw Variance explained by measures

SDi – Secondary Dimensions

SRMR – Standardized Root Mean Square Residual

ST – A very special trip

SVR – Simple View of Reading

T – Total of items

TAF – Teste de Avaliação da Fluência

TCTML-n – Teste de Compreensão de Textos na Modalidade de Leitura–Narrativo

TFL – Teste de Fluência de Leitura

TLC-n – Test of Listening Comprehension of Narrative texts

TLI – Tucker-Lewis Index

TLP – Teste de Leitura de Palavras

TMSFA – Texas Middle School Fluency Assessment

TPRI – Texas Primary Reading Inventory

TRC-n – Test of Reading Comprehension of Narrative texts

TRF – Test of Reading Fluency

U – Unique items

VOC – Vocabulary

VR – Verbal Reasoning

WISC–III – Wechsler Intelligence Scale for Children–III

WP – The history of white pencil

INDEX OF TABLES

Table 1. Items in Each Test Form of the TRC-n and TLC-n 24

Table 2. PCA of the Residuals and Reliability Coefficients by Test Form..... 27

Table 3. Descriptive Statistics of the Estimated Parameters by Test Form 28

Table 4. Descriptive Statistics and Correlations in the Fifth and Sixth Grades 36

Table 5. Descriptive Statistics of the Raw Scores 53

Table 6. Descriptive Statistics of the Equated Scores 55

Table 7. Descriptive Statistics, Test-retest Correlatins and Results of the Paired-samples t-test 59

Table 8. Descriptive Statistics and Correlations between TAF Test Forms and External Criteria from Grades 1 to 6 60

Table 9. Descriptive Statistics and Correlations between TAF Test Forms and TFL from Grades 1 to 6 61

Table 10. Descriptive Statistics and Correlations for Oral Reading Fluency (ORF), Listening Comprehension (LC), Reading Comprehension (RC), Vocabulary (VOC), and Reading Strategy Use (RSU) in Grades 4 (G4), 5 (G5) and 6 (G6) 78

Table 11. Fit Indexes for the Cross-lagged Models..... 79

INDEX OF FIGURES

Figure 1. Person-item Variable Map for the TRC-n_5 (left) and TRC-n_6 (right) 30

Figure 2. Person-item Variable Map for the TLC-n_5 (left) and TLC-n_6 (right) 32

Figure 3. NEAT Design 50

Figure 4. Model 2: Test of the Reciprocal Relations between ORF and RC and between LC and RC .. 80

Figure 5. Model 3: Test of the Reciprocal Relations between ORF and RC and between VOC and RC 80

Figure 6. Model 4: Test of the Reciprocal Relations between ORF and RC and between RSU and RC 81

INTRODUCTION

Introduction

Reading is a skill with a strong impact on the individuals' development and a powerful tool for lifelong learning and active citizenship. It is now widely recognized that reading literacy levels have a significant effect on the socio-economic development of societies (OECD, 2019a), which explains why literacy levels are considered one of the countries' growth indicators (Schleicher, 2019). The results of international reports point out that in the European space, one in five students has a poor reading performance (European Comission, 2012; OECD, 2019b). These two conflicting aspects – the unequivocal importance of literacy skills and the rate of students with low reading skills – reveal a problem that policymakers and researchers are seeking to address through solutions grounded in scientific knowledge. Specifically, national and international efforts are visible with regard to the assessment of students' literacy skills and the consequent analysis and interpretation of the results. An effort has been made by the Progress in International Reading Literacy Study (PIRLS) is an international study developed by the International Association for the Evaluation of Educational Achievement (IEA), conducted every 5 years, to assess the reading literacy performance of students in fourth grade. Regarding the results of Portuguese students in the 2016 assessment cycle, the following conclusions should be highlighted: (a) although Portugal scored 528 points (an average score significantly higher than the midpoint of the scale – 500 points), this score represents a significant drop of 13 points compared to the 2011 assessment cycle; (b) Portugal's overall average was 13 and 12 points below the overall average of the countries of the Organization for Economic Cooperation and Development (OECD) and the European Union, respectively; (c) only 7% of Portuguese students reached the "Advanced" level of performance and 38% reached the "High" level, percentages that are below the international median; (d) between 2011 and 2016, there was a significant decrease of 9% in the number of students at the "High" level and 5% of the number of students at the "Intermediate" level (Equipa dos Estudos Internacionais, 2017). In Portugal, all students at the end of fifth grade are assessed with a National Exam of Portuguese Language (*Prova de Aferição de Português*). The results in this exam are classified as: students were able to respond as expected (C); students were able to respond as expected, but they can still improve (CM); students revealed difficulty in answering (RD); students failed to respond as expected (NC) or did not respond (NR). In the school year 2020/2021, 70.2% and 86.6% of the students, who were part of the study obtained scores classified as "RD" or ""NC/NR", in the domains "speaking" and "reading and literary education", respectively. These results show that Portuguese Language is one of the subjects in which students have more difficulties in achieving the expected performance (i.e., scores classified as "C" or "CM") at the end of fifth grade (Serrão et al., 2021). The

Programme for International Student Assessment (PISA) is another international comparative study, conducted every 3 years by the OECD, to assess three domains, including reading literacy, in 15-year-old students attending at least seventh grade. In the 2018 edition, Portugal scored 492 points in reading literacy, being slightly above the OECD average (487 points). Despite an improving trend in the results, it is important to note the occurrence of a small decrease (although not statistically significant) in the students' achievement in the reading domain from the 2015 to the 2018 assessment cycle (Lourenço et al., 2019). It should also be noted that about 7% of students obtained the highest levels of reading proficiency, i.e., levels 5 and 6, and that approximately 20% of students did not reach the elementary level of reading proficiency, i.e., level 2 (Lourenço et al., 2019). Taken together, the results obtained by Portuguese students in the PIRLS, in the National Exam of Portuguese Language and in the PISA confirm the existence of a considerable percentage of students with reading difficulties between fourth and seventh grades.

These data highlight the need to continue to assess and monitor students' reading performance after the conclusion of the 1st cycle of basic education. For these purposes, it is important to recognize that reading comprehension is a complex activity that involves the activation of lower- and higher-level cognitive processes in order to assign meaning to written language (Spencer et al., 2020). The Simple View of Reading (SVR) is one of the most widespread conceptual models in research, advocating that decoding and listening comprehension are necessary skills but neither per se will be sufficient to explain reading comprehension performance (Hoover & Gough, 1990). The original model assumes that the components of the equation can be understood as dichotomous variables for the classification of students' abilities (has difficulties or lacks proficiency / has no difficulties or is proficient), which allows four reading profiles to be defined: (1) students with no difficulties in decoding, listening comprehension, and reading comprehension, (2) students with no difficulties in decoding but with difficulties in listening comprehension and reading comprehension, (3) students with no difficulties in listening comprehension but with difficulties in decoding and reading comprehension, (4) students with difficulties in decoding, listening comprehension, and reading comprehension (Duff & Clarke, 2011). Taking into account a developmental perspective, Hoover and Tunmer (2020d) revised the SVR considering decoding and listening comprehension as continuous variables. Thus, skill levels ranging from 0 (no skill) to 1 (perfect skill) are considered for each of the model's components. Throughout the 21st century, the results of empirical studies conducted on orthographies with distinct levels of depth have provided empirical evidence for the addition of oral reading fluency to the SVR (Cadime et al., 2017; Cutting & Scarborough, 2006; Padeliadu & Antoniou, 2014; Ribeiro et al., 2016; Salceda et al.,

2014). Therefore, it is important that the students identify words correctly, but also that they do it automatically, i.e., accurately and with proper speed. Considering the significant role of oral reading fluency in reading comprehension, from early to more advanced stages of reading development (Calet et al., 2015; Fernandes et al., 2017; Little et al., 2017; Padeliadu & Antoniou, 2014), as well as the heterogeneity that characterizes the students with reading comprehension difficulties, it is also relevant to consider this third component in the identification of students with reading difficulties. Thus, the characterization of students' profile of difficulties requires a rigorous and accurate assessment that allows the identification and description of their levels of proficiency in decoding, oral reading fluency, listening comprehension, and reading comprehension.

In Portugal, some recent legislation has fostered the implementation of measures aimed at promoting school success (Resolution of the Council of Ministers no. 23/2016 of 11 April, 2016), reading achievement (Resolution of the Council of Ministers no. 48-D/2017 of 31 March, 2017) and inclusive education (Decree-Law no. 54/2018 of July 6, 2018). The Decree-Law no. 54/2018 is based on a multilevel approach for the intervention in learning difficulties, where universal screenings and monitoring are essential conditions for its implementation. Therefore, the availability of measures that allow a large-scale assessment is crucial. These measures should also meet a set of conditions, such as (a) to be based on rigorous criteria in their construction process, both at the theoretical and methodological levels; (b) to be quick and easy to administer, score and interpret; and (c) to have adequate psychometric characteristics. Over the past few years, there has been a growing investment in the development and validation of reading assessment measures for the Portuguese school population, particularly for the 1st cycle of basic education (Cadime et al., 2013; Carvalho, 2010; Ribeiro et al., 2014; Silva et al., 2020; Sucena & Castro, 2008, 2012). However, there is still a scarcity of validated reading assessment instruments for Portuguese students attending later cycles of education.

In addition to the need of standardized measures to assess students' reading skills over time, the definition of guidelines for the promotion of students' literacy levels requires the collection of empirical evidence about the variables that explain performance in reading comprehension. In this context, several studies have been developed with samples of students in different grade levels and varying orthographic depths that attest to the predictive role of listening comprehension, oral reading fluency, vocabulary, and the use of reading strategies in reading comprehension (Bonifacci & Tobia, 2017; Fernandes et al., 2017; Frid & Friesen, 2020; Gomes-Koban et al., 2017; Høien-Tengesdal, 2010; Kershaw & Schatschneider, 2012; Kim et al., 2021; Kim & Wagner, 2015; Köse & Güneş, 2021; Lee & Chen, 2019; Liao et al., 2021; Swart et al., 2017; Verhoeven et al., 2019). There is also a range

of studies (although smaller) whose results suggest that reading comprehension also predicts the aforementioned reading, linguistic, and metacognitive skills (Berninger & Abbott, 2010; Kim, 2015; Klauda & Guthrie, 2008; Little et al., 2017; Muijselaar et al., 2017; Santos et al., 2020; Seigneuric & Ehrlich, 2005; Verhoeven et al., 2011; Verhoeven & van Leeuwe, 2008, 2012). Moreover, the changes in the strength of the relations between reading comprehension and its predictors as the students advance in schooling and in languages of different orthographic depth (Caravolas et al., 2019; Florit & Cain, 2011; Lervåg et al., 2018; Salceda et al., 2014; Verhoeven & van Leeuwe, 2008) justify the development of longitudinal studies that assess these skills at different timepoints. Thus, this thesis aims to provide a contribution to address the challenges inherent to the assessment of listening comprehension, reading comprehension and oral reading fluency, as well as analyze the interrelations between reading comprehension, listening comprehension, oral reading fluency, vocabulary, and metacognitive skills from fourth to sixth grades.

Reading Assessment in the 2nd Cycle of Basic Education

The construction of measures for text comprehension assessment (either in oral or reading modality) is a process that comprises a set of decisions related to the task itself, but also to the statistical procedures underlying the selection of the final items. The use of unpublished texts is an essential requirement to ensure the equality of circumstances for all students during assessment. Regarding the items, it is recommended to include questions that address different comprehension levels considering its potential influence on students' performance (Eason et al., 2012). With respect to the response format, multiple-choice items have been widely used in standardized comprehension tests (Magliano et al., 2007). Some of the main advantages of this type of response format are the reduced time to administer, the fidelity of the scoring procedures and the non-influence of the students' writing skills (Campbell, 2005; Kolen & Brennan, 2014; Spinillo et al., 2016).

The development of tests to assess reading skills requires original texts that are grade-level appropriate in terms of their linguistic characteristics, such as vocabulary and sentence complexity (Begeny & Greene, 2014). In addition, it is important to address other challenges as these are skills in which development is expected to occur over time as a result of the learning and training processes (Arnesen et al., 2017; Oakhill et al., 2019). One of these challenges relates to assessing students at multiple timepoints, within the same grade level or between different grade levels. Therefore, it is important to use standardized tests that allow for the assessment of inter- and intra-individual differences in order to compare each student's performance with their normative group but also to

monitor their progress over time (Cain, 2010). Using the same test to assess students in different grade levels can lead either to the occurrence of ceiling effects in the more advanced grade levels or to floor effects in early years of schooling, depending on the difficulty of the test material (Chaparro et al., 2018; McCoach & Yu, 2016). In contrast, administering distinct tests in different assessment moments does not allow the analysis of intra-individual changes if the test scores are not in the same metrics (de Ayala, 2009). The use of various test forms coupled with the application of statistical equating procedures has been strongly recommended as they avoid the aforementioned measurement biases and allow the direct comparison between students' scores collected at different timepoints (Kolen & Brennan, 2014). Horizontal equating places on a same metric scores from test forms constructed to be similar in difficulty level, allowing their interchangeable use, while vertical linking aligns the metrics from test forms with different difficulty levels, creating comparable score scales over time (Kolen & Brennan, 2014; Santi et al., 2016). In this sense, Item Response Theory models provide a viable solution to deal with the challenges underlying the use of comprehension tests that aimed at capturing intra- and inter-individual differences. In particular, the use of the Rasch model has several advantages, such as allowing the selection of items that are more appropriate to the skill level of the group they are intended to, the implementation of differential item functioning analyses and vertical equating of different forms of the same test (Kolen & Brennan, 2014; Prieto & Delgado, 2003). In the case of tests to assess oral reading fluency, the classical statistical equating methods (i.e., mean, linear, and equipercentile equating) have frequently been applied since the texts are constructed to be similar in terms of features such as difficulty, length, type, and word frequency (Albano & Rodriguez, 2012).

Reliability and validity are fundamental aspects underlying developing and evaluating tests. Reliability refers to consistency and stability of test scores and it is evaluated mainly in terms of reliability coefficients such as internal-consistency coefficients and test-retest coefficients (American Educational Research Association et al., 2014a). Validity refers to the degree to which theoretical and empirical evidences sustain proposed interpretations of test scores for a specific use (American Educational Research Association et al., 2014b). An important source of validity derives from analyses of the relation of the scores obtained in the created test to external variables, i.e., evidence based on relations to other variables (McCoach et al., 2013). Measures with high reliability and robust evidence of validity provide confidence to the inferences that can be drawn from the test scores. Therefore, these are crucial aspects to be addressed in measurement instruments' development.

In Portugal, the Battery of Reading Assessment [BAL – *Bateria de Avaliação da Leitura*], which is composed of tests to assess word reading, listening comprehension, and reading comprehension in

students attending the 1st cycle of basic education, was developed according to the aforementioned guidelines (Chaves-Sousa, Ribeiro, et al., 2017; Chaves-Sousa, Santos et al., 2015, 2016, 2017; Viana et al., 2015). Regarding the 2nd cycle of basic education, there are no validated tests to evaluate the comprehension of texts in oral and reading modalities with these same characteristics. To our knowledge, there is also no oral reading fluency test validated for Portuguese students with horizontally and vertically equated test forms.

The Development of Reading Comprehension throughout the 2nd Cycle of Basic Education

Decoding and listening comprehension explain between 40% and 80% of the reading comprehension variance in orthographies such as English (Kershaw & Schatschneider, 2012), Greek (Protopapas et al., 2007), Spanish (Nakamoto et al., 2008), French (Megherbi et al., 2006), Dutch (Verhoeven et al., 2019), Italian (Tobia & Bonifacci, 2015), European Portuguese (Cadime et al., 2017), Hebrew (Joshi et al., 2015), Arabic (Asadi & Ibrahim, 2018), and Mandarin/Cantonese (Peng et al., 2021). The empirical evidence about the explanatory power of these two skills with regard to reading comprehension difficulties is unequivocal, but it should not be dissociated from the recognition of the involvement of multiple processes and the simultaneous and concerted application of varied skills (Hoover & Tunmer, 2020c). Hence, more complex conceptual models that contain other components (e.g., oral reading fluency, vocabulary, previous knowledge, grammatical knowledge, comprehension monitoring) have been proposed in the scientific literature (Ahmed et al., 2016; Cervetti et al., 2020; Duke & Cartwright, 2021; Kim, 2017). Furthermore, the SVR is a static model, suggesting the existence of unidirectional relations between decoding/listening comprehension and reading comprehension (Hoover & Tunmer, 2020b). In order to explain the developmental changes observed in reading comprehension, a new model based on SVR, named Cognitive Foundations of Reading, was recently proposed by Hoover and Tunmer (2020d). The model is based on two central assumptions: (1) some mastery of lower-level skills is necessary so that higher-level skills can be learned; (2) many of the skills and processes involved in reading occur simultaneously, facilitating each other (Hoover & Tunmer, 2020a). In contrast to the original SVR, this model considers the existence of reciprocal relations between decoding/listening comprehension and reading comprehension. The results of previous studies provide empirical evidence for the existence of reciprocal relations between the components of the SVR and reading comprehension, but also between oral reading fluency and reading comprehension, in a wide range of orthographies and stages of reading development (Berninger & Abbott, 2010; Kim, 2015; Klauda & Guthrie, 2008; Little et al., 2017; Verhoeven & van Leeuwe, 2008,

2012). Studies conducted in Portugal have focused mainly on the early grade levels (Santos et al., 2020). The relative contribution of decoding, oral reading fluency, and listening comprehension to reading comprehension seems to be influenced not only by the stage of reading development, but also by the orthographic depth (Florit & Cain, 2011; Little et al., 2017; Padeliadu & Antoniou, 2014; Ribeiro et al., 2016; Santos et al., 2020). Additionally, as students advance across education system, they are expected to be able to read and understand texts with increasingly higher levels of difficulty (Clemens et al., 2017; Sanacore & Palumbo, 2008). In the specific case of the Portuguese educational system, the transition from the 1st to the 2nd cycle of basic education (i.e., the transition from fourth grade to fifth grade) may be a critical moment in the students' school life due to the pedagogical changes associated with this transition. From fifth grade on, students are challenged to acquire and apply knowledge in multiple domains (e.g., history and geography of Portugal) and are regularly confronted with significantly more complex texts (e.g., narrative texts but also poetic and expository texts). At the same time, the focus on formal instruction and training of reading finishes after the first four grade levels. Given these circumstances, a percentage of readers may begin to experience significant difficulties (Catts et al., 2012; Etmanskie et al., 2016). At this stage of learning, it becomes more evident that in addition to reading fluently and performing at high levels in terms of listening comprehension, students are required to have sophisticated inference skills, think critically about what they read, possess a rich vocabulary, and use reading strategies efficiently (Duke & Cartwright, 2021; Spinillo & Hodges, 2012). Previous research findings suggest the existence of reciprocal relations between vocabulary and reading comprehension (Seigneuric & Ehrlich, 2005; Verhoeven et al., 2011; Verhoeven & van Leeuwe, 2008). In contrast, studies focused on analyzing the reciprocal relations between reading strategies and reading comprehension are still scarce (Muijselaar et al., 2017).

Thesis Aims

The general aim of this thesis is to analyze the developmental and reciprocal relations between reading comprehension, listening comprehension, oral reading fluency, vocabulary and metacognitive skills across grades 4 to 6. This required the prior construction and validation of standardized measures for the assessment of listening and reading comprehension, as well as oral reading fluency. Specifically, the aims of the present thesis are:

(1) the construction and validation of the Test of Listening Comprehension and the Test of Reading Comprehension, using narrative texts, for grades 5 and 6 (TLC-$n_{5/6}$ and TRC-$n_{5/6}$, respectively),

following up the work initiated with the BAL that already includes tests from grades 1 to 4 (Ribeiro et al., 2014);

(2) the construction and validation of the Test of Fluency Assessment [TAF – *Teste de Avaliação da Fluência*] for grades 5 and 6 ($TAF_{5/6}$), continuing the work developed in the scope of the construction of a test to assess oral reading fluency for students in the 1st cycle of basic education (Freitas, 2020);

(3) the analysis of the reciprocal relations between reading comprehension, listening comprehension, oral reading fluency, vocabulary and metacognitive skills in a sample of students followed from grades 4 to 6.

Thesis Structure

The present thesis is organized into three chapters, corresponding to each of the aforementioned aims. Each of the chapters is published in or submitted to peer-reviewed international journals.

The first chapter includes two studies related to the construction and collection of validity evidence for the Test of Listening Comprehension of Narrative Texts and the Test of Reading Comprehension of Narrative Texts for students in the 2nd cycle of basic education (TLC-$n_{5/6}$ and TRC-$n_{5/6}$, respectively). Following the previous work of development of the BAL (Ribeiro et al., 2014), the construction of the test forms for the 2nd cycle of basic education was carried out using also the Rasch model. In the first study, the results concerning the psychometric qualities of each test form, such as unidimensionality, local independence of the items and reliability are reported. In the second study, the validity evidence based on the relations with other variables is presented. The results obtained in both studies are discussed considering empirical findings from previous studies, the limitations of the study and the implications for practice and future research.

In the second chapter, the development and validation process of a test to assess oral reading fluency in students from grades 1 to 6 ($TAF_{1\text{-}6}$) is presented. This is part of a larger project on reading fluency assessment across grades 1–6, and the construction of the test material for students in grades 1–4 was previously developed in the scope of another research project (Freitas, 2020). The test forms for fifth and sixth graders were constructed within the scope of the present doctoral thesis, and the equating procedures here presented were applied to the data collected for the six grade levels. This chapter includes two studies. The first study focuses on the process of horizontal and vertical equating of the test forms of the TAF. The second study was centered on collecting evidence of validity based on the relations with other variables, obtained through correlations between scores on the TAF and scores

on measures of decoding and reading comprehension, other measure of oral reading fluency, as well as students' academic achievement and teachers' ratings. Test-retest reliability was also analyzed through the correlations between the results obtained in the test forms of the TAF in two timepoints separated by two weeks. The results and the limitations of the studies are analyzed and discussed and the main contributions to the practices of test development and reading assessment in the Portuguese school context are explored.

The third and last chapter aims to clarify the directionality of the relations between reading comprehension and other four components, namely listening comprehension, oral reading fluency, vocabulary, and metacognitive skills across grades 4, 5 and 6. A longitudinal cross-lagged design was used. The results are presented and discussed taking into account previous theory and research, the limitations of the study as well as its theoretical and practical implications.

The results obtained in the three chapters of this thesis are integrated in a final reflection concerning the process of constructing tests to assess listening comprehension, reading comprehension and oral reading fluency, and the longitudinal interrelations between reading comprehension, listening comprehension, oral reading fluency, vocabulary, and metacognitive skills from grades 4 to 6. Guidelines for future research are also pointed out. Finally, we sought to explore the social impact of this thesis, particularly with regard to its contributions both to reading research and psychoeducational practice in the fields of assessment and intervention in reading. Some final considerations are also be presented.

References

Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M., & Kulesz, P. (2016). Validation of the direct and inferential mediation (DIME) model of reading comprehension in grades 7 through 12. *Contemporary Educational Psychology*, *44*, 68–82. https://doi.org/10.1016/j.cedpsych.2016.02.002

Albano, A. D., & Rodriguez, M. C. (2012). Statistical equating with measures of oral reading fluency. *Journal of School Psychology*, *50*(1), 43–59. https://doi.org/10.1016/j.jsp.2011.07.002

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014a). Reliability/precision and errors of measurement. In *Standards for educational and psychological testing* (pp. 33–48). American Educational Research Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014b). Validity. In *Standards for educational and psychological testing* (pp. 11–31). American Educational Research Association.

Arnesen, A., Braeken, J., Baker, S. K., Meek-Hansen, W., Ogden, T., & Melby-Lervåg, M. (2017). Growth in oral reading fluency in a semitransparent orthography: Concurrent and predictive relations with reading proficiency in Norwegian, grades 2–5. *Reading Research Quarterly*, *52*(2), 177–201. https://doi.org/10.1002/rrq.159

Asadi, I. A., & Ibrahim, R. (2018). The simple view of reading model in the transparent and deep versions of Arabic orthography. *Reading Psychology*, *39*(6), 537–552. https://doi.org/10.1080/02702711.2018.1481477

Begeny, J. C., & Greene, D. J. (2014). Can readability formulas be used to successfully gauge difficulty of reading materials? *Psychology in the Schools*, *51*(2), 198–215. https://doi.org/10.1002/pits.21740

Berninger, V. W., & Abbott, R. D. (2010). Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language systems in grades 1, 3, 5, and 7. *Journal of Educational Psychology*, *102*(3), 635–651. https://doi.org/10.1037/a0019319

Bonifacci, P., & Tobia, V. (2017). The simple view of reading in bilingual language-minority children acquiring a highly transparent second language. *Scientific Studies of Reading*, *21*(2), 109–119. https://doi.org/10.1080/10888438.2016.1261869

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*(3), 384–389. https://doi.org/10.7334/psicothema2012.288

Cadime, I., Rodrigues, B., Santos, S., Viana, F. L., Chaves-Sousa, S., Cosme, M. C., & Ribeiro, I. (2017). The role of word recognition, oral reading fluency and listening comprehension in the simple view of reading: A study in an intermediate depth orthography. *Reading and Writing*, *30*(3), 591–611. https://doi.org/10.1007/s11145-016-9691-3

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

Calet, N., Gutiérrez-Palma, N., & Defior, S. (2015). A cross-sectional study of fluency and reading comprehension in Spanish primary school children. *Journal of Research in Reading*, *38*(3), 272–285. https://doi.org/10.1111/1467-9817.12019

Campbell, S. (2005). Single instrument, multiple measures: Considering the use of multiple item formats to assess reading comprehension. In S. Paris & S. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 347–368). Lawrence Erlbaum Associates.

Caravolas, M., Lervåg, A., Mikulajová, M., Defior, S., Seidlová-Málková, G., & Hulme, C. (2019). A cross-linguistic, longitudinal study of the foundations of decoding and reading comprehension ability. *Scientific Studies of Reading*, *23*(5), 386–402.

https://doi.org/10.1080/10888438.2019.1580284

Carvalho, A. (2010). *Teste de avaliação da fluência e precisão da leitura: O Rei [Test of reading fluency and accuracy: The King]*. Edipsico.

Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology*, *104*(1), 166–181. https://doi.org/10.1037/a0025323

Cervetti, G. N., Pearson, P. D., Palincsar, A. S., Afflerbach, P., Kendeou, P., Biancarosa, G., Higgs, J., Fitzgerald, M. S., & Berman, A. I. (2020). How the reading for understanding initiative's research complicates the simple view of reading invoked in the science of reading. *Reading Research Quarterly*, *55*, S161–S172. https://doi.org/10.1002/rrq.343

Chaparro, E. A., Stoolmiller, M., Park, Y., Baker, S. K., Basaraba, D., Fien, H., & Smith, J. L. M. (2018). Evaluating passage and order effects of oral reading fluency passages in second grade: A partial replication. *Assessment for Effective Intervention*, *44*(1), 3–16. https://doi.org/10.1177/1534508417741128

Chaves-Sousa, S., Ribeiro, I., Viana, F. L., Vale, A. P., Santos, S., & Cadime, I. (2017). Validity evidence of the Test of Word Reading for Portuguese elementary students. *European Journal of Psychological Assessment*, *33*(6), 460–466. https://doi.org/10.1027/1015-5759/a000307

Chaves-Sousa, S., Santos, S., Viana, F. L., Vale, A. P., Cadime, I., Prieto, G., & Ribeiro, I. (2017). Development of a word reading test: Identifying students at-risk for reading problems. *Learning and Individual Differences*, *56*, 159–166. https://doi.org/10.1016/j.lindif.2016.11.008

Clemens, N. H., Simmons, D., Simmons, L. E., Wang, H., & Kwok, O. (2017). The prevalence of reading fluency and vocabulary difficulties among adolescents struggling with reading comprehension. *Journal of Psychoeducational Assessment*, *35*(8), 785–798. https://doi.org/10.1177/0734282916662120

Cutting, L. E., & Scarborough, H. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, *10*(3), 277–299. https://doi.org/10.1207/s1532799xssr1003_5

de Ayala, R. J. (2009). *The theory and practice of item response theory*. The Guilford Press.

Decree-Law no. 54/2018 of July 6 (2018). *Diário Da República: I Série, n.º 129*, 2918–2928. Retrieved July 8, 2021, from https://dre.pt/dre/detalhe/decreto-lei/54-2018-115652961

Duff, F. J., & Clarke, P. J. (2011). Practitioner review: Reading disorders: What are the effective interventions and how should they be implemented and evaluated? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *52*(1), 3–12. https://doi.org/10.1111/j.1469-7610.2010.02310.x

Duke, N. K., & Cartwright, K. B. (2021). The science of reading progresses: Communicating advances beyond the simple view of reading. *Reading Research Quarterly*, *56*, S25–S44. https://doi.org/10.1002/rrq.411

Eason, S. H., Goldberg, L. F., Young, K. M., Geist, M. C., & Cutting, L. E. (2012). Reader-text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. *Journal of Educational Psychology*, *104*(3), 515–528. https://doi.org/10.1037/a0027182

Equipa dos Estudos Internacionais (2017). *Resultados Globais PIRLS 2016, ePIRLS 2016 – Portugal. Literacia de leitura & literacia de leitura online [PIRLS 2016, ePIRLS 2016 global results – Portugal. Reading literacy & online reading literacy]*. Instituto de Avaliação Educativa, I.P. Retrieved January 12, 2022, from https://iave.pt/wp-content/uploads/2020/02/Relat_PIRLS_2016_Resultados.pdf

Etmanskie, J. M., Partanen, M., & Siegel, L. S. (2016). A longitudinal examination of the persistence of

late emerging reading disabilities. *Journal of Learning Disabilities*, *49*(1), 21–35. https://doi.org/10.1177/0022219414522706

European Comission (2012). *Final report. EU high level group of experts on literacy*. Publications Office of the European Union.

Fernandes, S., Querido, L., Verhaeghe, A., Marques, C., & Araújo, L. (2017). Reading development in European Portuguese: Relationships between oral reading fluency, vocabulary and reading comprehension. *Reading and Writing*, *30*(9), 1987–2007. https://doi.org/10.1007/s11145-017-9763-z

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. https://doi.org/10.1007/s10648-011-9175-6

Freitas, T. (2020). *Desenvolvimento da fluência de leitura e o seu papel mediador na relação entre variáveis cognitivo-linguísticas e a compreensão da leitura [Unpublished doctoral dissertation]*. Universidade of Minho.

Frid, B., & Friesen, D. C. (2020). Reading comprehension and strategy use in fourth- and fifth-grade French immersion students. *Reading and Writing*, *33*(5), 1213–1233. https://doi.org/10.1007/s11145-019-10004-5

Gomes-Koban, C., Simpson, I. C., Valle, A., & Defior, S. (2017). Oral vocabulary training program for Spanish third-graders with low socio-economic status: A randomized controlled trial. *PLoS ONE*, *12*(11), Article e0188157. https://doi.org/10.1371/journal.pone.0188157

Høien-Tengesdal, I. (2010). Is the simple view of reading too simple? *Scandinavian Journal of Educational Research*, *54*(5), 451–469. https://doi.org/10.1080/00313831.2010.508914

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

Hoover, W. A., & Tunmer, W. E. (2020a). Summary of the Cognitive Foundations Framework. In *The cognitive foundations of reading and its acquisition: A framework with applications connecting teaching and learning* (pp. 85–88). Springer.

Hoover, W. A., & Tunmer, W. E. (2020b). The cognitive foundations of reading. In *The cognitive foundations of reading and its acquisition: A framework with applications connecting teaching and learning* (pp. 23–40). Springer. https://doi.org/10.1007/978-3-030-44195-1

Hoover, W. A., & Tunmer, W. E. (2020c). The cognitive foundations of reading acquisition. In *The cognitive foundations of reading and its acquisition: A framework with applications connecting teaching and learning* (pp. 41–84). Springer. https://doi.org/10.1007/978-3-030-44195-1

Hoover, W. A., & Tunmer, W. E. (2020d). *The cognitive foundations of reading and its acquisition: A framework with applications connecting teaching and learning*. Springer. https://doi.org/10.1007/978-3-030-44195-1

Joshi, R. M., Ji, X. R., Breznitz, Z., Amiel, M., & Yulia, A. (2015). Validation of the simple view of reading in Hebrew—A semitic language. *Scientific Studies of Reading*, *19*(3), 243–252. https://doi.org/10.1080/10888438.2015.1010117

Kershaw, S., & Schatschneider, C. (2012). A latent variable approach to the simple view of reading. *Reading and Writing*, *25*(2), 433–464. https://doi.org/10.1007/s11145-010-9278-3

Kim, Y.-S. G. (2015). Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. *Child Development*, *86*(1), 128–144. https://doi.org/10.1111/cdev.12293

Kim, Y.-S. G. (2017). Why the simple view of reading is not simplistic: Unpacking component skills of reading using a direct and indirect effect model of reading (DIER). *Scientific Studies of Reading*, *21*(4), 310–333. https://doi.org/10.1080/10888438.2017.1291643

Kim, Y.-S. G., Quinn, J. M., & Petscher, Y. (2021). What is text reading fluency and is it a predictor or an outcome of reading comprehension? A longitudinal investigation. *Developmental Psychology*,

57(5), 718–732. https://doi.org/10.1037/dev0001167

Kim, Y.-S. G., & Wagner, R. K. (2015). Text (oral) reading fluency as a construct in reading development: An investigation of its mediating role for children from grades 1 to 4. *Scientific Studies of Reading*, *19*(3), 224–242. https://doi.org/10.1080/10888438.2015.1007375

Klauda, S. L., & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology*, *100*(2), 310–321. https://doi.org/10.1037/0022-0663.100.2.310

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). Springer. https://doi.org/10.1007/978-1-4939-0317-7

Köse, N., & Güneş, F. (2021). Undergraduate students' use of metacognitive strategies while reading and the relationship between strategy use and reading comprehension skills. *Journal of Education and Learning*, *10*(2), 99. https://doi.org/10.5539/jel.v10n2p99

Lee, K., & Chen, X. (2019). An emergent interaction between reading fluency and vocabulary in the prediction of reading comprehension among French immersion elementary students. *Reading and Writing*, *32*(7), 1657–1679. https://doi.org/10.1007/s11145-018-9920-z

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*, *89*(5), 1821–1838. https://doi.org/10.1111/cdev.12861

Liao, X., Zhu, X., & Zhao, P. (2021). The mediating effects of reading amount and strategy use in the relationship between intrinsic reading motivation and comprehension: Differences between grade 4 and grade 6 students. *Reading and Writing*. Advance online publication. https://doi.org/10.1007/s11145-021-10218-6

Little, C. W., Hart, S. A., Quinn, J. M., Tucker-Drob, E. M., Taylor, J., & Schatschneider, C. (2017). Exploring the co-development of reading fluency and reading comprehension: A twin study. *Child Development*, *88*(3), 934–945. https://doi.org/10.1111/cdev.12670

Lourenço, V., Duarte, A., Nunes, A., Amaral, A., Gonçalves, C., Mota, M., & Mendes, R. (2019). *PISA 2018 – Portugal. Relatório nacional [PISA 2018 – Portugal. National report]*. Instituto de Avaliação Educativa, I.P. Retrieved January 12, 2022, from http://www.pnl2027.gov.pt/np4/file/1205/RELATORIO_NACIONAL_PISA2018_IAVE.pdf

Magliano, J. P., Millis, K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 107–134). Erlbaum.

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). Evidence based on relations to other variables: Bolstering the empirical validity arguments for constructs. In *Instrument development in the affective domain: School and corporate applications* (3rd ed., pp. 209–248). Springer. https://doi.org/10.1007/978-1-4614-7135-6

McCoach, D. B., & Yu, H. (2016). Using individual growth curves to model reading fluency. In K. Cummings & Y. Petscher (Eds.), *The fluency construct: Curriculum-based measurement concepts and applications* (pp. 269–308). Springer. https://doi.org/10.1007/978-1-4939-2803-3

Megherbi, H., Seigneuric, A., & Ehrlich, M. F. (2006). Reading comprehension in French 1st and 2nd grade children: Contribution of decoding and language comprehension. *European Journal of Psychology of Education*, *21*(2), 135–147. https://doi.org/10.1007/BF03173573

Muijselaar, M. M. L., Swart, N. M., Steenbeek-planting, E. G., Droop, M., Verhoeven, L., & de Jong, P. F. (2017). Developmental relations between reading comprehension and reading strategies. *Scientific Studies of Reading*, *21*(3), 194–209. https://doi.org/10.1080/10888438.2017.1278763

Nakamoto, J., Lindsey, K. A., & Manis, F. R. (2008). A cross-linguistic investigation of English language

learners' reading comprehension in English and Spanish. *Scientific Studies of Reading*, *12*(4), 351–371. https://doi.org/10.1080/10888430802378526

Oakhill, J., Cain, K., & Elbro, C. (2019). Reading comprehension and reading comprehension difficulties. In D. A. Kilpatrick, R. M. Joshi, & R. K. Wagner (Eds.), *Reading development and difficulties: Bridging the gap between research and practice* (pp. 83–115). Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-26550-2_5

OECD (2019a). *PISA 2018 assessment and analytical framework*. OECD Publishing. https://doi.org/10.1787/b25efab8-en

OECD (2019b). *PISA 2018 results (volume I): What students know and can do*. OECD Publishing. https://doi.org/10.1787/5f07c754-en

Padeliadu, S., & Antoniou, F. (2014). The relationship between reading comprehension, decoding, and fluency in Greek: A cross-sectional study. *Reading & Writing Quarterly*, *30*(1), 1–31. https://doi.org/10.1080/10573569.2013.758932

Peng, P., Lee, K., Luo, J., Li, S., Joshi, R. M., & Tao, S. (2021). Simple view of reading in Chinese: A one-stage meta-analytic structural equation modeling. *Review of Educational Research*, *91*(1), 3–33. https://doi.org/10.3102/0034654320964198

Prieto, G., & Delgado, A. R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, *15*(1), 94–100.

Protopapas, A., Sideridis, G. D., Mouzaki, A., & Simos, P. G. (2007). Development of lexical mediation in the relation between reading comprehension and word reading skills in Greek. *Scientific Studies of Reading*, *11*(3), 165–197. https://doi.org/10.1080/10888430701344322

Resolution of the Council of Ministers no. 23/2016 of 11 April (2016). Programa Nacional de Promoção do Sucesso Escolar [National Program for the Promotion of School Success]. *Diário Da República: I Série, n.º 70*, 1195–1196. Retrieved January 11, 2022, from https://dre.pt/dre/detalhe/resolucao-conselho-ministros/23-2016-74094661

Resolution of the Council of Ministers no. 48-D/2017 of 31 March (2017). Linhas orientadoras para o Plano Nacional de Leitura 2027 [Guidelines for the National Reading Plan 2027]. *Diário Da República: I Série, n.º 65*, 1692-(5)-1692-(8). Retrieved January 11, 2022, from https://dre.pt/dre/detalhe/resolucao-conselho-ministros/48-d-2017-106816250

Ribeiro, I., Cadime, I., Freitas, T., & Viana, F. L. (2016). Beyond word recognition, fluency, and vocabulary: The influence of reasoning on reading comprehension. *Australian Journal of Psychology*, *68*(2), 107–115. https://doi.org/10.1111/ajpy.12095

Ribeiro, I., Viana, F. L., Santos, S., Cadime, I., Chaves-Sousa, S., Vale, A. P., & Spinillo, A. G. (2014). *BAL – Bateria de Avaliação da Leitura: Abordagens teóricas e opções metodológicas [BAL – Battery of Reading Assessment: Theoretical approaches and methodological options]*. CEGOC-TEA.

Salceda, J. C. R., Alonso, G. A., & Castilla-Earls, A. P. (2014). The simple view of reading in elementary school: A systematic review. *Revista de Logopedia, Foniatria y Audiologia*, *34*(1), 17–31. https://doi.org/10.1016/j.rlfa.2013.04.006

Sanacore, J., & Palumbo, A. (2008). Understanding the fourth-grade slump: Our point of view. *The Educational Forum*, *73*(1), 67–74. https://doi.org/10.1080/00131720802539648

Santi, K. L., Barr, C., Khalaf, S., & Francis, D. J. (2016). Different approaches to equating oral reading fluency passages. In K. Cummings & Y. Petscher (Eds.), *The fluency construct: Curriculum-based measurement concepts and applications* (pp. 223–265). Springer. https://doi.org/10.1007/978-1-4939-2803-3

Santos, S., Cadime, I., Viana, F. L., Chaves-Sousa, S., Gayo, E., Maia, J., & Ribeiro, I. (2017). Assessing reading comprehension with narrative and expository texts: Dimensionality and relationship with fluency, vocabulary and memory. *Scandinavian Journal of Psychology*, *58*(1), 1–8.

https://doi.org/10.1111/sjop.12335
Santos, S., Cadime, I., Viana, F. L., Prieto, G., Chaves-Sousa, S., Spinillo, A. G., & Ribeiro, I. (2016). An application of the Rasch model to reading comprehension measurement. *Psicologia: Reflexão e Crítica*, *29*(38), 1–19. https://doi.org/10.1186/s41155-016-0044-6
Santos, S., Cadime, I., Viana, F. L., & Ribeiro, I. (2020). Cross-lagged relations among linguistic skills in European Portuguese: A longitudinal study. *Reading Research Quarterly*, *55*(2), 177–192. https://doi.org/10.1002/rrq.261
Santos, S., Viana, F. L., Ribeiro, I., Prieto, G., Brandão, S., & Cadime, I. (2015). Development of listening comprehension tests with narrative and expository texts for Portuguese students. *Spanish Journal of Psychology*, *18*(e5), 1–7. https://doi.org/10.1017/sjp.2015.7
Schleicher, A. (2019). *PISA 2018: Insights and interpretations*. OECD Publishing.
Seigneuric, A., & Ehrlich, M. F. (2005). Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. *Reading and Writing*, *18*(7), 617–656. https://doi.org/10.1007/s11145-005-2038-0
Serrão, A., Simões, P., & Pires, R. (2021). *Estudo de aferição amostral do Ensino Básico 2021. Volume I – Resultados nacionais [Study of Basic Education 2021. Volume I – National results]*. Instituto de Avaliação Educativa, I.P. Retrieved December 28, 2021, from https://iave.pt/relatorios/
Silva, A. F., Marques, C., & Sucena, A. (2020). Validity evidence of the Reading Screening Test for Portuguese first graders. *Frontiers in Education*, *5*, Article ID 570639. https://doi.org/10.3389/feduc.2020.570639
Spencer, M., Richmond, M. C., & Cutting, L. E. (2020). Considering the role of executive function in reading comprehension: A structural equation modeling approach. *Scientific Studies of Reading*, *24*(3), 179–199. https://doi.org/10.1080/10888438.2019.1643868
Spinillo, A. G., & Hodges, L. V. S. D. (2012). Análise de erros e compreensão de textos: Comparações entre diferentes situações de leitura. *Psicologia: Teoria e Pesquisa*, *28*(4), 381–388. https://doi.org/10.1590/S0102-37722012000400006
Spinillo, A. G., Hodges, L. V. S. D., & Arruda, A. S. (2016). Reflexões teórico-metodológicas acerca da pesquisa em compreensão de textos com crianças. *Psicologia: Teoria e Pesquisa*, *32*(1), 45–51. https://doi.org/10.1590/0102-37722016011986045051
Sucena, A., & Castro, S. L. (2008). *Aprender a ler e avaliar a leitura. O TIL: Teste de Idade de Leitura [Learning to read and assess reading. The TIL: Reading Age Test]*. Almedina.
Sucena, A., & Castro, S. L. (2012). *ALEPE – Avaliação da leitura em Português Europeu [ALEPE – Reading assessment in European Portuguese]*. CEGOC-TEA.
Swart, N. M., Muijselaar, M. M. L., Steenbeek-Planting, E. G., Droop, M., de Jong, P. F., & Verhoeven, L. (2017). Differential lexical predictors of reading comprehension in fourth graders. *Reading and Writing*, *30*(3), 489–507. https://doi.org/10.1007/s11145-016-9686-0
Tobia, V., & Bonifacci, P. (2015). The simple view of reading in a transparent orthography: The stronger role of oral comprehension. *Reading and Writing*, *28*(7), 939–957. https://doi.org/10.1007/s11145-015-9556-1
Verhoeven, L., & van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, *22*(3), 407–423. https://doi.org/10.1002/acp
Verhoeven, L., & van Leeuwe, J. (2012). The simple view of second language reading throughout the primary grades. *Reading and Writing*, *25*(8), 1805–1818. https://doi.org/10.1007/s11145-011-9346-3
Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading*, *15*(1), 8–25. https://doi.org/10.1080/10888438.2011.536125

Verhoeven, L., Voeten, M., & Vermeer, A. (2019). Beyond the simple view of early first and second language reading: The impact of lexical quality. *Journal of Neurolinguistics*, *50*, 28–36. https://doi.org/10.1016/j.jneuroling.2018.03.002

Viana, F. L., Santos, S., Ribeiro, I., Chaves-Sousa, S., Brandão, S., Cadime, I., & Maia, J. (2015). Listening comprehension assessment: Validity studies of two vertically scaled tests for Portuguese students. *Universitas Psychologica*, *14*(1), 345–354. https://doi.org/10.11144/Javeriana.upsy14-1.lcav

CHAPTER 1

DEVELOPING AND VALIDATING TESTS OF READING AND LISTENING COMPREHENSION FOR FIFTH AND SIXTH GRADE STUDENTS IN PORTUGAL

Chapter 1

Developing and Validating Tests of Reading and Listening Comprehension for Fifth and Sixth Grade Students in Portugal[1]

Introduction

The product of listening and reading comprehension is an integrated mental representation of the meaning of a text (Oakhill et al., 2019). The processes necessary to extract meaning from written or oral language are generally similar: integration of information, making inferences, association of what one reads/hears with one's previous knowledge, and construction of the meaning of the material (Cain & Oakhill, 2008; Perfetti et al., 2005).

The assessment of these skills allows us to identify at-risk readers, to support the development of intervention and teaching programmes, and to monitor the students' progress in these areas over time (Santos et al., 2016). For this purpose, the use of standardized measures with robust psychometric qualities is essential (Salvia et al., 2010).

The overall aim of this paper is to describe the development and validation of two vertically scaled forms of a reading comprehension and a listening comprehension test for Portuguese students in the fifth and sixth grades.

Reading and listening comprehension tests developed to assess specific age or grade level groups are useful tools to compare inter-individual differences (i.e., to compare a student's performance with a normative group). However, when the goal is to compare the achievement of the same student across different time points (intra-individual differences), the administration of the same test at different educational levels has several disadvantages. The use of the same test across a wide range of academic grades is problematic due to the learning effects and reactivity effects of the measures. Moreover, the results can be influenced by extreme floor or ceiling effects in lower and upper grades, respectively (Santos et al., 2016). The solution for these problems is to use different and specific test forms for each academic grade with equated scores. Equating is a statistical process that allows the conversion of the scores obtained in different test forms into a single metric, so that these test forms can be used at different points of time and the scores can be directly compared to assess the development of these skills in the same individual over time (Kolen & Brennan, 2014).

---

[1] This work is published in the journal *Frontiers in Psychology.*

Rodrigues, B., Cadime, I., Viana, F. L., & Ribeiro, I. (2020). Developing and validating tests of reading and listening comprehension for fifth and sixth grade students in Portugal. *Frontiers in Psychology, 11,* Article ID 610876. https://doi.org/10.3389/fpsyg.2020.610876

Equating models based on item response theory analyses are widely used (Wilson & Moore, 2011). Item response theory analyses, including Rasch model analyses, allow researchers to assess "not only the difficulty level of a specific item, but also permit interval scaling for the assessment of change, assessment of the dimensionality of a set of items, and specification of the range of items (in terms of 'ability scores') that characterize a particular measurement device" (Francis et al., 2005, p. 375). Item response theory analyses also allow researchers to perform differential item functioning (DIF) analyses. Differential item functioning ensures equity in testing because identifying items that favor one group over another on a test prevents bias in the comparison of test scores between different groups (Walker & Beretvas, 2001). The development of standardized tests following item response theory, specifically based on Rasch model, has several advantages such as allowing the selection of the most appropriate items to the level of competence of the group that is intended to evaluate and performing the vertical equating of different versions of the same test (Kolen & Brennan, 2014; Prieto & Delgado, 2003).

Collecting empirical evidence of validity is also crucial for the development of tests. According to the American Educational Research Association et al. (2014), "validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11). Evidence based on relationships with other variables is one of the sources of validity evidence and refers to "the degree to which these relationships are consistent with the construct underlying the proposed test score interpretations" (American Educational Research Association et al., 2014, p. 16). It implies the identification of relevant variables for the construct to be measured and the analysis of the relationships between them. Reading comprehension requires the development of basic reading skills, such as oral reading fluency. A fluent reading ability is mandatory so that higher-level processes of reading comprehension can take place. Therefore, medium-to-large correlation coefficients between these skills have been found across a wide range of orthographies with varying depths, in students up to the sixth grade (Fernandes et al., 2017; Padeliadu & Antoniou, 2014; Yovanoff et al., 2005). However, as the automaticity of the basic reading processes increases across schooling years, successful text comprehension becomes more dependent on higher order skills, such as vocabulary, memory, reasoning, and comprehension monitoring (Fernandes et al., 2017; Nouwens et al., 2016; Ouellette & Beers, 2010; Ribeiro et al., 2016; Sesma et al., 2009; Yovanoff et al., 2005).

With regard to listening comprehension, given that it involves linguistic processes similar to the ones used in reading comprehension, similar results have been observed for the relationship between

listening comprehension, vocabulary, and working memory (Florit et al., 2014; Jiang & Farquharson, 2018; Kim, 2016; Ouellette & Beers, 2010; Tighe et al., 2015).

Analogical reasoning also seems to play an important role in solving comprehension tasks, since it enables processes for making inference (Tzuriel & George, 2009). In this sense, previous studies have shown that verbal and non-verbal reasoning had medium-to-large sized correlations with reading and listening comprehension in several orthographies (Potocki et al., 2017; Ribeiro et al., 2016; Tighe et al., 2015).

Moreover, readers who can successfully comprehend the text employ planning strategies (e.g., evaluate the text's difficulty before reading) to begin reading metacognitively, and monitoring strategies (e.g., summarize information in the text) to make sense of what they read (Botsas, 2017). However, empirical studies seem to yield mixed results when the use of reading strategies is assessed by self-report measures. For example, in a sample of Croatian students from the fifth to eighth grades, perceived use of reading strategies was significantly associated with reading comprehension only in eighth-grade students (Kolić-Vehovec & Bajšanski, 2006). However, in another study conducted with Chinese students, perceived use of reading strategies was also moderately correlated with reading comprehension among fifth graders (Law, 2009).

Finally, previous studies have also found medium-to-large correlation coefficients between teachers' ratings of students' reading skills and students' performance on standardized tests that assess reading and listening comprehension from kindergarten to the fifth grade (Feinberg & Shapiro, 2003, 2009; Gilmore & Vance, 2007; Santos et al., 2017; Viana et al., 2015).

The Present Study

Various measures of reading assessment for elementary school students have been developed in Portugal. One of these measures was the Battery of Reading Assessment (Santos et al., 2015, 2016), which is composed of vertically scaled forms to assess word reading, listening, and reading comprehension from the first to the fourth grade. The special attention paid to the lower grades of elementary schools for assessing reading and listening comprehension can be explained by the importance and the impact of learning across the primary school years on the subsequent years. However, results from national level reports in Portugal have shown that the number of children who have reading difficulties past lower grades of elementary school is still high (Monteiro et al., 2014). These data raise growing concerns about reading difficulties emerging in the later years of schooling: students who succeed in learning to read in the primary grades, but then fall behind in the upper

elementary or middle school grades (Catts et al., 2005; Leach et al., 2003; Lipka et al., 2006). This phenomenon imposes the need for the development of robust measures that not only allow further development of research in this field, but also help to assess and monitor comprehension performance beyond lower elementary school grades. Therefore, the present study intended to expand the Battery of Reading Assessment for fifth and sixth graders in Portugal.

This paper reports the procedure and results of two studies. The purpose of the first study was to develop listening and reading comprehension tests, with two vertically scaled test forms for European Portuguese students in the fifth and sixth grades, using Rasch model analyses. The second study aimed to collect validity evidence for the two vertically scaled forms of each test based on the relationship of test scores to other variables by analysing the relationships between the developed test forms and measures used as external criteria for oral reading fluency, vocabulary, working memory, comprehension monitoring, verbal and abstract reasoning, teachers' ratings, and academic achievement. Based on the research literature, it was expected that the scores on the test of reading comprehension will be positively correlated with all the other variables and that the scores on the test of listening comprehension will be correlated with measures of vocabulary, working memory, verbal and abstract reasoning, teachers' ratings, and academic achievement.

Study 1

Materials and Methods

Participants

All participants were native speakers of European Portuguese, attending schools located in northern Portugal. The sample included 222 fifth graders (M_{age} = 10.95 years, SD = 0.58; 52.3% were boys; 77% were attending public schools) and 232 sixth graders (M_{age} = 11.98 years, SD = 0.42; 52.6% were boys; 89.2% were attending public schools). Students who qualified for educational intervention at the selective and/or additional levels were not included in the sample. With regard to socioeconomic status, 43.7% of the fifth graders and 26.7% of the sixth graders benefited from scholar social support (i.e., reduced-price meals at school, access to a loan service for books, and support for the acquisition of school supplies). Regarding maternal education of the fifth graders, 16.2% of the mothers had completed a university degree, 28.8% had completed high school, and 55% had a lower educational degree. In relation to the sample of sixth graders, 30.2% of their mothers had completed a university degree, 27.6% had completed high school, and 37.9% had a lower educational degree (4.3% missing information).

Study Design and Measures

Non-equivalent groups with anchor test design is the most appropriate equating procedure in the construction of measures involving at least two groups that differ in ability level, responding to different test forms (de Ayala, 2009; Kolen & Brennan, 2014). For this purpose, test forms should include a set of common items between adjacent grades, and a set of unique items for each test form, that allows researchers to calibrate each test form separately as well as sequentially using vertical equating (Kolen & Brennan, 2014). This kind of equating is used when groups of subjects differ in ability level and tests differ in level of difficulty (Baker, 1984). This technique is used when the goal is to compare performances in skills that are expected to develop over time, such as listening and reading comprehension (de Ayala, 2009). Specific test forms for each grade were developed to assess reading and listening comprehension at the end of fifth and sixth grades, namely, the Test of Reading Comprehension of Narrative Texts (TRC-$n_{5/6}$) and the Test of Listening Comprehension of Narrative Texts (TLC-$n_{5/6}$). Each test form included a booklet with three original texts (two unique texts for each grade, and one common text between adjacent grades) authored by Portuguese writers of literature for children and a worksheet containing the test items. The length of the texts for TRC-n_5 and TLC-n_5 ranged from 551 to 882 words, and for TRC-n_6 and TLC-n_6 from 574 to 700 words. Each test form comprised unique and common items between the test forms for the adjacent grades (see Table 1). Regarding TRC-n_5 and TLC-n_5, anchor items (and the respective text) were derived from the test forms previously validated for fourth graders (Santos et al., 2015, 2016). To select the anchor items to be included in the test forms for the fifth grade, the mean difficulty level of the items of the texts that composed the fourth-grade test forms was computed. The text whose items had the highest mean difficulty was selected and the respective items were used as anchor items. Anchor items (and the respective text) of the TRC-n_6 and the TLC-n_6 were derived from the TRC-n_5 and the TLC-n_5, respectively. Test items were multiple-choice questions with three options. Prior studies have shown that three options are optimal for multiple-choice items, being as psychometrically efficient as four or five options (Delgado & Prieto, 1998; Rodriguez, 2005). Items were developed to assess four levels of comprehension (literal, inferential, critical, and reorganization) described in the taxonomy by Català et al. (2001), that was used in the development of the test forms for primary school students (Santos et al., 2015, 2016). The items and the options were developed by the researchers and were later revised by text comprehension experts, who have extensive experience in teachers' training. In the TRC-n, the student silently reads the text passages that are followed by multiple-choice questions and marks the chosen option on the answer sheet (pencil-and-paper format). In the TLC-n, the student listens to the texts divided in short

passages and the respective items that are only presented orally though an audiotaped recording and marks the chosen option on the answer sheet. The testing procedure included two example items for all test forms. There was no time limit to complete each test. In the TLC-$n_{5/6}$, the students listened to the texts divided in short passages and the respective items that are only presented orally though an audiotaped recording. The audiotaped recording was stopped after each item so that all students had time to mark their response. The presentation of the next item proceeded only after all students have marked the chosen option on their answer sheet or decided not to respond.

Table 1

Items in Each Test Form of the TRC-n and TLC-n

Test form	Initial pool of items							Final pool of items						
	LC	IC	CC	R	U	A	T	LC	IC	CC	R	U	A	T
TRC-n_5	7	21	3	8	32	7	39	6	15	3	8	26	6	32
TRC-n_6	13	22	3	8	32	14	46	7	14	2	7	22	8	30
TLC-n_5	12	25	3	8	35	13	48	10	16	2	7	27	8	35
TLC-n_6	11	19	3	6	27	12	39	7	15	3	4	23	6	29

Note. LC = number of items of Literal Comprehension, IC = number of items of Inferential Comprehension, CC = number of items of Critical Comprehension, R = number of items of Reorganization, U = number of Unique items, A = number of Anchor items, T = Total of items.

Procedure

Legal authorizations for data collection were obtained from the ethics committee of the University of Minho and the Portuguese Ministry of Education, and from the respective school boards. Informed consent for the test administration was previously collected from parents or legal tutors. The anonymity and confidentiality of the data were guaranteed. Children were informed of the goals and characteristics of the study and were told that they could drop the study at any time. All tests were administered collectively in the classroom by trained psychologists.

Data Analyses

Ten cases had missing data in the TRL-$n_{6,}$ but the number of missing values was only 0.14% of the total data. Five outliers for TRC-n_6 were found in the exploratory data analyses and were, therefore, removed. Unidimensionality of the test forms was tested using principal component analyses (PCA) of

the linearized Rasch residuals. Eigenvalues less than 2.0 and/or explained variance less than 5% for secondary dimensions support this requisite (Linacre, 2018). Correlations between the items' linearized Rasch residuals were calculated to examine the assumption of local independence of the items. Correlations higher than .70 may indicate that response to an item does not exclusively depend on the persons' ability and is influenced by the performance on another item (Linacre, 2018). The reliability was analysed by calculating Rasch coefficients for the person- and item-separation reliability (PSR and ISR), as well as Kuder-Richardson formula 20 (KR20). All coefficients values should be higher than .70 (Nunnally & Bernstein, 1994). Infit and outfit mean square statistics were analysed to assess person and item fit to the model. These values should be smaller than 1.5 (Linacre, 2002). Mean ability for the students who selected each answer option was also computed for each item of the test forms. It was expected that the students with highest ability levels would choose the correct options (Linacre, 2018). Differential item functioning analysis was conducted to test the invariance of measurement as a function of sex and socioeconomic status for all items of each test form, using the Rasch-Welch procedure and considering a significance level of 5% (Linacre, 2018). Besides statistical significance, DIF size was also considered for practical significance: it was considered notable if the DIF contrast was $\geq |1.0|$ logit (Wright & Douglas, 1975, 1976). The displacement of the anchor items was also analyzed in order to evaluate the stability of the common items' difficulty between adjacent grades. Values of the anchor items' displacement can assume values as large as 1.0 logits without causing much impact on measurement (Linacre, 2018). The literature also suggests a minimum of 20% of anchor items in tests with 40 or more items for equating purposes (Kolen & Brennan, 2014). Both criteria were taken into account in the decision of deleting anchor items.

The TRC-n and TLC-n forms were linked according to three steps. In a first step, the calibration of the versions for the fifth grade was performed fixing the parameters of the common items with the values obtained in the versions for the fourth grade. In a second step, the items with inappropriate psychometric characteristics were removed. The psychometric characteristics considered were: item misfit, point-measure correlation (correlation between the response to the item and the construct that is being measured by the set of items) lower than .15, correct option not chosen by participants with higher levels of the latent trait, presence of DIF as a function of sex and of socioeconomic status, and, in the case of anchor items, displacement higher than 1.0. In a third step, the number of unique items were reduced taken into account the same criteria adopted in the development of the test forms for the primary school students (Santos et al., 2015, 2016): the spread of difficulty (the items distributed along the continuum of ability of each grade sample were chosen), the redundancy (the number of items of

similar difficulty levels was reduced by discarding some redundant items), and the comprehension level (the proportion of items of each comprehension level in the initial pool of items was maintained in the final pool of items, and when two or more items were of similar difficulty levels and they measured the same comprehension level, the item with a higher point-measure correlation was selected). The same steps were followed for the versions of the sixth grade, with the scores of the anchor items obtained in the fifth grade being used in the first step. The test forms were again linked through a final set of calibrations using the unique and anchor items selected in the previous steps. Finally, the reliability coefficients were recalculated. All these analyses were conducted using the Rasch measurement software Winsteps 3.92.1 (Linacre, 2018). Descriptive statistics and one-way analysis of variance to test differences between the grades in the scaled scores obtained on each test form were performed through the statistical program IBM® SPSS Statistics 26. Statistically significant differences were expected in the means between the three grade levels, with successively higher values in upper grade levels.

Results

Dimensionality and Local Independence of the Items

Results of the PCA of the residuals revealed that all the secondary dimensions had eigenvalues close to 2.0 for the initial forms and these secondary dimensions explained less than 5% of the variance (see Table 2). The explained variance by measures was about four times higher than the variance explained by the first secondary dimension. The residuals' correlations were much lower than .70. Therefore, these results support the assumptions of unidimensionality and local independence of the items for implementing Rasch model analyses.

Table 2

PCA of the Residuals and Reliability Coefficients by Test Form

Test form	Highest eigenvalue of the SDi (EV)	RV by measures	Highest correlation of the IR	PSR	KR20	ISR
Initial pool of items						
TRC-n_5	2.17 (4.4%)	21.6%	.25	.77	.78	.97
TRC-n_6	2.33 (4.0%)	21.0%	.34	.78	.80	.97
TLC-n_5	2.10 (3.6%)	17.8%	.26	.73	.73	.97
TLC-n_6	2.29 (4.6%)	21.2%	.23	.74	.76	.97
Final pool of items						
TRC-n_5	1.84 (4.7%)	18.7%	.22	.75	.76	.95
TRC-n_6	1.92 (4.9%)	24.1%	.18	.72	.76	.97
TLC-n_5	1.90 (4.5%)	17.1%	.22	.70	.72	.95
TLC-n_6	2.03 (5.4%)	22.7%	.21	.70	.72	.97

Note. SDi = Secondary Dimensions, EV = Variance Explained by the secondary dimensions, RV = Raw Variance explained by measures, IR = Items' Residuals, PSR = Person Separation Reliability, KR20 = Kuder-Richardson 20, ISR = Item Separation Reliability.

Item Analyses

Table 3 presents descriptive statistics for the Rasch estimated parameters for each test form.

Table 3

Descriptive Statistics of the Estimated Parameters by Test Form

Test form	Person ability				Item difficulty				Item infit				Item outfit				% person misfit	
	M	*SD*	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*	Min	Max	*M*	*SD*	Min	Max	Infit>1.5 (%)	Outfit>1.5 (%)
Initial pool of items																		
TRC-n_5	1.17	0.78	–0.85	3.19	0.71	0.89	–0.81	3.56	1.00	0.07	0.89	1.14	1.00	0.12	0.83	1.47	0.5	4.5
TRC-n_6	2.03	0.85	–0.39	5.13	0.84	0.99	–1.92	3.74	1.01	0.14	0.71	1.65	0.97	0.26	0.49	2.14	0	4.3
TLC-n_5	1.86	0.66	0.17	3.56	1.10	0.89	–0.74	3.91	1.02	0.14	0.66	1.51	1.03	0.17	0.64	1.60	0	2.3
TLC-n_6	2.46	0.82	0.02	4.64	1.31	0.99	–0.64	3.27	1.02	0.10	0.88	1.27	0.99	0.15	0.66	1.27	0.9	5.2
Final pool of items																		
TRC-n_5	1.17	0.83	–1.20	3.55	0.64	0.69	–0.61	1.85	1.00	0.08	0.89	1.16	0.99	0.11	0.83	1.27	0	0.5
TRC-n_6	2.08	0.98	–0.53	4.67	0.70	1.09	–1.98	3.96	1.01	0.12	0.71	1.37	0.97	0.24	0.55	1.49	0	7.0
TLC-n_5	2.03	0.78	0.26	4.05	0.99	0.75	–0.64	3.04	1.01	0.10	0.68	1.40	1.01	0.13	0.66	1.43	0	2.3
TLC-n_6	2.50	0.90	–0.49	5.10	1.35	0.95	–0.30	3.33	1.00	0.09	0.89	1.28	0.97	0.15	0.68	1.24	0.7	6.5

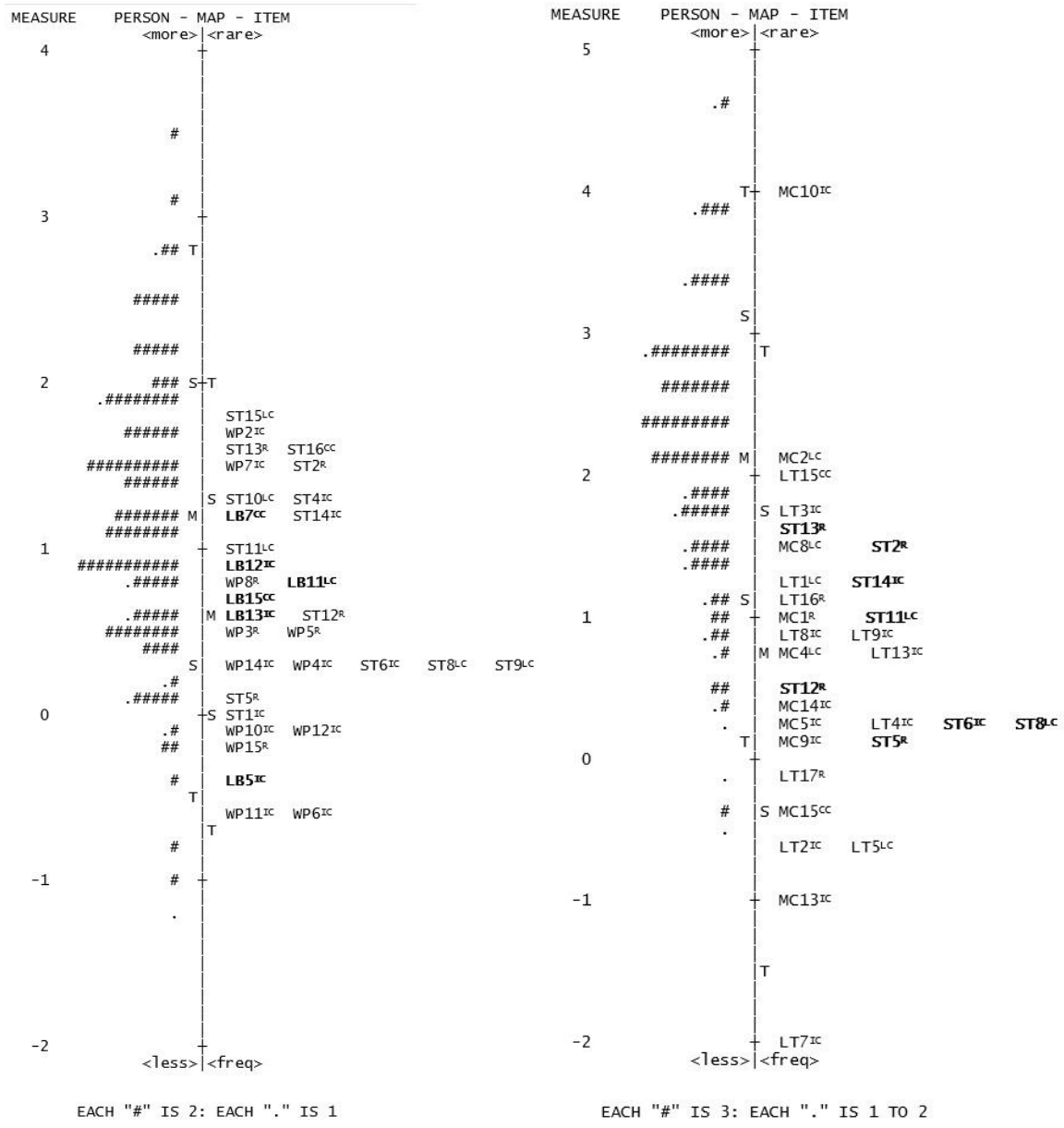
Note. M = Mean, *SD* = Standard Deviation, Min = Minimum, Max = Maximum.

Tests of Reading Comprehension. In the TRC-n_5 none of the items exceeded the reference value of 1.5 for infit and outfit statistics (see Table 3) and the highest mean ability value was obtained by students who chose the correct answer option for all 39 items. However, one item exhibited a difficulty value higher than the maximum value for person ability, meaning that it was too difficult for fifth graders. The same item presented a point-measure correlation lower than .15. Moreover, four items were flagged with DIF as a function of sex and two items were flagged with DIF as a function of socioeconomic status. Two of these six items were anchor items. Consequently, only one out of these two (the one with the highest DIF contrast) was eliminated in order to maintain the percentage of anchor items close to the minimum value of 20%. The item that was maintained in the measure obtained a DIF contrast of 0.65 and, therefore, its impact was considered not notable. In addition to the six items with inappropriate psychometric characteristics mentioned above, one more item was deleted according to the criteria for selection of unique items. Therefore, seven items were removed from the initial version of TRC-n_5. Thus, the final version of TRC-n_5 was composed of 32 items with six anchor items (18.8% of the total number of items).

In the TRC-n_6 initial pool of 46 items, four items presented difficulty levels lower than the minimum person ability value, meaning that they were too easy for fifth graders. Further, one item had infit and outfit values higher than 1.5 and a negative point-measure correlation. In the same item, students who chose the correct answer option were not the ones with the highest average ability levels. A second item had a point-measure correlation lower than .15. This one and more two items were also flagged with DIF as a function of sex. Additionally, four items were flagged with DIF as a function of socioeconomic status. Therefore, eight items were removed from TRC-n_6. According to the criteria for selection of unique items, eight other items were also removed. Therefore, the final version of TRC-n_6 was composed of 30 items with eight anchor items (26.7% of the total number of items). Figure 1 presents the item and person parameter locations in the vertical scale resulting from the final recalibration of the TRC-n_5 (left) and the TRC-n_6 (right). Mean values of the person ability standardized scores for the TRC-n were 111 (SD = 10) for the TRC-n_5, and 120 (SD = 10) for the TRC-n_6. In the validation study of the version for the fourth grade (TRC-n_4, Santos et al., 2016), the mean was 108 (SD = 10). With the progress in grade levels (lower to higher), person ability values were significantly greater, $F(2, 670) = 90.874$, $p < .001$. Post-hoc tests revealed significant differences ($p < .001$) between the scaled scores obtained on the three TRC-n test forms.

Figure 1

Person-item Variable Map for TRC-n$_5$ (left) and TRC-n$_6$ (right)



Note. Items are identified by the text to which they are related (WP = The history of white pencil, ST = A very special trip, LB = The lost bread, MC = A mysterious chest, LT = Loose thoughts), followed by the item's number. The comprehension level assessed by each item is presented in superscript (LC = Literal Comprehension, IC = Inferential Comprehension, R = Reorganization, CC = Critical Comprehension). Anchor items appear in bold.

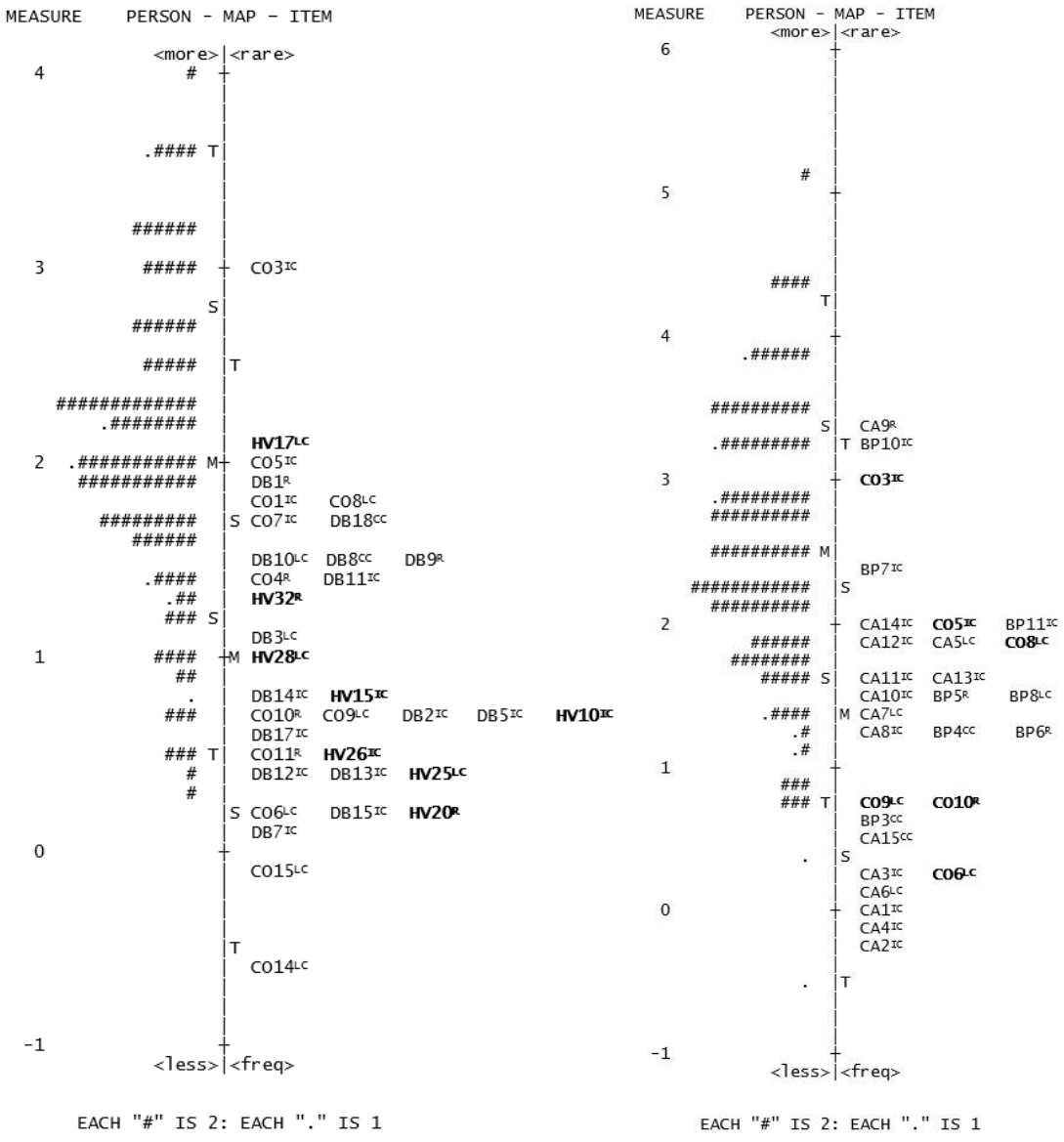
Tests of Listening Comprehension. Of the TLC-n$_5$ initial pool of 48 items, six items exhibited difficulty values lower than the minimum value for person ability (see Table 3), meaning that they were very easy for fifth graders. Additionally, one item exhibited difficulty value higher than the

maximum value for person ability, indicating that it was very difficult for fifth graders. Regarding the infit and outfit statistics, two items exceeded the reference value of 1.5. Three items had point-measure correlations lower than .15. In one of these three items, the highest mean ability value was not obtained by students who chose the correct answer option, suggesting that the students with greater reading comprehension abilities chose an incorrect alternative. Additionally, four items were removed because they were flagged as having DIF regarding sex or socioeconomic status. According to the criteria for the selection of unique items, four additional items were also removed. Therefore, a total of 13 items were removed. Thus, the final version of the TLC-n_5 was composed of 35 items with eight anchor items (22.9% of the total number of items).

In the TLC-n_6 initial pool of 39 items, the minimum person ability value exceeded the minimum value of items difficulty for four items, meaning that these items were easy for sixth graders. Regarding the fit statistics for the items, none of the items exceed the reference value of 1.5. Only one item had point-measure correlation lower than .15. Two items were flagged as having DIF both as a function of sex and socioeconomic status and, therefore, were eliminated. Additionally, four items were flagged as having DIF only as a function of sex and six as having DIF only as a function of socioeconomic status. Only seven out of these 10 items were removed in order to maintain acceptable reliability coefficients (PSR, KR20 and ISR). The three with the lowest DIF contrast, ranging between 0.61 and 0.78, were maintained in the test form. As a summary, 10 items were removed and the final version of the TLC-n_6 was composed of 29 items, among which six were anchor items (20.69% of the total number of items). Figure 2 presents the item and person parameter locations in the vertical scale resulting from the final recalibration of the TLC-n_5 (left) and the TLC-n_6 (right). Mean values of the person ability standardized scores for the TLC-n were 124 (SD = 10) for the TLC-n_5, and 128 (SD = 10) for the TLC-n_6. In the validation study of the version for the fourth grade (TLC-n_4; Santos et al., 2015) the mean was 122 (SD = 10). With the progress in grade levels (lower to higher), person ability values were significantly greater, $F(2, 711) = 24.721$, $p < .001$. Post-hoc tests revealed significant differences ($p < .05$) between the scaled scores obtained on the three TLC-n test forms.

Figure 2

Person-item Variable Map for TLC-n$_5$ (left) and TLC-n$_6$ (right)



Note. Items are identified by the text to which they are related (DB = The Dentuça Bigodaça, CO = A composition, HV = Holidays in the Village, CA = The camping, BP = The bird and the pine), followed by the item's number. The comprehension level assessed by each item is presented in superscript (LC = Literal Comprehension, IC = Inferential Comprehension, R = Reorganization, CC = Critical Comprehension). Anchor items appear in bold.

Reliability

The PSR and KR20 values were moderate, and the ISR coefficients were very high for the initial and final versions of all test forms. The elimination of items from the initial test forms to the final test forms did not cause a sharp decrease in reliability (see Table 2).

Study 2

Method

Participants

A group of 179 students participated in the study of validity evidence of the TRC-n and the TLC-n forms: 94 were fifth graders (M_{age} = 10.96 years, SD = 0.55; 52.1% were girls) and 85 were sixth graders (M_{age} = 12.01 years, SD = 0.50; 52.9% were boys). All students attended public schools located in northern Portugal and were native speakers of European Portuguese. Students with special educational needs who were identified for selective and/or additional intervention were not included in the sample.

Measures

TRC-n$_{5/6}$ and TLC-n$_{5/6}$

In this study, the final versions of these scales as developed in Study 1, were used.

Test of Reading Fluency

This test assesses oral reading fluency in students from the first to sixth grade, with an unpublished text composed of 1,160 words. Students were asked to read the text aloud within 3 min. Word omissions, substitutions, and mispronunciations were scored as errors. Self-corrections within 3 s after the error, repeated words, mispronunciations due to dialect or regional differences, hesitations, or words read slowly but correctly were not scored as errors. The number of words read correctly per minute was calculated using the mean of the 3 min.

Vocabulary Subtest of the Wechsler Intelligence Scale for Children–III (WISC–III; Wechsler, 2003)

It has 30 items composed of words presented orally which the students were asked to define orally, as completely as possible. Each item can be scored with zero, one, or two points, depending on the quality of the response, and the raw scores are converted to standardized scores. The administration is interrupted after four consecutive failures. High reliability and good indicators of validity were found in the Portuguese version of the WISC–III (Simões & Albuquerque, 2002).

Digit Span Subtest of the WISC–III (Wechsler, 2003)

This subtest includes two tests of digit span (forward and backward). Each series consists of two rehearsals and the test is interrupted after both rehearsals of the same series fail. Items are scored with zero or one point, and the raw scores were converted to standardized scores. Reliability coefficients for the Portuguese Digit Span subtest ranged between .71 and .90 (Simões & Albuquerque, 2002).

Reading Strategy Use (Ribeiro et al., 2015)

This scale is composed of 22 items that assess cognitive and metacognitive strategy use (10 and 12 items, respectively). Each item consists of a proposition that represents a reading strategy, and the student's task is to mark the frequency of its use on a 7-point Likert scale ranging from 1 (*never*) to 7 (*always*). The adaptation and validation studies of this scale for the Portuguese population supported the one-dimensional structure assumption. Cronbach's alpha coefficient was .85 for the 22 items (Ribeiro et al., 2015).

Verbal Reasoning Subtest (Almeida & Lemos, 2006)

This subtest assesses the ability to infer and apply relationships using tasks of verbal content. It consists of 20 multiple-choice items with four options (one correct), involving analogies between words. The test administration has a time limit of 4 min. Items are scored with zero or one point. Reliability coefficient for this test was .72 (Almeida & Lemos, 2006).

Abstract Reasoning Subtest (Almeida & Lemos, 2006)

It is composed of 20 items representing analogies with geometrical figures that have to be answered within a 5-min time limit. Items are multiple-choice questions with four options (one correct). Items are scored with zero or one point. Reliability coefficient for this test was .71, and statistically significant correlation coefficients were obtained with school achievement in subjects such as "Portuguese" and "mathematics" (Almeida & Lemos, 2006).

Teacher Ratings of Students' Reading Skills

Teachers were asked to rate students' performance in oral reading fluency, listening, and reading comprehension in the classroom, using a scale ranging from 1 (*very poor*) to 5 (*excellent*).

Academic Results

Two scores of academic achievement were collected: the classification in the subject "Portuguese" and the grade point average obtained at the end of the academic year.

Procedure

Similar procedures to the ones used in Study 1 were employed. The TRC-$n_{5/6}$, TLC-$n_{5/6}$, both subtests of reasoning and the strategy use scale were administered collectively, while the remaining tests were administered individually.

Data Analyses

Descriptive statistics and Pearson and Spearman correlation coefficients between all variables were calculated using IBM® SPSS Statistics 26. The size of the correlations was evaluated according to the following criteria: .10, .30, and .50 indicates a small, a medium, and a large effect, respectively (Cohen, 1992).

Results

Table 4 provide the descriptive statistics and the correlations between the scores on the TRC-$n_{5/6}$ and TLC-$n_{5/6}$ and the other measures used as external criteria for the students' abilities.

High correlations were found between the TRC-n and TLC-n test forms in both grades. The correlations between the TRC-n forms and oral reading fluency assessed with the TRF, were moderate in the fifth grade and high in the sixth grade. The scores on the TRC-n were also moderately correlated with the use of reading strategies in the fifth grade, but not in the sixth. All of the TRC-n and TLC-n test forms were highly correlated with measures of vocabulary and with academic results of the students. The correlations of the TRC-n and the TLC-n with measures of working memory and abstract reasoning were moderate in the fifth grade and low in the sixth grade. Correlation coefficients with verbal reasoning were medium-to-large in the fifth grade, but small in the sixth grade. Moderate-to-high correlations with teacher ratings were found for all test forms in both grades.

Table 4

Descriptive Statistics and Correlations in the Fifth and the Sixth Grades

Tests	M_{grade5}	SD_{grade5}	M_{grade6}	SD_{grade6}	1	2	3	4	5	6	7	8	9	10	11	12	13
1.TRC-n	111.30	10.99	116.58	11.87	–	.57***	.54***	.77***	.26*	.16	.21	.20	.62***	.55***	.65****	.55***	.62***
2.TLC-n	124.14	10.63	124.67	9.63	.55***	–	.35***	.53***	.18	.10	.26*	.22*	.49***	.39***	.45***	.46***	.53***
3. TRF	126.84	23.38	137.40	25.60	.37***	.18	–	.48***	.24*	.12	.14	.13	.52***	.48***	.45***	.39***	.51***
4. V	9.88	3.13	9.70	2.98	.53***	.49***	.34***	–	.39***	.13	.16	.17	.61***	.55***	.64***	.61***	.65***
5. DS	9.53	3.03	8.80	2.39	.31**	.31**	.22*	.32**	–	–.04	.18	.13	.30**	.14	.16	.16	.19
6. RSU	102.28	21.60	99.18	18.18	.42***	.15	.34***	.14	.05	–	–.13	.00	.07	.12	.10	.10	.10
7. AR	12.70	3.09	13.51	2.78	.49***	.45***	.14	.38***	.34***	.11	–	.48***	.09	.05	.10	.06	.09
8. VR	12.59	3.36	13.80	3.12	.54***	.36***	.37***	.32**	.23*	.22*	.44***	–	.22*	.12	.17	.11	.17
9. GPA	3.66	0.70	3.82	0.73	.59***	.58***	.35***	.45***	.32**	.27**	.34***	.45***	–	.45***	.48***	.61***	.75***
10. ORF	3.44	0.90	3.56	0.92	.57***	.52***	.62***	.50***	.21*	.28**	.36***	.42***	.57***	–	.79***	.77***	.65***
11. LC	3.60	0.79	3.41	0.88	.58***	.45***	.59***	.55***	.25*	.17	.41***	.51***	.53***	.76***	–	.82***	.55***
12. RC	3.30	0.87	3.24	0.92	.64***	.58***	.54***	.51***	.23*	.22*	.37***	.45***	.51***	.72***	.78***	–	.53***
13. PC	3.42	0.77	3.53	0.67	.53***	.55***	.43***	.44***	.22*	.17	.35***	.47***	.74***	.61***	.57***	.71***	–

Note. TRC-n = Test of Reading Comprehension of Narrative Texts, TLC-n = Test of Listening Comprehension of Narrative Texts, TRF = Test of Reading Fluency, V = Vocabulary, DS = Digit Span, RSU = Reading Strategy Use, AR = Abstract Reasoning, VR = Verbal Reasoning, GPA = Grade Point Average, ORF = Oral Reading Fluency assessed by teachers, LC = Listening Comprehension assessed by teachers, RC = Reading Comprehension assessed by teachers, PC = Classification in the subject of "Portuguese" at the final of academic year. Correlation coefficients below the diagonal are for students in the fifth grade and above the diagonal are for students in the sixth grade. The first nine correlation coefficients for each grade are Pearson correlations; the remaining four correlation coefficients are Spearman correlations.

*** $p < .001$. ** $p < .01$. * $p < .05$.

Discussion

The aim of the first study was to develop vertically scaled test forms for listening and reading comprehension for Portuguese students in the fifth and sixth grades, through the application of Rasch model analyses. In the second study, evidence for the validity of these forms was collected, based on relationships of the test scores with other variables.

Regarding the first study, the selection of items for each form of the TRC-n and TLC-n took into account the items with misfit, point-measure correlations lower than .15, unexpected responses (i.e., the highest mean ability value was obtained by students who chose an incorrect answer option), and the presence of DIF. Items flagged with any of these problems were removed from the respective test forms. Anchor items with high displacement values were also eliminated from each test form. Unique items for each test form were selected by considering the same criteria used in the development of the TRC-n and TLC-n for primary school, that is, first to fourth grades, in previous studies (Santos et al., 2015, 2016). The TRC-n_5 and the TRC-n_6 were composed of 32 and 30 items, with six and eight anchor items, respectively. Regarding the TLC-n, the test form for the fifth grade included 35 items, 8 of which were anchor items, and the test form for sixth grade included 29 items, with 6 anchor items. Anchor items represented about 20% of the final pool of items in all test forms, as recommended by Kolen and Brennan (2014). All items from the TRC-n and TLC-n final forms revealed appropriate psychometric characteristics. In order to maintain this percentage of anchor items in the final versions, one item with socioeconomic status-related DIF was maintained in the TRC-n_5. Additionally, to maintain acceptable reliability, another three items with DIF (one with gender-related DIF and two with socioeconomic status-related DIF) were retained in the TLC-n_6. This number of items with DIF represents a low percentage of the test forms (3.1% and 6.7% respectively), and most likely has a low impact on the validity of test forms' scores, given that their size was not considered notable, according to Wright and Douglas criteria (1975, 1976) and it is common to have about 15% of items with DIF in achievement tests (Buzick & Stone, 2011; Narayanan & Swaminathan, 1994). Post-hoc test results showed that the TRC-n and the TLC-n were able to capture the reading and listening comprehension improvements of students across subsequent grades. Additionally, evidence of adequate reliability was obtained for these final test forms. In summary, the results from the first study provided evidence for good psychometric properties for all forms of the TRC-n and the TLC-n.

Regarding the second study, the TRC-n forms were highly correlated with the TLC-n forms. Correlations of moderate magnitudes between listening and reading comprehension in the same grade levels have been found in other studies (Diakidoy et al., 2005; Ouellette & Beers, 2010; Tobia &

Bonifacci, 2015). The close relationship between reading and listening comprehension is congruent with the idea that cognitive processes involved in both skills are the same (Cain & Oakhill, 2008; Perfetti et al., 2005). Additionally, most correlations between the developed test forms and external criteria were positive and statistically significant. Moderate to high correlations between the TRC-n and the oral reading fluency measure are consistent with findings from other studies with samples of speakers of a wide range of languages and orthographies, enrolled in the fifth and sixth grades (Fernandes et al., 2017; Padeliadu & Antoniou, 2014; Yovanoff et al., 2005). The TRC-n was moderately correlated with reading strategy use in the fifth grade. This result is particularly similar to the one obtained by Law (2009). In contrast, in the sixth grade, this relationship was not statistically significant, similar to the results reported in the study by Kolić-Vehovec and Bajšanski (2006). These mixed findings can be explained by a decrease of the influence of the use of reading strategies on reading comprehension with progress in schooling from lower to higher grades, when other variables, such as vocabulary gain more influence on reading comprehension (Ouellette & Beers, 2010). Congruent with this idea, the results of this study suggest large correlations between vocabulary and the TRC-n and TLC-n forms. Similar results were observed in previous research conducted with fifth and sixth graders (Fernandes et al., 2017; Nouwens et al., 2016; Ouellette & Beers, 2010; Yovanoff et al., 2005). This finding suggests that vocabulary has a strong influence on both reading and listening comprehension in this phase of development.

Most correlations between the TRC-n and TLC-n forms, working memory, and reasoning skills were statistically significant, similar to what has been reported in other studies with students of different ages and countries (Jiang & Farquharson, 2018; Potocki et al., 2017; Ribeiro et al., 2016; Sesma et al., 2009; Tighe et al., 2015). However, the effect size of these relationships was lower in the sixth grade compared to the fifth grade. These results might suggest that cognitive variables, such as working memory and reasoning, assume less weight in comprehension as students reach upper elementary and middle school grades.

Finally, medium-to-large correlation coefficients were found between the TRC-n and TLC-n forms and the teachers' ratings regarding students' performance in oral reading fluency, listening, and reading comprehension, and high correlations were found with academic achievement indicators. Prior studies reported correlations of similar magnitude between these variables (Feinberg & Shapiro, 2003, 2009; Gilmore & Vance, 2007; Santos et al., 2017; Viana et al., 2015). These results suggest that the scores in the developed test forms are a fair representation of actual school achievement of the students. Overall, the findings of the second study provide evidence of validity for the developed test forms.

The development of the TRC-$n_{5/6}$ and the TLC-$n_{5/6}$ forms is an important contribution for the assessment of reading and listening comprehension in the Portuguese educational context; these two vertically scaled test forms, used in combination with the test forms developed for primary school (Santos et al., 2015, 2016), allow for the assessment and monitoring of students' performance in these skills across multiple time points from the first to the sixth grade, allowing the direct comparison of the scores, and avoiding learning effects. Students that obtained low scores in these tests should be referred to a more comprehensive assessment, including the assessment of cognitive abilities, and to an appropriate intervention. Students may present difficulties in listening comprehension or in reading comprehension, or even in both skills (Nation, 2005). Activities focused on the explicit teaching of vocabulary, the activation of previous knowledge, the teaching of comprehension strategies, questioning, comprehension monitoring, making inferences and retelling, are usually effective in the promotion of comprehension skills in both modalities (oral or written) (Hogan et al., 2014; Snowling & Hulme, 2012). Moreover, given that different types of comprehension are assessed, the results in these tests can also inform instructional decision-making, by indicating which types – literal, inferential, reorganization and/or critical comprehension – need further attention. For example, students might be able to respond correctly to questions that require only the comprehension of the information explicitly stated in the text but not be able to make inferences when asked to. In this case, teachers can design lessons to foster inferential comprehension, including strategies such as the expansion, activation and mobilization of relevant previous knowledge or strategies targeting the integration of in-text and out-of-text information (Barth & Elleman, 2017).

These tests are also adequate tools for large-scale testing in schools. Large-scale testing can positively affect different stakeholders. Abu-Alhija (2007) pointed their benefits for students, teachers, parents, administrators and policymakers. Regarding the students, large-scale testing results can encourage students to work more efficiently by describing their knowledge at the time of the assessment and by flagging what needs to be further studied and learned. Teachers can also benefit of the use of this type of tests because large-scale assessments can help to recognize strengths/weakness in the curriculum, to detect deficit areas in the students' knowledge and to redirect instruction. Moreover, this information can motivate teachers to invest in their professional development and to enhance instruction. One of the possible positive effects of large-scale testing in parents is to stimulate their involvement in school activities. The use of these tests can also help administrators, for example, in the evaluation of the qualities of their programs. Additionally, large-scale testing can help

policymakers' in the evaluation of the effectiveness of the educational polices and in the promotion of a better allocation of resources (Abu-Alhija, 2007).

This study also had certain limitations, such as the use of a convenience sampling technique and the centralisation of the data collection in northern Portugal. Thus, the results of the study should be generalised with caution. Another limitation was the maintenance of some items that showed DIF. Although the size of the differences was not very large, these items can introduce some bias in testing, especially when comparing the results of students from different socioeconomic levels.

The collection of evidence based on consequences of testing can be an important aim for future studies. The administration of tests in educational contexts is based on the idea that the interpretation of the scores should be used for realising some benefits, such as improvement of learning and motivation for students through the selection of efficacious instructional strategies. Interviews and focus groups with teachers and students as well as classroom observations can be used to obtain data about this type of validity. Further research may analyse the accuracy classification indices of these tests in identifying students with reading difficulties in the fifth and sixth grades. Taking into account that prior literature has shown that different results can be obtained for comprehension tests with texts of distinct typologies (RAND Reading Group, 2002), future research may also focus on development of the TRC-n and the TLC-n for the same grade levels using expository texts. Consequently, the separate or combined use of reading and listening comprehension tests with narrative and expository texts may contribute to the identification of specific comprehension difficulties and then guide intervention programmes more centred on training strategies for extracting meaning from texts of one or both typologies.

References

Abu-Alhija, F. N. (2007). Large-scale testing: Benefits and pitfalls. *Studies in Educational Evaluation*, *33*(1), 50–68. https://doi.org/10.1016/j.stueduc.2007.01.005

Almeida, L. S., & Lemos, G. (2006). *Bateria de Provas de Raciocínio: Manual técnico [Battery of Reasoning Tests: Technical manual]*. Centro de Investigação em Psicologia, Universidade do Minho.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). Validity. In *Standards for educational and psychological testing* (pp. 11–31). American Educational Research Association.

Baker, F. B. (1984). Ability metric transformations involved in vertical equating under item response theory. *Applied Psychological Measurement*, *8*(3), 261–271. https://doi.org/10.1177/014662168400800302

Barth, A. E., & Elleman, A. (2017). Evaluating the impact of a multistrategy inference intervention for middle-grade struggling readers. *Language, Speech, and Hearing Services in Schools*, *48*(1), 31–41. https://doi.org/10.1044/2016_LSHSS-16-0041

Botsas, G. (2017). Differences in strategy use in the reading comprehension of narrative and science texts among students with and without learning disabilities. *Learning Disabilities: A Contemporary Journal*, *15*(1), 139–162. https://doi.org/10.4324/9780203876428

Buzick, H., & Stone, E. (2011). *Recommendations for conducting Differential Item Functioning (DIF) analyses for students with disabilities based on previous DIF studies*. Educational Testing Service.

Cain, K., & Oakhill, J. (2008). Cognitive bases of children's language comprehension difficulties. Where do we go from here? In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language. A cognitive perspective* (pp. 283–295). The Guilford Press.

Català, G., Català, M., Molina, E., & Monclús, R. (2001). *Evaluación de la comprensión lectora: Pruebas ACL (1º - 6º de primaria)*. Editorial Graó.

Catts, H. W., Hogan, T. P., & Adlof, S. M. (2005). Developmental changes in reading and reading disabilities. In H. W. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 25–40). Lawrence Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

de Ayala, R. J. (2009). *The theory and practice of item response theory*. The Guilford Press.

Delgado, A. R., & Prieto, G. (1998). Further evidence favoring three-option items in multiple-choice tests. *European Journal of Psychological Assessment*, *14*(3), 197–201. https://doi.org/10.1027/1015-5759.14.3.197

Diakidoy, I. A. N., Stylianou, P., Karefillidou, C., & Papageorgiou, P. (2005). The relationship between listening and reading comprehension of different types of text at increasing grade levels. *Reading Psychology*, *26*(1), 55–80. https://doi.org/10.1080/02702710590910584

Feinberg, A. B., & Shapiro, E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly*, *18*(1), 52–65. https://doi.org/10.1521/scpq.18.1.52.20876

Feinberg, A. B., & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research*, *102*(6), 453–462. https://doi.org/10.3200/JOER.102.6.453-462

Fernandes, S., Querido, L., Verhaeghe, A., Marques, C., & Araújo, L. (2017). Reading development in European Portuguese: Relationships between oral reading fluency, vocabulary and reading comprehension. *Reading and Writing*, *30*(9), 1987–2007. https://doi.org/10.1007/s11145-017-9763-z

Florit, E., Roch, M., & Levorato, M. C. (2014). Listening text comprehension in preschoolers: A longitudinal study on the role of semantic components. *Reading and Writing*, *27*(5), 793–817. https://doi.org/10.1007/s11145-013-9464-1

Francis, D. J., Fletcher, J. M., Catts, H. W., & Tomblin, J. B. (2005). Dimensions affecting the assessment of reading comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 369–394). Lawrence Erlbaum Associates.

Gilmore, J., & Vance, M. (2007). Teacher ratings of children's listening difficulties. *Child Language Teaching and Therapy*, *23*(2), 133–156. https://doi.org/10.1177/0265659007073876

Hogan, T. P., Adlof, S. M., & Alonzo, C. N. (2014). On the importance of listening comprehension. *International Journal of Speech-Language Pathology*, *16*(3), 199–207. https://doi.org/10.3109/17549507.2014.904441

Jiang, H., & Farquharson, K. (2018). Are working memory and behavioral attention equally important for both reading and listening comprehension? A developmental comparison. *Reading and Writing*, *31*(7), 1449–1477. https://doi.org/10.1007/s11145-018-9840-y

Kim, Y.-S. G. (2016). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *Journal of Experimental Child Psychology*, *141*, 101–120. https://doi.org/10.1016/j.jecp.2015.08.003

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). Springer. https://doi.org/10.1007/978-1-4939-0317-7

Kolić-Vehovec, S., & Bajšanski, I. (2006). Metacognitive strategies and reading comprehension in elementary-school students. *European Journal of Psychology of Education*, *21*(4), 439–451. https://doi.org/10.1007/bf03173513

Law, Y. K. (2009). The role of attribution beliefs, motivation and strategy use in Chinese fifth-graders' reading comprehension. *Educational Research*, *51*(1), 77–95. https://doi.org/10.1080/00131880802704764

Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology*, *95*(2), 211–224. https://doi.org/10.1037/0022-0663.95.2.211

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, *16*(2), 878.

Linacre, J. M. (2018). *A user's guide to WINSTEPS and MINISTEP: Rasch-model computer programs. Program manual 4.3.1.* Retrieved January 15, 2020, from https://www.winsteps.com

Lipka, O., Lesaux, N. K., & Siegel, L. S. (2006). Retrospective analyses of the reading development of grade 4 students with reading disabilities: Risk status and profiles over 5 years. *Journal of Learning Disabilities*, *39*(4), 364–378. https://doi.org/10.1177/00222194060390040901

Monteiro, A., Fonseca, D., Rodrigues, E. M., Monteiro, I., Rebelo, I., Silva, M. B., & Ferreira, R. (2014). *Processo de Avaliação Externa da Aprendizagem – Provas Finais de Ciclo e Exames Nacionais 2014 [External assessment process of learning – End of cycle exams and national exams 2014]*. Direção-Geral da Educação, Ministério da Educação e Ciência. Retrieved March 4, 2020, from http://www.dge.mec.pt/sites/default/files/JNE/relatorio_anual_do_jne_2014.pdf

Narayanan, P., & Swaminathan, H. (1994). Performance of the Mantel-Haenszel and simultaneous item bias procedures for detecting Differential Item Functioning. *Applied Psychological Measurement*, *18*(4), 315–328. https://doi.org/10.1177/014662169401800403

Nation, K. (2005). Children's reading comprehension difficulties. In M. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 248–265). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470757642

Nouwens, S., Groen, M. A., & Verhoeven, L. (2016). How storage and executive functions contribute to children's reading comprehension. *Learning and Individual Differences*, *47*, 96–102. https://doi.org/10.1016/j.lindif.2015.12.008

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill, Inc.

Oakhill, J., Cain, K., & Elbro, C. (2019). Reading comprehension and reading comprehension difficulties. In D. A. Kilpatrick, R. M. Joshi, & R. K. Wagner (Eds.), *Reading development and difficulties: Bridging the gap between research and practice* (pp. 83–115). Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-26550-2_5

Ouellette, G. P., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing*, *23*(2), 189–208. https://doi.org/10.1007/s11145-008-9159-1

Padeliadu, S., & Antoniou, F. (2014). The relationship between reading comprehension, decoding, and fluency in Greek: A cross-sectional study. *Reading & Writing Quarterly*, *30*(1), 1–31. https://doi.org/10.1080/10573569.2013.758932

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Blackwell. https://doi.org/10.1002/9780470757642

Potocki, A., Sanchez, M., Ecalle, J., & Magnan, A. (2017). Linguistic and cognitive profiles of 8- to 15-year-old children with specific reading comprehension difficulties: The role of executive functions. *Journal of Learning Disabilities*, *50*(2), 128–142. https://doi.org/10.1177/0022219415613080

Prieto, G., & Delgado, A. R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, *15*(1), 94–100.

RAND Reading Group (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. RAND Education. https://doi.org/10.1044/leader.ov.22102017.np

Ribeiro, I., Cadime, I., Freitas, T., & Viana, F. L. (2016). Beyond word recognition, fluency, and vocabulary: The influence of reasoning on reading comprehension. *Australian Journal of Psychology*, *68*(2), 107–115. https://doi.org/10.1111/ajpy.12095

Ribeiro, I., Dias, O., Oliveira, Í. M., Miranda, P., Ferreira, G., Saraiva, M., Paulo, R., & Cadime, I. (2015). Adaptação e Validação da Escala Reading Strategy use para a População Portuguesa. *Revista Iberoamericana de Diagnostico y Evaluacion Psicologica*, *1*(40), 25–36.

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, *24*(2), 3–13. https://doi.org/10.1111/j.1745-3992.2005.00006.x

Salvia, J., Ysseldyke, J. E., & Bolt, S. (2010). Technical adequacy. In J. Salvia, J. E. Ysseldyke, & S. Bolt (Eds.), *Assessment in special and inclusive education* (11th ed., pp. 53–71). Wadsworth Cengage Learning.

Santos, S., Cadime, I., Viana, F. L., Chaves-Sousa, S., Gayo, E., Maia, J., & Ribeiro, I. (2017). Assessing reading comprehension with narrative and expository texts: Dimensionality and relationship with fluency, vocabulary and memory. *Scandinavian Journal of Psychology*, *58*(1), 1–8. https://doi.org/10.1111/sjop.12335

Santos, S., Cadime, I., Viana, F. L., Prieto, G., Chaves-Sousa, S., Spinillo, A. G., & Ribeiro, I. (2016). An application of the Rasch model to reading comprehension measurement. *Psicologia: Reflexão e Crítica*, *29*(38), 1–19. https://doi.org/10.1186/s41155-016-0044-6

Santos, S., Viana, F. L., Ribeiro, I., Prieto, G., Brandão, S., & Cadime, I. (2015). Development of listening comprehension tests with narrative and expository texts for Portuguese students. *Spanish Journal of Psychology*, *18*(e5), 1–7. https://doi.org/10.1017/sjp.2015.7

Sesma, H. W., Mahone, E. M., Levine, T., Eason, S. H., & Cutting, L. E. (2009). The contribution of executive skills to reading comprehension. *Child Neuropsychology*, *15*(3), 232–246. https://doi.org/10.1080/09297040802220029

Simões, M. R., & Albuquerque, C. P. (2002). Estudos com a versão portuguesa da WISC-III no âmbito

da validade concorrente e preditiva: Relação com as classificações escolares [Predictive validity of the Portuguese version of the WISC-III: Relationship with school grades]. *Psychologica*, *29*, 153–168.

Snowling, M. J., & Hulme, C. (2012). Interventions for children's language and literacy difficulties. *International Journal of Language and Communication Disorders*, *47*(1), 27–34. https://doi.org/10.1111/j.1460-6984.2011.00081.x

Tighe, E. L., Spencer, M., & Schatschneider, C. (2015). Investigating predictors of listening comprehension in third-, seventh-, and tenth-grade students: A dominance analysis approach. *Reading Psychology*, *36*(8), 700–740. https://doi.org/10.1080/02702711.2014.963270

Tobia, V., & Bonifacci, P. (2015). The simple view of reading in a transparent orthography: The stronger role of oral comprehension. *Reading and Writing*, *28*(7), 939–957. https://doi.org/10.1007/s11145-015-9556-1

Tzuriel, D., & George, T. (2009). Improvement of analogical reasoning and academic achievements by the Analogical Reasoning Programme (ARP). *Educational and Psychology Child*, *26*(3), 71–94.

Viana, F. L., Santos, S., Ribeiro, I., Chaves-Sousa, S., Brandão, S., Cadime, I., & Maia, J. (2015). Listening comprehension assessment: Validity studies of two vertically scaled tests for Portuguese students. *Universitas Psychologica*, *14*(1), 345–354. https://doi.org/10.11144/Javeriana.upsy14-1.lcav

Walker, C. M., & Beretvas, S. N. (2001). An empirical investigation demonstrating the multidimensional DIF paradigm: A cognitive explanation for DIF. *Journal of Educational Measurement*, *38*(2), 147–163. https://doi.org/10.1111/j.1745-3984.2001.tb01120.x

Wechsler, D. (2003). *Escala de Inteligência de Wechsler para Crianças – Terceira edição (WISC–III) [Wechsler Intelligence Scale for Children – Third edition (WISC–III)]*. CEGOC-TEA.

Wilson, M., & Moore, S. (2011). Building out a measurement model to incorporate complexities of testing in the language domain. *Language Testing*, *28*(4), 441–462. https://doi.org/10.1177/0265532210394142

Wright, B. D., & Douglas, G. A. (1975). *Best test design and self-tailored testing (MESA Research Memorandum No. 19).* Statistical Laboratory, Department of Education. University of Chicago.

Wright, B. D., & Douglas, G. A. (1976). *Rasch item analysis by hand (MESA Research Memorandum No. 21)*. Statistical Laboratory, Department of Education. University of Chicago.

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice*, *24*(3), 4–12. https://doi.org/10.1111/j.1745-3992.2005.00014.x

CHAPTER 2

ASSESSING ORAL READING FLUENCY WITHIN AND ACROSS GRADE LEVELS: DEVELOPMENT OF EQUATED TEST FORMS

Chapter 2

Assessing Oral Reading Fluency Within and Across Grade Levels: Development of Equated Test Forms[2]

Introduction

Learning to read is a central goal during elementary school. Oral reading fluency (ORF) implies to read with accuracy, speed and expressiveness (Hudson et al., 2005). However, accuracy and speed are the dimensions most commonly assessed in psychoeducational research and practice (e.g., Deeney, 2010; Valencia et al., 2010). Generally, the student's task consists of reading text(s) aloud for 1 min as the evaluator records the errors and signals the last word read and then calculates the number of words read correctly (Hasbrouck & Tindal, 2006). Thus, 1-min ORF measures have been widely used in the last two decades, largely due to international recommendation to evaluate ORF regularly (National Institute of Child Health and Human Development, 2000).

The development of ORF measures raises particular issues, given that this is a skill that is expected to develop across time as a consequence of learning and training processes (Arnesen et al., 2017). One of these challenges is related to its assessment at multiple timepoints in a same grade or across distinct grades. Applying the same text(s) across more than one timepoint allows to directly compare ORF scores but these may be affected by learning effects (Chaparro et al., 2018). Moreover, using the same text(s) across very different grade levels, mostly likely will lead to floor or ceiling effects because the difficulty level will not be adequate for all grade levels (McCoach & Yu, 2016). Therefore, the use of multiple texts or test forms and the application of statistical equating procedures has been recommended as a viable solution to avoid these measurement biases and to guarantee the scores' equivalence (Albano & Rodriguez, 2012; Ardoin et al., 2010; Ardoin & Christ, 2009; Betts et al., 2009; Cummings et al., 2013; Denton et al., 2011; Francis et al., 2008; Petscher & Kim, 2011; Stoolmiller et al., 2013).

Horizontal equating places scores from multiple texts or test forms on a single metric, enabling their interchangeable use, while vertical linking aligns the metrics from test forms with distinct difficulty levels, such as forms for different grade levels, producing comparable score scales over time (Kolen & Brennan, 2014; Santi et al., 2016).

---

[2] This work is published in the journal *Behavior Research Methods*.

Rodrigues, B., Cadime, I., Freitas, T., Choupina, C., Baptista, A., Viana, F. L., & Ribeiro, I. (2022). Assessing oral reading fluency within and across grade levels: Development of equated test forms. *Behavior Research Methods.* Advance online publication. https://doi.org/10.3758/s13428-022-01806-7

Reliability and validity are fundamental requisites for tests' development. Test-retest is frequently used as method for estimate reliability of ORF tests (American Educational Research Association et al., 2014; Salvia et al., 2010). Test-retest correlation coefficients for these type of measures are usually higher than .80 (Alonzo et al., 2012; Anderson, Lai, et al., 2012; Anderson, Park, et al., 2012; Center on Teaching and Learning, 2018; Lai et al., 2012; McGlinchey & Hixson, 2004; Park et al., 2012). Although the time intervals used for the test-retest of ORF tests are largely varied, it is recommended that students are tested and then retested 2 weeks later because ORF's scores will be more stable across small intervals of time (Salvia et al., 2010).

The collection of validity evidence based on relations to other variables requires the identification of pertinent variables that are related to ORF (American Educational Research Association et al., 2014). Accuracy in word identification is one of the ORF dimensions and therefore moderate-to-high correlations between ORF and word reading measures have been consistently obtained in prior research (Adlof et al., 2006; Cadime et al., 2017; Chaves-Sousa, Ribeiro, et al., 2017; Padeliadu & Antoniou, 2014; Spear-Swerling, 2006; Speece et al., 2010). ORF is also a condition for successful reading comprehension, given that the automaticity in reading frees the cognitive resources for the extraction and construction of meaning from text (National Institute of Child Health and Human Development, 2000). Thus, correlations between ORF and reading comprehension measures have also been found in students from grades 1 to 6 across a wide variety of orthographies (Cadime et al., 2017; Calet et al., 2015; Fernandes et al., 2017; Kim et al., 2012; Little et al., 2017; Padeliadu & Antoniou, 2014; Ribeiro et al., 2016; Sabatini et al., 2019; Santos et al., 2017; Shapiro et al., 2008; Silberglitt et al., 2006; Valencia et al., 2010; Yovanoff et al., 2005). Regarding convergent validity, high correlations between the scores obtained on different tests of ORF have consistently been found (e.g., Lai, Alonzo, & Tindal, 2013). Previous studies have also shown medium-to-large correlation coefficients between reading performance of students and teachers' evaluations (Begeny et al., 2008, 2011; Bigozzi et al., 2017; Feinberg & Shapiro, 2003, 2009).

The Present Study

Several international measures of reading have been developed for elementary school students (e.g., Texas Primary Reading Inventory – TPRI, Children's Learning Institute & Texas Institute for Measurement, Evaluation, and Statistics, 2010; Texas Middle Fluency Assessment – TMSFA, Francis et al., 2010; Dynamic Indicators of Basic Early Literacy Skills – DIBELS, Good & Kaminski, 2011; Gray Oral Reading Test – GORT-5; Hall & Tannebaum, 2013; AIMSweb Standard Reading Assessment

Passages, Howe & Shinn, 2002) as a consequence of the test developers' efforts for responding the need of screening and progress-monitoring students' performance in skills that, like ORF, undergo rapid increases and must be consistently monitored over time. Most studies conducted with ORF's measures have only performed horizontal equating procedures (e.g., Albano & Rodriguez, 2012; Betts et al., 2009; Francis et al., 2008; Stoolmiller et al., 2013). However, Chaparro et al. (2018) referred that test developers should also consider the use of vertical linking (or vertical equating) across grade levels considering that it will produce a continuous progress-monitoring scale that can be used over various years of students' reading development. In the Portuguese context, the existing ORF standardized tests (e.g., Test of Reading Fluency and Accuracy: The King [*Teste de Avaliação da Fluência e Precisão da Leitura: O Rei*], Carvalho, 2010) include only a single text to be used with students from different grades and no equating procedures were applied, and therefore are inadequate for monitoring across time. Thus, the main goal of this study was to develop and validate an ORF measure with horizontally and vertically equated test forms for Portuguese elementary students: the test of fluency assessment (TAF – *Teste de Avaliação da Fluência*). For this purpose, two studies were conducted in order to describe the development and equating process of the TAF test forms from grades 1 to 6 (study 1), as well as to examine test-retest reliability of the TAF test forms and gather validity evidence based on relationship to other variables that empirical studies have systematically related to ORF, namely word reading, ORF measured by other test, reading comprehension, teachers' judgments and school outcomes (study 2).

Study 1

Method

Participants

The total sample consisted of 1,166 students attending public (n = 970, 83.19%) and private (n = 196, 16.81%) schools located in northern Portugal. Two hundred thirty-three first graders (53.2% female; M_{age} = 6.83, SD = 0.41), 220 second graders (46.4% female; M_{age} = 7.80, SD = 0.56), 218 third graders (54.6% female; M_{age} = 8.54, SD = 0.50), 218 fourth graders (50.9% female; M_{age} = 9.52, SD = 0.53), 139 fifth graders (48.9% female; M_{age} = 10.51, SD = 0.53) and 138 sixth graders (47.1% female; M_{age} = 11.48, SD = 0.62) participated in this study. All students were native European Portuguese speakers. Students identified with support measures for learning and inclusion at the secondary and/or tertiary levels were not included in this study (Decree-Law no. 54/2018 of July 6).

Instrument

The TAF is an original test, developed to assess ORF from grades 1 to 6. The test forms per grade were as following: grade 1 – one test form; grades 2 to 4 – three test forms per grade; grades 5 and 6 – two test forms per grade. The main motivation for the development of the TAF was the possibility to assess oral reading fluency of the students across the first 6 years of basic education and at different time points. Grade 1 is the first year of formal reading instruction and students first master decoding at the word level, and later at the phrase and text level. Thus, assessing text fluency is, in most cases, only possible in the last months of grade 1. This is the reason why there is only one test form for grade 1. As students advance from lower to higher grade levels, students increasingly improve their text reading fluency skills. However, longitudinal studies focused on the growth trajectories in oral reading fluency have provided empirical evidence of a deceleration in oral reading fluency growth in most advanced grade levels (e.g., Arnesen et al., 2017). Therefore, in more advanced grades, ORF measurements are conducted fewer times across the school year than in earlier grades. Based on this assumption, we developed three test forms for grades 2, 3 and 4 and only two test forms for grades 5 and 6. Each test form includes three unpublished texts from three typologies: one narrative with dialogue, one narrative without dialogue and one expository. The length of these texts increased with grade level: for grade 1, the texts length ranged from 197 to 243 words, for grade 2 from 236 to 275 words, for grade 3 from 352 to 390 words, for grade 4 from 422 to 507 words, for grade 5 from 364 to 643 words and for grade 6 from 440 to 659 words.

Students were asked to read aloud each text during 1 min as the evaluator registered the errors (i.e., omissions, substitutions, mispronunciations, and hesitations of more than 3 s) and signalled the last word read. The number of correct words read per minute in the three texts was averaged for each test form.

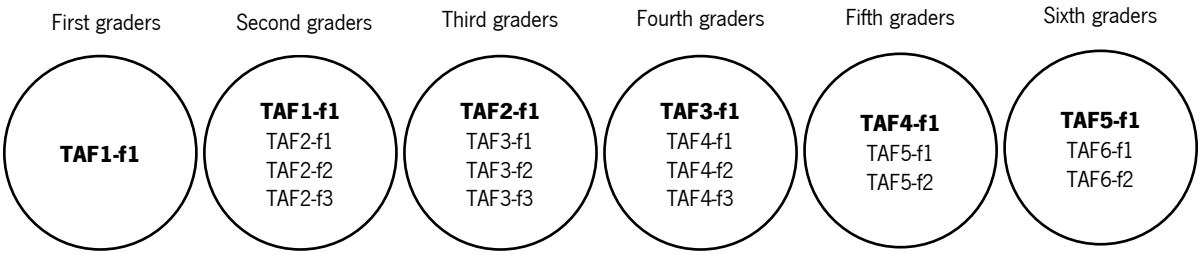
Design and Procedures

A non-equivalent groups with anchor test (NEAT) design was adopted (de Ayala, 2009) in order to perform vertical linking from the TAF1 test forms to the TAF6 test forms. This design involves two or more groups with different ability levels that respond to different test forms (Kolen & Brennan, 2014). The groups are non-equivalent because they comprise students in different grade levels that differ in ORF ability. Therefore, students read aloud all texts that compose the test forms for their grade – unique test forms – and three texts of one test form designed for the adjacent grade – anchor or common test form (see Figure 3).

Data collection was authorized by the Portuguese Education Ministry and school boards. Informed consent was obtained from students' parents or legal tutors. The administration of the TAF texts was random and individual and conducted by trained psychologists.

Figure 3

NEAT Design



Note. TAF1 = TAF for grade 1, TAF2 = TAF for grade 2, TAF3 = TAF for grade 3, TAF4 = TAF for grade 4, TAF5 = TAF for grade 5, TAF6 = TAF for grade 6, f1 = test form 1, f2 = test form 2, f3 = test form 3. Common test forms appear in bold.

Statistical Analyses

Equipercentile equating was performed to equate test forms of the TAF within each grade (Albano, 2011). When it is intended to obtain accurate scores along the full score scale, and not only near the mean, equipercentile equating is the most adequate equating method (Kolen & Brennan, 2014). Equipercentile equating controls for variance and it is the only method that adjusts differences in the overall distribution of raw scores across texts or test forms (Cummings et al., 2013). Moreover, it accommodates any degree of linearity or nonlinearity across texts or test forms (Albano & Rodriguez, 2012), creating psychometrically equivalent texts or test forms (Stoolmiller et al., 2013). Large samples (N = 1500) are preferred for performing this statistical equating (Kolen & Brennan, 2014). When it is not possible to attain this sample size, a smoothing step is particularly necessary to eliminate irregularities in the distribution of the raw scores but maintaining the same location, spread and shape (Livingston, 2014).

In a first step, TAF test forms from grades 2 to 6 were horizontally equated using equipercentile equating in order to place test forms of each grade level on the same metric. The reference test forms were TAF2-f1 for grade 2, TAF3-f1 for grade 3, TAF4-f1 for grade 4, TAF5-f1 for grade 5, TAF6-f1 for grade 6. Log-linear pre-smoothing was applied to smooth sample score distributions, reducing equating error (Albano, 2011). In a second step, to ensure the comparability of the scores between grades, the

chained equipercentile method with NEAT design was used to vertical link the TAF test forms. Kolen and Brennan (2014) suggested that this method should be considered when the equating groups differ significantly in ability. Empirical research has also suggested that chained equipercentile method tends to produce slightly more accurate scores (smaller bias) when compared with other equating methods in the context of the NEAT design (Holland et al., 2008; Sinharay & Holland, 2010; Wang et al., 2008).

The third and final step of the equating process was focused in the resolution of a practical issue: "equating relationship cannot be determined for the parts of the score range above the highest score you observe and below the lowest score you observe" (Livingston, 2014, p. 22). In these cases, one of the possible solutions is to fix these scores at particular values (Kolen & Brennan, 2014). Therefore, in this study, the lowest/highest possible equated scores were assigned to the extreme, non-observed, raw scores. These analyses, as well as descriptive statistics of the raw and equated scores, were performed with "equate" R package (Albano, 2016).

Correlation coefficients and repeated measures analyses of variance were performed to test for differences in raw and equated scores within each grade level. One-way analyses of variance were computed to test between-grade differences in the equated scores of the TAF. These analyses were performed with R statistical software. The correlations' magnitude was evaluated using Cohen's (1992) recommendations: .10 indicates a small effect, .30 a medium effect and .50 a large effect. Cohen's *d* was used as a measure of effect size for the analyses of variance and was interpreted according to the following guidelines: 0.20 suggests a small effect, 0.50 a medium effect and 0.80 a large effect (Cohen, 1992).

Results

Correlation coefficients between TAF test forms within each grade level were high, ranging between .96 and .98, and statistically significant ($p < .001$), suggesting strong linear relationships. Table 5 presents the descriptive statistics of the raw scores in each test form. The skewness and kurtosis values were low, suggesting no robust violations to the normality of the distributions. Results from repeated measures analysis of variance showed statistically significant differences in the raw scores between test forms within each grade from grades 2 to 4: $F(2,438) = 46.81$, $p < .001$ for the TAF2 test forms, $F(2,434) = 107.98$, $p < .001$ for the TAF3 test forms, and $F(2,434) = 103.98$, $p < .001$ for the TAF4 test forms. Bonferroni post-hoc comparisons showed the mean's number of words read correctly per minute obtained in the TAF2-f2 and the TAF2-f3 was significantly higher than that obtained in the TAF2-f1 ($d = 0.12$ for both pairwise comparisons). There were no statistically significant

differences between the TAF2-f2 and the TAF2-f3 (d = 0.00). The mean's number of words read correctly per minute obtained in the TAF3-f2 was significantly higher than those obtained in the TAF3-f1 (d = 0.09) and the TAF3-f3 (d = 0.22), as well the mean's number of words read correctly per minute obtained in the TAF3-f1 was significantly higher than that obtained in the TAF3-f3 (d = 0.12). The mean's number of words read correctly per minute obtained in the TAF4-f1 and the TAF4-f2 was significantly higher than that obtained in the TAF4-f3 (d = 0.20 and d = 0.23 for each pairwise comparison, respectively). There were no statistically significant differences between the TAF4-f1 and the TAF4-f2 (d = 0.04). Results from repeated measures analysis of variance showed statistically significant differences in the raw scores between test forms within grade 5: F (1,138) = 393.79, p < .001, d = 0.39, for the TAF5 test forms and marginally significant differences in the raw scores between test forms within grade 6: F (1,137) = 3.154, p = .078, d = 0.04, for the TAF6 test forms. The effect size was small for all pairwise comparisons within grade levels. These results suggest that the raw scores of the different test forms cannot be directly compared within grade levels.

Table 5

Descriptive Statistics of the Raw Scores

Test form	Total number of words[a]	*M*	*SD*	Min.	Max.	Percentile			Skewness	Kurtosis
						25	50	75		
Grade 1										
TAF1-f1	219.33	45.22	21.93	3.00	144.00	31.00	42.00	58.00	0.90	1.73
Grade 2										
TAF1-f1	**219.33**	**77.61**	**26.79**	**19.00**	**184.00**	**58.75**	**74.00**	**93.00**	**0.69**	**1.03**
TAF2-f1	260.00	75.30	26.10	19.00	189.00	59.00	71.00	88.25	0.99	1.99
TAF2-f2	265.67	78.63	27.89	17.00	195.00	59.00	77.50	95.25	0.69	1.15
TAF2-f3	263.33	78.60	27.88	19.00	188.00	59.75	76.00	92.00	0.67	0.85
Grade 3										
TAF2-f1	**260.00**	**102.47**	**28.74**	**43.00**	**208.00**	**83.00**	**98.50**	**120.00**	**0.60**	**0.38**
TAF3-f1	378.67	96.66	30.09	41.00	191.00	74.25	92.50	116.75	0.43	–0.25
TAF3-f2	377.00	99.33	29.15	44.00	193.00	78.00	95.00	119.00	0.51	–0.07
TAF3-f3	374.33	93.06	28.42	43.00	185.00	72.00	87.50	109.00	0.66	–0.02
Grade 4										
TAF3-f1	**378.67**	**118.02**	**25.64**	**37.00**	**175.00**	**101.00**	**118.00**	**135.75**	**–0.13**	**–0.26**
TAF4-f1	474.33	114.57	22.58	36.00	171.00	100.00	115.00	130.75	–0.05	0.02
TAF4-f2	473.67	115.51	24.41	34.00	172.00	101.00	115.00	131.00	–0.14	–0.10
TAF4-f3	474.33	109.88	23.90	34.00	174.00	93.00	109.00	127.00	–0.11	–0.24
Grade 5										
TAF4-f1	**474.33**	**127.32**	**25.29**	**75.00**	**227.00**	**112.50**	**124.00**	**143.00**	**0.53**	**0.93**
TAF5-f1	537.67	121.32	25.97	51.00	217.00	106.50	120.00	138.00	0.21	0.79
TAF5-f2	534.67	131.58	26.60	65.00	233.00	116.00	128.00	148.00	0.35	0.77
Grade 6										
TAF5-f1	**537.67**	**137.59**	**24.31**	**82.00**	**191.00**	**121.00**	**136.50**	**155.00**	**–0.14**	**–0.55**
TAF6-f1	585.33	136.55	23.03	77.00	209.00	122.25	133.50	151.00	0.06	–0.04
TAF6-f2	573.67	137.48	22.95	80.00	183.00	122.00	137.00	154.00	–0.17	–0.59

Note. [a]Average across the texts that compose each test form, *M* = Mean, *SD* = Standard Deviation, Min. = Minimum, Max. = Maximum, TAF1 = TAF for grade 1, TAF2 = TAF for grade 2, TAF3 = TAF for grade 3, TAF4 = TAF for grade 4, TAF5 = TAF for grade 5, TAF6 = TAF for grade 6, f1 = test form 1, f2 = test form 2, f3 = test form 3. The raw scores correspond to the number of correct words read per minute. Anchor test forms are in bold.

Descriptive statistics of the equated scores by test form are displayed in Table 6. No statistically significant differences were found between equated scores on test forms within each grade: $F(2,438) =$

0.152, $p = .859$, for the TAF2 test forms; $F(2,434) = 0.077$, $p = .926$, for the TAF3 test forms; $F(2,434) = 0.004$, $p = .996$, for the TAF4 test forms; $F(1,138) = 0.012$, $p = .912$, for the TAF5 test forms; $F(1,137) = 0.040$, $p = .842$, for the TAF6 test forms. These results provide empirical evidence for the interchangeable use of the TAF test forms within grade levels because equating procedures guaranteed the scores' equivalence.

The mean of equated scores was significantly greater as grade levels increased (see Table 6): $F(5,3682) = 996.70$, $p < .001$. Tukey post-hoc tests showed statistically significant differences between all means of equated scores obtained from the TAF1 to the TAF6 ($p < .001$). The effect sizes for the means of equated scores between adjacent grade levels ranged from large to medium: $d = 1.23$ (grades 1 and 2), $d = 1.02$ (grades 2 and 3), $d = 0.76$ (grades 3 and 4), $d = 0.50$ (grades 4 and 5) and $d = 0.68$ (grades 5 and 6).

Table 6

Descriptive Statistics of the Equated Scores

Test form	*M*	*SD*	Min.	Max.	Percentile			Skewness	Kurtosis
					25	50	75		
Grade 1									
TAF1-f1	45.22	21.93	3.00	144.00	31.00	42.00	58.00	0.90	1.73
Grade 2									
TAF1-f1	**77.61**	**26.79**	**19.00**	**184.00**	**58.75**	**74.00**	**93.00**	**0.69**	**1.03**
TAF2-f1	77.04	26.88	18.00	182.00	58.00	73.00	92.25	0.73	0.78
TAF2-f2	77.15	27.09	20.00	191.00	57.00	76.50	92.25	0.76	1.24
TAF2-f3	77.25	27.10	17.00	191.00	58.75	74.00	91.00	0.73	1.29
Mean TAF2	77.26	26.73	18.50	187.00	57.06	74.25	92.38	0.73	1.10
Grade 3									
TAF2-f1	**104.95**	**27.86**	**44.00**	**191.00**	**86.00**	**104.00**	**123.00**	**0.27**	**–0.20**
TAF3-f1	104.74	27.31	45.00	184.00	85.25	102.50	123.75	0.22	–0.32
TAF3-f2	104.87	27.40	44.00	184.00	85.00	102.50	124.00	0.21	–0.27
TAF3-f3	104.88	27.21	44.00	184.00	86.00	102.50	123.00	0.24	–0.36
Mean TAF3	104.86	27.21	46.50	185.75	85.75	102.00	122.81	0.24	0.30
Grade 4									
TAF3-f1	**124.04**	**22.32**	**43.00**	**174.00**	**110.00**	**125.00**	**139.00**	**–0.28**	**0.11**
TAF4-f1	124.04	22.83	43.00	178.00	110.00	125.00	140.75	–0.20	0.07
TAF4-f2	124.06	22.95	43.00	176.00	110.00	124.00	139.00	–0.22	0.24
TAF4-f3	124.02	22.95	43.00	178.00	108.00	124.00	141.00	–0.26	0.06
Mean TAF4	124.04	22.48	43.00	176.50	109.75	124.38	140.38	–0.25	0.14
Grade 5									
TAF4-f1	**135.94**	**22.92**	**82.00**	**178.00**	**122.50**	**134.00**	**152.00**	**–0.10**	**–0.50**
TAF5-f1	135.09	21.63	81.00	174.00	121.50	135.00	152.00	–0.27	–0.54
TAF5-f2	135.04	21.83	82.00	174.00	122.00	132.00	152.00	–0.23	–0.52
Mean TAF5	135.35	21.87	81.67	175.33	122.33	134.33	153.67	–0.19	–0.52
Grade 6									
TAF5-f1	**148.69**	**19.18**	**99.00**	**173.00**	**136.00**	**150.50**	**165.00**	**–0.71**	**–0.41**
TAF6-f1	150.25	20.91	99.00	177.00	137.25	151.50	168.00	–0.64	–0.63
TAF6-f2	150.17	21.13	99.00	177.00	134.00	154.00	169.00	–0.61	–0.67
Mean TAF6	149.71	20.20	99.00	175.33	135.25	153.00	167.50	–0.65	–0.56

Note. M = Mean, *SD* = Standard Deviation, Min. = Minimum, Max. = Maximum, TAF1 = TAF for grade 1, TAF2 = TAF for grade 2, TAF3 = TAF for grade 3, TAF4 = TAF for grade 4, TAF5 = TAF for grade 5, TAF6 = TAF for grade 6, f1 = test form 1, f2 = test form 2, f3 = test form 3. Mean TAF = mean of the equated scores obtained on TAF test forms of each grade. Anchor test forms are in bold.

Study 2

Method

Participants

A subsample of the previous study, composed of 223 students, was used in order to analyze validity evidence of the TAF1-4 based on test-criterion relationships. The distribution per grade was as following: 54 first graders (M_{age} = 6.91, *SD* = 0.35; 59.3% female), 61 second graders (M_{age} = 7.52, *SD* = 0.67; 50.8% female), 57 third graders (M_{age} = 8.65, *SD* = 0.48; 64.9% female) and 51 fourth graders (M_{age} = 9.57, *SD* = 0.50; 66.7% female). Another sample of 105 students participated in the study of convergent evidence and reliability of the TAF_{1-4}: 27 first graders (M_{age} = 6.37, *SD* = 0.49; 44.4% female), 26 second graders (M_{age} = 7.12, *SD* = 0.33; 69.2% female), 26 third graders (M_{age} = 8.23, *SD* = 0.51; 50% female) and 26 fourth graders (M_{age} = 9.08, *SD* = 0.27; 61.5% female). Regarding the study of convergent and test-criterion validity of the TAF_{5-6}, a sample of 94 fifth graders (M_{age} = 10.96, *SD* = 0.55; 52.1% female) and 85 sixth graders (M_{age} = 12.01, *SD* = 0.50; 47.1% female) was used. This sample had already participated in a previous study to validate comprehension tests (Rodrigues et al., 2020). Another sample of 21 fifth graders (M_{age} = 10.48, *SD* = 0.51; 52.4% female) and 21 sixth graders (M_{age} = 11.43, *SD* = 0.51; 47.6% female) participated in the reliability study of TAF_{5-6}. The students' eligibility criteria in this study were similar to those adopted in study 1.

Measures

Fluency Assessment Test [TAF, Teste de Avaliação da Fluência]

TAF test forms described in study 1 were used to evaluate ORF in students attended from grades 1 to 6. The raw scores were converted to equated scores through a conversion file developed using Microsoft Excel.

Test of Word Reading [TLP, Teste de Leitura de Palavras]

This test comprises four vertically scaled forms that assess word reading accuracy in Portuguese students of grades 1–4 (Chaves-Sousa, Santos, et al., 2017). Each form includes 30 items and students were asked to read aloud them, without time limits. Each word read accurately is scored as 1 (correct answer), and omissions, substitutions, additions and no response are scored as 0 (incorrect answer). The raw score is calculated by counting the number of words correctly read and is subsequently converted to a standardized score. The reliability coefficients for TLP test forms were high to very high. Correlation coefficients between the test forms of the TLP and other tests of word reading, ORF, reading comprehension and teachers' judgments were medium-to-large. Moderate-to-strong

correlations were also found between the scores in the TLP and the scores in vocabulary and digit span tests in all grades, excepting grade 1 (Chaves-Sousa, Ribeiro, et al., 2017).

Test of Reading Fluency [TFL, Teste de Fluência de Leitura]

This test assesses ORF in students from first to sixth grades (Ribeiro et al., 2014). Students were asked to read an unpublished text aloud during 3 min. The mean of the 3 min of reading was used to calculate the number of words read correctly per minute.

Test of Reading Comprehension of Narrative Texts [TCTML-n, Teste de Compreensão de Textos na Modalidade de Leitura-Narrativo]

It is composed of five vertically scaled forms that assesses reading comprehension of narrative texts in students from second to sixth grades (Rodrigues et al., 2020; Santos et al., 2016). Students were instructed to read silently a set of texts and answer a set of multiple-choice questions (three options) about them. The responses are scored as 0 (incorrect) and 1 (correct). Raw scores are computed by adding the number of correct responses, and are then converted to standardized scores. The reliability coefficients for TCTML-n test forms were moderate to very high. Regarding validity evidence, medium-to-large correlation coefficients were reported between the results in all test forms of the TCTML-n and those obtained in others reading tests and teachers' evaluations (Rodrigues et al., 2020; Santos et al., 2017).

Teachers' Evaluations about their Students' Reading Skills

These dimensions were assessed by asking teachers to rate decoding, ORF and reading comprehension for each of their students using a rating scale (1 = *very poor* to 5 = *excellent*). School outcome in the subject "Portuguese" and grade point average at the end of the academic year were also collected.

Procedures

Ethical procedures were similar to those used in study 1. The TLP was administered to students from grades 1 to 4, the TCTML-n to students from grades 2 to 6, whereas the TAF and the TFL to all students. The TCTML-n was administered collectively in classrooms, and the remaining tests individually in a separate room in the students' schools. The application of all tests was ensured by trained psychologists.

To test reliability, the TAF test forms were administered across two testing sessions separated by 2 weeks.

Statistical Analyses

Descriptive statistics and correlation coefficients between all variables were calculated using IBM® SPSS Statistics 27. The correlations' magnitude was evaluated using Cohen's (1992) recommendations: .10 suggests a small effect, .30 a medium effect and .50 a large effect. To investigate the reliability of the scores, besides computing correlation coefficients between the scores obtained at the two timepoints, paired-samples *t*-tests were also computed to analyse differences between the means of the equated scores in TAF from the first to the second timepoint. Cohen's *d* was used as a measure of effect size: values of 0.20, 0.50 and 0.80 indicate a small, medium or large effect, respectively (Cohen, 1992).

Results

Table 7 displays descriptive statistics, test-retest correlations and results of the paired-samples *t*-test for each test form of the TAF. Results from paired-samples *t*-tests showed a statistically significant increase in the means of all equated scores from timepoint 1 to timepoint 2. However, the effect size of the differences was small for all TAF test forms. Test-retest correlation coefficients for all TAF test forms were very high.

Regarding the relationship with other variables, medium-to-large correlation coefficients were found between all TAF test forms and external criteria, with exception of the correlation coefficient between TAF5-f1 and grade point average that was small-to-medium (see Tables 8 and 9). The correlations' magnitude between scores from the TAF and the TLP and the TCTML-n diminishes across grade levels, changing from a large effect for grades 1–3 to a medium effect for grades 4–6.

Table 7

Descriptive Statistics, Test-retest Correlations and Results of the Paired-samples t-test

Test form		T1		T2		*t* (df)	*d*	*r*
	n	*M*	*SD*	*M*	*SD*			
TAF1-f1	27	42.11	17.00	47.70	16.74	–5.86 (26)***	0.33	.957***
TAF2-f1	26	63.65	19.38	69.54	21.53	–4.56 (25)***	0.29	.954***
TAF2-f2	26	61.65	17.88	69.96	23.18	–5.76 (25)***	0.40	.969***
TAF2-f3	26	63.85	18.24	70.31	21.67	–4.61 (25)***	0.32	.950***
TAF3-f1	26	97.85	26.08	104.46	24.74	–5.45 (25)***	0.26	.972***
TAF3-f2	26	96.27	23.77	101.08	21.84	–3.27 (25)**	0.21	.950***
TAF3-f3	26	91.62	21.37	96.88	23.84	–4.21 (25)***	0.23	.966***
TAF4-f1	26	115.54	21.30	122.35	23.60	–4.43 (25)***	0.30	.944***
TAF4-f2	26	111.58	23.50	119.92	25.93	–6.41 (25)***	0.34	.969***
TAF4-f3	26	103.65	20.70	111.35	24.75	–4.28 (25)***	0.34	.934***
TAF5-f1	21	135.14	25.40	142.24	25.79	–4.82 (20)***	0.28	.965***
TAF5-f2	21	137.52	25.22	145.95	25.71	–4.23 (20)***	0.33	.936***
TAF6-f1	21	145.95	21.63	152.90	19.63	–4.68 (20)***	0.34	.950***
TAF6-f2	21	145.05	22.57	151.05	23.19	–3.11 (20)**	0.26	.926***

Note. M = Mean of the equated scores, *SD* = Standard deviation of the equated scores, df = Degrees of freedom, TAF1 = TAF for grade 1, TAF2 = TAF for grade 2, TAF3 = TAF for grade 3, TAF4 = TAF for grade 4, TAF5 = TAF for grade 5, TAF6 = TAF for grade 6, f1 = test form 1, f2 = test form 2, f3 = test form 3, T1 = Timepoint 1, T2 = Timepoint 2.

*** $p < .001$. ** $p < .01$.

Table 8

Descriptive Statistics and Correlations between TAF Test Forms and External Criteria from Grades 1 to 6

Test form	*n*	*M*	*SD*	TLP	TCTML-n	GPA	POR	DEC	ORF	RC
TAF1-f1	54	55.43	24.88	.728***	n.a.	.633**	.765***	.728***	.806***	.665***
TAF2-f1	61	75.13	25.57	.635***	.517***	.515***	.612***	.528***	.597***	.558***
TAF2-f2	61	75.36	26.62	.628***	.558***	.555***	.652***	.577***	.635***	.595***
TAF2-f3	61	74.54	24.66	.633***	.543***	.519***	.618***	.547***	.609***	.572***
TAF3-f1	57	99.60	26.40	.590***	.671***	.712***	.699***	.633***	.700***	.663***
TAF3-f2	57	98.42	24.65	.584***	.645***	.708***	.684***	.668***	.678***	.663***
TAF3-f3	57	92.40	23.88	.560***	.608***	.649***	.636***	.582***	.653***	.615***
TAF4-f1	51	124.12	24.72	.395**	.404**	.599***	.498**	.591***	.611***	.506***
TAF4-f2	51	120.76	28.30	.345*	.412**	.560***	.432*	.541***	.597***	.462***
TAF4-f3	51	113.59	27.57	.350*	.417**	.570***	.474**	.543***	.557***	.449***
TAF5-f1	94	139.02	22.95	n.a.	.455***	.280**	.432***	n.a.	.685***	.593***
TAF5-f2	94	142.65	23.40	n.a.	.473***	.328**	.455***	n.a.	.668***	.604***
TAF6-f1	85	151.21	21.38	n.a.	.443***	.490***	.514***	n.a.	.526***	.390***
TAF6-f2	85	148.05	23.29	n.a.	.462***	.494***	.511***	n.a.	.552***	.430***

Note. M = Mean of the equated scores, *SD* = Standard Deviation of the equated scores, TAF1 = TAF for grade 1, TAF2 = TAF for grade 2, TAF3 = TAF for grade 3, TAF4 = TAF for grade 4, TAF5 = TAF for grade 5, TAF6 = TAF for grade 6, f1 = test form 1, f2 = test form 2, f3 = test form 3, TLP = Test of Word Reading (standardized scores), TCTML-n = Test of Reading Comprehension of Narrative Texts (standardized scores), GPA = Grade Point Average, POR = classification in the subject "Portuguese", DEC = Decoding rated by teacher, ORF = Oral Reading Fluency rated by teacher, RC = Reading Comprehension rated by teacher, n.a. = not applicable.

***$p < .001$. **$p < .01$. *$p < .05$.

Table 9

Descriptive Statistics and Correlations between TAF Test Forms and TFL from Grades 1 to 6

Test form	*n*	*M*	*SD*	TFL
TAF1-f1	27	42.11	17.00	.969***
TAF2-f1	26	63.65	19.38	.960***
TAF2-f2	26	61.65	17.88	.973***
TAF2-f3	26	63.85	18.24	.942***
TAF3-f1	26	97.85	26.08	.964***
TAF3-f2	26	96.27	23.77	.954***
TAF3-f3	26	91.62	21.37	.952***
TAF4-f1	26	115.54	21.30	.927***
TAF4-f2	26	111.58	23.50	.934***
TAF4-f3	26	103.65	20.70	.912***
TAF5-f1	94	139.02	22.95	.923***
TAF5-f2	94	142.65	23.40	.901***
TAF6-f1	85	151.21	21.38	.899***
TAF6-f2	85	148.05	23.29	.902***

Note. M = Mean of the equated scores, *SD* = Standard Deviation of the equated scores, TAF1 = TAF for grade 1, TAF2 = TAF for grade 2, TAF3 = TAF for grade 3, TAF4 = TAF for grade 4, TAF5 = TAF for grade 5, TAF6 = TAF for grade 6, f1 = test form 1, f2 = test form 2, f3 = test form 3, TFL = Test of Reading Fluency.

*** $p < .001$.

Discussion and Conclusions

The main goal of this study was to develop and validate the TAF, an ORF measure with horizontally and vertically equated test forms for Portuguese elementary students.

In study 1, the TAF test forms showed statistically significant form effects suggesting that ORF scores of the students are influenced by test form that they are asked to read. These results showed that the use of grade-level texts, with control of their linguistic characteristics, is not sufficient for guaranteeing that they are equivalent, as reported in several previous research studies (e.g., Ardoin et al., 2010; Betts et al., 2009; Cummings et al., 2013; Francis et al., 2008). A formal statistical process for equating scores from different texts is mandatory to ensure that ORF score differences reflect only differences in student's ability (e.g., Albano & Rodriguez, 2012; Stoolmiller et al., 2013). Consequently,

a teacher, a psychologist or a researcher applying a set of equivalent texts at multiple timepoints in order to quantify the progress made by students may be confident that scores obtained from one time to the next not increased because students read easier texts, or decreased because students read more difficult texts, in determined assessments (Santi et al., 2016). Therefore, smoothed equipercentile equating was used to horizontally equate test forms of the TAF from each grade. Results from tests of differences showed that this equating procedure assured the scores' equivalence between test forms within each grade. Therefore, these forms can be used interchangeably. Considering NEAT design, vertical linking of the TAF using chained equipercentile method was conducted to yield scores, obtained from test forms of increasing difficulty between grade levels, on a common metric. As expected with vertically linked tests, the mean of equated scores increases from grades 1 to 6, meaning that the ORF skill improves as students advance in schooling. These results demonstrated the strengths of use of vertical equating in the development of ORF's measures, beyond horizontal equating statistical analysis frequently performed in a large number of previous studies (e.g., Albano & Rodriguez, 2012; Betts et al., 2009; Francis et al., 2008; Stoolmiller et al., 2013).

In study 2, test forms of the TAF demonstrated good reliability and validity evidence. As reported in others studies about reliability of the ORF tests (e.g., Center on Teaching and Learning, 2018; McGlinchey & Hixson, 2004), a very strong test-retest correlation was found for all test forms of the TAF. Results from paired-samples *t*-test showed statistically significant differences between the means of the equated scores obtained in the TAF test forms in the two timepoints, being higher at timepoint 2. Nonetheless, Cohen's *d* values ranged between 0.21 and 0.40, indicating that these differences are small-sized. Even so, it is important to reflect about the increase of number of words correctly read from the first measurement timepoint to the next. On the one side, this increase might reflect some learning effects, given than the students were exposed to the same test forms, in two testing sessions, separated by two weeks. On the other side, it is also probable that students might have simply experienced a growth in ORF, given that they underwent regular reading instruction during the two weeks that separated the two measurement timepoints.

Correlations of the TAF test forms with measures of word reading and reading comprehension were positive, of moderate-to-high magnitude, and statistically significant, consistently with previous research (Adlof et al., 2006; Cadime et al., 2017; Calet et al., 2015; Chaves-Sousa, Ribeiro, et al., 2017; Fernandes et al., 2017; Kim et al., 2012; Padeliadu & Antoniou, 2014; Sabatini et al., 2019; Santos et al., 2017; Shapiro et al., 2008; Silberglitt et al., 2006; Spear-Swerling, 2006; Speece et al., 2010; Valencia et al., 2010; Yovanoff et al., 2005). Furthermore, our results also suggest that the

relationship of the ORF with word reading and reading comprehension becomes weaker as the grade level increases. Similar results have been reported in prior research (Calet et al., 2015; Fernandes et al., 2017; Little et al., 2017; Padeliadu & Antoniou, 2014; Ribeiro et al., 2016; Yovanoff et al., 2005). As expected, the TAF test forms were also highly correlated with another test of ORF. Additionally, medium-to-large correlation coefficients with teachers' judgments and school outcomes were found for all test forms of the TAF (except for the TAF5-f1 in relation to grade point average in which correlation coefficient was slightly below the reference value .30 to be considered medium), as reported in other empirical studies (Begeny et al., 2008, 2011; Bigozzi et al., 2017; Feinberg & Shapiro, 2003, 2009).

The results of both studies provide empirical support for the use of this test to assess ORF in elementary Portuguese students.

The small sample size in study 2 is the main limitation of this work. Future validity studies may consider its replication with additional samples in order to confirm the reported relationships. Moreover, further studies are needed to provide additional validity evidence, such as predictive validity. For instance, it may be tested to what extent scores of the TAF test forms are predictors of students' results in the national exams of Portuguese.

Nonetheless, this work has significant implications for tests' development. In this study, we applied both horizontal and vertical equating procedures in the development of an ORF test, which constitutes a particular innovation, with potential transference for the construction of new tests that measure variables with similar characteristics to those of the ORF, i.e., competences that suffer major changes in time and that need to be monitored frequently. Moreover, the TAF can be an important tool for research in the Portuguese context, where there are no ORF tests that allow direct comparisons across time without repeating the reading of the same text. The TAF test forms may be used interchangeably within the same grade level, avoiding learning effects. Furthermore, it can be especially useful in longitudinal studies that cover ORF development across all or part of the first six years of schooling.

References

Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component? *Reading and Writing*, *19*(9), 933–958. https://doi.org/10.1007/s11145-006-9024-z

Albano, A. D. (2011). Statistical equating methods. *The R Project for Statistical Computing*. Retrieved April 1, 2021, from http://www2.uaem.mx/r-mirror/web/packages/equate/vignettes/equatevignette.pdf

Albano, A. D. (2016). equate: An R package for observed-score linking and equating. *Journal of Statistical Software*, *74*(8), 1–36. https://doi.org/10.18637/jss.v074.i08

Albano, A. D., & Rodriguez, M. C. (2012). Statistical equating with measures of oral reading fluency. *Journal of School Psychology*, *50*(1), 43–59. https://doi.org/10.1016/j.jsp.2011.07.002

Alonzo, J., Lai, C.-F., Anderson, D., Park, B. J., & Tindal, G. (2012). *An examination of test-retest, alternate form reliability, and generalizability theory study of the easyCBM passage reading fluency assessments: Grade 4 (Technical Report No. 1219)*. Retrieved July 5, 2021, from https://files.eric.ed.gov/fulltext/ED532323.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Anderson, D., Lai, C.-F., Park, B. J., Alonzo, J., & Tindal, G. (2012). *An examination of test-tetest, alternate form reliability, and generalizability theory study of the easyCBM reading assessments: Grade 2 (Technical Report No. 1217)*. Retrieved July 5, 2021, from https://files.eric.ed.gov/fulltext/ED532279.pdf

Anderson, D., Park, B. J., Lai, C.-F., Alonzo, J., & Tindal, G. (2012). *An examination of test-retest, alternate form reliability, and generalizability theory study of the easyCBM reading assessments: Grade 1 (Technical Report No. 1216)*. Retrieved July 5, 2021, from https://files.eric.ed.gov/fulltext/ED532278.pdf

Ardoin, S. P., & Christ, T. J. (2009). Curriculum-based measurement of oral reading: Standard errors associated with progress monitoring outcomes from DIBELS, AIMSweb, and an experimental passage set. *School Psychology Review*, *38*(2), 266–283. https://doi.org/10.1080/02796015.2009.12087837

Ardoin, S. P., Williams, J. C., Christ, T. J., Klubnik, C., & Wellborn, C. (2010). Examining readability estimates' predictions of students' oral reading rate: Spache, lexile, and forcast. *School Psychology Review*, *39*(2), 277–285. https://doi.org/10.1080/02796015.2010.12087778

Arnesen, A., Braeken, J., Baker, S. K., Meek-Hansen, W., Ogden, T., & Melby-Lervåg, M. (2017). Growth in oral reading fluency in a semitransparent orthography: Concurrent and predictive relations with reading proficiency in Norwegian, grades 2–5. *Reading Research Quarterly*, *52*(2), 177–201. https://doi.org/10.1002/rrq.159

Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, *23*(1), 43–55. https://doi.org/10.1037/1045-3830.23.1.43

Begeny, J. C., Krouse, H. E., Brown, K. G., & Mann, C. M. (2011). Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures. *School Psychology Review*, *40*(1), 23–38. https://doi.org/10.1080/02796015.2011.12087726

Betts, J., Pickart, M., & Heistad, D. (2009). An investigation of the psychometric evidence of CBM-R passage equivalence: Utility of readability statistics and equating for alternate forms. *Journal of School Psychology*, *47*(1), 1–17. https://doi.org/10.1016/j.jsp.2008.09.001

Bigozzi, L., Tarchi, C., Vagnoli, L., Valente, E., & Pinto, G. (2017). Reading fluency as a predictor of

school outcomes across grades 4-9. *Frontiers in Psychology*, *8*(200), 1–9. https://doi.org/10.3389/fpsyg.2017.00200

Cadime, I., Rodrigues, B., Santos, S., Viana, F. L., Chaves-Sousa, S., Cosme, M. C., & Ribeiro, I. (2017). The role of word recognition, oral reading fluency and listening comprehension in the simple view of reading: A study in an intermediate depth orthography. *Reading and Writing*, *30*(3), 591–611. https://doi.org/10.1007/s11145-016-9691-3

Calet, N., Gutiérrez-Palma, N., & Defior, S. (2015). A cross-sectional study of fluency and reading comprehension in Spanish primary school children. *Journal of Research in Reading*, *38*(3), 272–285. https://doi.org/10.1111/1467-9817.12019

Carvalho, A. (2010). *Teste de avaliação da fluência e precisão da leitura: O Rei [Test of reading fluency and accuracy: The King]*. Edipsico.

Center on Teaching and Learning (2018). *Understanding the research behind DIBELS® 8th Edition (Technical Report No. 1801)*. Retrieved July 6, 2021, from https://dibels.uoregon.edu/sites/dibels1.uoregon.edu/files/DIBELS8thEdition_TechRpt1801_ResearchBrief.pdf

Chaparro, E. A., Stoolmiller, M., Park, Y., Baker, S. K., Basaraba, D., Fien, H., & Smith, J. L. M. (2018). Evaluating passage and order effects of oral reading fluency passages in second grade: A partial replication. *Assessment for Effective Intervention*, *44*(1), 3–16. https://doi.org/10.1177/1534508417741128

Chaves-Sousa, S., Ribeiro, I., Viana, F. L., Vale, A. P., Santos, S., & Cadime, I. (2017). Validity evidence of the Test of Word Reading for Portuguese elementary students. *European Journal of Psychological Assessment*, *33*(6), 460–466. https://doi.org/10.1027/1015-5759/a000307

Chaves-Sousa, S., Santos, S., Viana, F. L., Vale, A. P., Cadime, I., Prieto, G., & Ribeiro, I. (2017). Development of a word reading test: Identifying students at-risk for reading problems. *Learning and Individual Differences*, *56*, 159–166. https://doi.org/10.1016/j.lindif.2016.11.008

Children's Learning Institute, & Texas Institute for Measurement, Evaluation, and Statistics (2010). *Technical report: Texas Primary Reading Inventory (2010–2014 Edition)*. Retrieved July 6, 2021, from https://www.tpri.org/

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cummings, K., Park, Y., & Schaper, H. A. B. (2013). Form effects on DIBELS Next oral reading fluency progress-monitoring passages. *Assessment for Effective Intervention*, *38*(2), 91–104. https://doi.org/10.1177/1534508412447010

de Ayala, R. J. (2009). *The theory and practice of item response theory*. The Guilford Press.

Decree-Law no. 54/2018 of July 6 (2018). *Diário Da República: I Série, n.º 129*, 2918–2928. Retrieved July 8, 2021, from https://dre.pt/dre/detalhe/decreto-lei/54-2018-115652961

Deeney, T. A. (2010). One-minute fluency measures: Mixed messages in assessment and instruction. *The Reading Teacher*, *63*(6), 440–450. https://doi.org/10.1598/RT.63.6.1

Denton, C. A., Barth, A. E., Fletcher, J. M., Wexler, J., Vaughn, S., Cirino, P. T., Romain, M., & Francis, D. J. (2011). The relations among oral and silent reading fluency and comprehension in middle school: Implications for identification and instruction of students with reading difficulties. *Scientific Studies of Reading*, *15*(2), 109–135. https://doi.org/10.1080/10888431003623546

Feinberg, A. B., & Shapiro, E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly*, *18*(1), 52–65. https://doi.org/10.1521/scpq.18.1.52.20876

Feinberg, A. B., & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research*,

102(6), 453–462. https://doi.org/10.3200/JOER.102.6.453-462

Fernandes, S., Querido, L., Verhaeghe, A., Marques, C., & Araújo, L. (2017). Reading development in European Portuguese: Relationships between oral reading fluency, vocabulary and reading comprehension. *Reading and Writing*, *30*(9), 1987–2007. https://doi.org/10.1007/s11145-017-9763-z

Francis, D. J., Barth, A., Cirino, P., Reed, D. K., & Fletcher, J. M. (2010). *Texas Middle School Fluency Assessment (version 2.0)*. University of Houston/Texas Education Agency.

Francis, D. J., Santi, K. L., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. R. (2008). Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology*, *46*(3), 315–342. https://doi.org/10.1016/j.jsp.2007.06.003

Good, R. H., & Kaminski, R. A. (2011). *DIBELS Next assessment manual*. Dynamic Measurement Group.

Hall, A. H., & Tannebaum, R. P. (2013). *Test Review: JL Wiederholt & BR Bryant.(2012). Gray Oral Reading Tests-Fifth Edition (GORT-5)*. Pro-Ed.

Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, *59*(7), 636–644. https://doi.org/10.1598/RT.59.7.3

Holland, P. W., Sinharay, S., von Davier, A. A., & Han, N. (2008). An approach to evaluating the missing data assumptions of the chain and post-stratification equating methods for the NEAT design. *Journal of Educational Measurement*, *45*(1), 17–43. https://doi.org/10.1111/j.1745-3984.2007.00050.x

Howe, K. B., & Shinn, M. M. (2002). *Standard Reading Assessment Passages (RAPs) for use in general outcome measurement: A manual describing development and technical features*. Edformation. Retrieved July 7, 2021, from http://www.aimsweb.com

Hudson, R. F., Lane, H. B., & Pullen, P. C. (2005). Reading fluency assessment and instruction: What, why, and how? *The Reading Teacher*, *58*(8), 702–714. https://doi.org/10.1598/rt.58.8.1

Kim, Y.-S. G., Wagner, R. K., & Lopez, D. (2012). Developmental relations between reading fluency and reading comprehension: A longitudinal study from grade 1 to grade 2. *Journal of Experimental Child Psychology*, *113*(1), 93–111. https://doi.org/10.1016/j.jecp.2012.03.002

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). Springer. https://doi.org/10.1007/978-1-4939-0317-7

Lai, C.-F., Alonzo, J., & Tindal, G. (2013). *easyCBM® reading criterion related validity evidence: Grades 2-5 (Technical Report No. 1310)*. Retrieved July 5, 2021, from https://files.eric.ed.gov/fulltext/ED545272.pdf

Lai, C.-F., Park, B. J., Anderson, D., Alonzo, J., & Tindal, G. (2012). *An examination of test-retest, alternate form reliability, and generalizability theory study of the easyCBM reading assessments: Grade 5 (Technical Report No. 1220)*. Retrieved July 5, 2021, from https://files.eric.ed.gov/fulltext/ED532327.pdf

Little, C. W., Hart, S. A., Quinn, J. M., Tucker-Drob, E. M., Taylor, J., & Schatschneider, C. (2017). Exploring the co-development of reading fluency and reading comprehension: A twin study. *Child Development*, *88*(3), 934–945. https://doi.org/10.1111/cdev.12670

Livingston, S. A. (2014). *Equating test scores (without IRT)* (2nd ed.). Educational Testing Service.

McCoach, D. B., & Yu, H. (2016). Using individual growth curves to model reading fluency. In K. Cummings & Y. Petscher (Eds.), *The fluency construct: Curriculum-based measurement concepts and applications* (pp. 269–308). Springer. https://doi.org/10.1007/978-1-4939-2803-3

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review*, *33*(2), 193–203. https://doi.org/10.1080/02796015.2004.12086242

National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4769)*. U.S. Government Printing Office.

Padeliadu, S., & Antoniou, F. (2014). The relationship between reading comprehension, decoding, and fluency in Greek: A cross-sectional study. *Reading & Writing Quarterly*, *30*(1), 1–31. https://doi.org/10.1080/10573569.2013.758932

Park, B. J., Anderson, D., Alonzo, J., Lai, C.-F., & Tindal, G. (2012). *An examination of test-retest, alternate form reliability, and generalizability theory study of the easyCBM word and passage reading fluency assessments: Grade 3 (Technical Report No. 1218)*. Retrieved July 5, 2021, from https://files.eric.ed.gov/fulltext/ED532326.pdf

Petscher, Y., & Kim, Y.-S. G. (2011). The utility and accuracy of oral reading fluency score types in predicting reading comprehension. *Journal of School Psychology*, *49*(1), 107–129. https://doi.org/10.1016/j.jsp.2010.09.004

Ribeiro, I., Cadime, I., Freitas, T., & Viana, F. L. (2016). Beyond word recognition, fluency, and vocabulary: The influence of reasoning on reading comprehension. *Australian Journal of Psychology*, *68*(2), 107–115. https://doi.org/10.1111/ajpy.12095

Ribeiro, I., Viana, F. L., Cadime, I., Chaves-Sousa, S., Santos, S., Silva, C., & Brandão, S. (2014). *Teste de Fluência da Leitura. Manual não publicado [The Test of Reading Fluency. Unpublished manual]*. Escola de Psicologia, Universidade do Minho.

Rodrigues, B., Cadime, I., Viana, F. L., & Ribeiro, I. (2020). Developing and validating tests of reading and listening comprehension for fifth and sixth grade students in Portugal. *Frontiers in Psychology*, *11*, Article ID 610876. https://doi.org/10.3389/fpsyg.2020.610876

Sabatini, J., Wang, Z., & O'Reilly, T. (2019). Relating reading comprehension to oral reading performance in the NAEP fourth-grade special study of oral reading. *Reading Research Quarterly*, *54*(2), 253–271. https://doi.org/10.1002/rrq.226

Salvia, J., Ysseldyke, J. E., & Bolt, S. (2010). Technical adequacy. In J. Salvia, J. E. Ysseldyke, & S. Bolt (Eds.), *Assessment in special and inclusive education* (11th ed., pp. 53–71). Wadsworth Cengage Learning.

Santi, K. L., Barr, C., Khalaf, S., & Francis, D. J. (2016). Different approaches to equating oral reading fluency passages. In K. Cummings & Y. Petscher (Eds.), *The fluency construct: Curriculum-based measurement concepts and applications* (pp. 223–265). Springer. https://doi.org/10.1007/978-1-4939-2803-3

Santos, S., Cadime, I., Viana, F. L., Chaves-Sousa, S., Gayo, E., Maia, J., & Ribeiro, I. (2017). Assessing reading comprehension with narrative and expository texts: Dimensionality and relationship with fluency, vocabulary and memory. *Scandinavian Journal of Psychology*, *58*(1), 1–8. https://doi.org/10.1111/sjop.12335

Santos, S., Cadime, I., Viana, F. L., Prieto, G., Chaves-Sousa, S., Spinillo, A. G., & Ribeiro, I. (2016). An application of the Rasch model to reading comprehension measurement. *Psicologia: Reflexão e Crítica*, *29*(38), 1–19. https://doi.org/10.1186/s41155-016-0044-6

Shapiro, E. S., Solari, E., & Petscher, Y. (2008). Use of a measure of reading comprehension to enhance prediction on the state high stakes assessment. *Learning and Individual Differences*, *18*(3), 316–328. https://doi.org/10.1016/j.lindif.2008.03.002

Silberglitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools*, *43*(5), 527–535. https://doi.org/10.1002/pits.20175

Sinharay, S., & Holland, P. W. (2010). A new approach to comparing several equating methods in the context of the NEAT design. *Journal of Educational Measurement*, *47*(3), 261–285.

https://doi.org/10.1111/j.1745-3984.2010.00113.x

Spear-Swerling, L. (2006). Children's reading comprehension and oral reading fluency in easy text. *Reading and Writing*, *19*(2), 199–220. https://doi.org/10.1007/s11145-005-4114-x

Speece, D. L., Ritchey, K. D., Silverman, R., Schatschneider, C., Walker, C. Y., & Andrusik, K. N. (2010). Identifying children in middle childhood who are at risk for reading problems. *School Psychology Review*, *39*(2), 258–276. https://doi.org/10.1080/02796015.2010.12087777

Stoolmiller, M., Biancarosa, G., & Fien, H. (2013). Measurement properties of DIBELS oral reading fluency in grade 2: Implications for equating studies. *Assessment for Effective Intervention*, *38*(2), 76–90. https://doi.org/10.1177/1534508412456729

Valencia, S. W., Smith, A. T., Reece, A. M., Li, M., Wixson, K. K., & Newman, H. (2010). Oral reading fluency assessment: Issues of construct, criterion, and consequential validity. *Reading Research Quarterly*, *45*(3), 270–291. https://doi.org/10.1598/rrq.45.3.1

Wang, T., Lee, W.-C., Brennan, R. L., & Kolen, M. J. (2008). A comparison of the frequency estimation and chained equipercentile methods under the common-item nonequivalent groups design. *Applied Psychological Measurement*, *32*(8), 632–651. https://doi.org/10.1177/0146621608314943

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice*, *24*(3), 4–12. https://doi.org/10.1111/j.1745-3992.2005.00014.x

CHAPTER 3

READING, LINGUISTIC, AND METACOGNITIVE SKILLS: ARE THE RELATIONS RECIPROCAL PAST THE FIRST SCHOOL YEARS?

Chapter 3

Reading, Linguistic, and Metacognitive Skills: Are the Relations Reciprocal Past the First School Years?[3]

Introduction

Decoding, oral reading fluency, listening comprehension, and vocabulary have been consistently identified as predictors of reading comprehension across a wide range of grade levels and orthographies (Cadime et al., 2017; Fernandes et al., 2017; Lervåg et al., 2018; Little et al., 2017; Muijselaar & de Jong, 2015; Ouellette & Beers, 2010; Padeliadu & Antoniou, 2014; Tilstra et al., 2009; Tobia & Bonifacci, 2015; Wolfgramm et al., 2016). However, the relative contributions of these skills to reading comprehension seem to be influenced by the phase of reading acquisition, as well as by the depth of the orthographies (e.g., Florit & Cain, 2011). As students advance from lower to higher grade levels, they are faced with more complex texts and greater demands for high-level reasoning and inferencing (Denton et al., 2011). Hence, beyond to read fluently and to have a broad and deep vocabulary knowledge and high levels of listening comprehension, it is increasingly necessary that students apply efficiently cognitive and metacognitive reading strategies in order to achieve better reading comprehension performance (Botsas, 2017; Muijselaar et al., 2017).

Developmental changes observed in the relations between reading comprehension and other reading-related skills, namely oral reading fluency, listening comprehension, vocabulary, and use of the reading strategies, justify their analysis past the early grade levels of reading development. Indeed, empirical studies are scarcer with samples of students beyond the fourth grade. This paper contributes to reduce this gap in the literature, addressing the longitudinal reciprocal interrelations between these skills in European Portuguese, an intermediate-depth orthography, using a sample of students assessed from grades 4 to 6.

Interrelations between Reading Skills

The relation between oral reading fluency and reading comprehension seems to be somewhat stronger at the early stages than at later stages of reading development (Benson, 2008; Kim & Wagner, 2015; Little et al., 2017; Padeliadu & Antoniou, 2014; Ribeiro et al., 2016; Silberglitt et al., 2006). Notwithstanding, previous research conducted in both transparent and opaque orthographies has shown that oral reading fluency continues to be a significant predictor of reading comprehension past the first four years of schooling (Fernandes et al., 2017; Padeliadu & Antoniou, 2014; Shapiro et al.,

[3] This work was submitted to the journal *Reading and Writing* in co-authorship with Iolanda Ribeiro and Irene Cadime.

2008; Valencia et al., 2010; Yildirim, Rasinski, et al., 2019; Yovanoff et al., 2005). Furthermore, results from other empirical studies suggested that reading comprehension also predicts oral reading fluency in samples of students from kindergarten to seventh grade across a wide range of orthographies such as English, Korean, Turkish and German, suggesting that there is a reciprocal relation between these skills (Berninger et al., 2010; Gebauer et al., 2013; Hudson et al., 2012; Jenkins et al., 2003; Kim, 2015; Klauda & Guthrie, 2008; Little et al., 2017; Yildirim, Ates, et al., 2019). In the case of European Portuguese, oral reading fluency and reading comprehension were reciprocally related between grades 2 and 3 but not between grades 3 and 4 (Santos et al., 2020). Highly efficient readers can direct their cognitive resources to developing complex comprehension skills (Baker et al., 2011; National Institute of Child Health and Human Development, 2000). In turn, good comprehenders are able to identify contextual cues that help to decipher the meanings of the phrases in a text, and they monitor efficiently the comprehension which allows fast correction of misread words during reading (Fuchs et al., 2001; Jenkins et al., 2003; Spear-Swerling, 2006).

Interrelations between Reading Comprehension and Linguistic Skills

Listening comprehension has been identified as a good predictor of reading comprehension (Cadime et al., 2017; Kim, 2015; Lervåg et al., 2018; Tilstra et al., 2009; Tobia & Bonifacci, 2015; Torppa et al., 2016). Nonetheless, listening comprehension seems to have a strong influence on reading comprehension since early stages of reading acquisition in transparent orthographies, but in deep orthographies, this effect is verified only in more advanced stages of reading acquisition (Florit & Cain, 2011). Moreover, some empirical studies have also suggested a reciprocal relation between these comprehension skills in orthographies of varying depth (Berninger & Abbott, 2010; Santos et al., 2020; Verhoeven & van Leeuwe, 2008, 2012; Wong, 2021). In a six-year cross-lagged panel study with Dutch students (grades 1-6), listening comprehension was reciprocally related to reading comprehension, only between grades 3 to 4, but not in the remaining grades (Verhoeven & van Leeuwe, 2008). Four years later, the same authors performed a longitudinal study in which data on listening comprehension were collected in grades 1, 3, and 5; and data on reading comprehension were collected in grades 2, 4, and 6. For first language learners (Dutch), listening comprehension in grades 1, 3 and 5 predicted reading comprehension in grades 2, 4, and 6, respectively, and reading comprehension in grades 2 and 4 predicted listening comprehension in grades 3 and 5, respectively (Verhoeven & van Leeuwe, 2012). Berninger and Abbott (2010) also observed reciprocal interrelations between these two comprehension skills in which they contributed unique variance to each other in grades 1, 3, 5 and 7. More recently,

cross-lagged relations were found between listening and reading comprehension in European Portuguese speakers in a longitudinal study involving grades 2 through 4 (Santos et al., 2020). Moreover, reciprocal relations between these two comprehensions skills were found in Chinese as a second-language, from grades 4 to 6 (Wong, 2021). Linguistic processes subjacent the two skills (e.g., processes related to phonology, lexis, grammar, syntax and semantics) may justify the bidirectionality of this relation (Hogan et al., 2014; Hoover & Tunmer, 2020; Kim & Phillips, 2014; Nation, 2005; Perfetti et al., 2005). Therefore, the development of these skills seems to be highly interdependent, regardless of the grade and the orthography.

Vocabulary is another linguistic skill that has been consistently found to account for unique reading comprehension variance, even after listening comprehension is statistically controlled, particularly for students past the early stages of reading acquisition (Braze et al., 2016; Goff et al., 2005; Muijselaar & de Jong, 2015; Ouellette & Beers, 2010; Protopapas et al., 2013; Seigneuric & Ehrlich, 2005; Tilstra et al., 2009; Tunmer & Chapman, 2012; Verhoeven & van Leeuwe, 2008). This finding is congruent with the lexical quality hypothesis, in which the quality of the words' representations, including a deep knowledge of their meanings, affects comprehension (Perfetti & Hart, 2002). Additionally, some previous studies have suggested that vocabulary depth measured by expressive measures may be more sensitive to specific aspects of linguistic skill than listening comprehension tests which are often influenced by memory (Ouellette & Beers, 2010; Protopapas et al., 2013; Tilstra et al., 2009). Results from prior research also showed that the additional contribution of vocabulary for reading comprehension seems to increase as students move from lower to upper grade levels, assuming a significant relative weight in grades 4, 5 and 6 both in opaque and more transparent orthographies (Fernandes et al., 2017; Ouellette & Beers, 2010; Swart et al., 2017; Tilstra et al., 2009; Wolfgramm et al., 2016; Yovanoff et al., 2005). Vocabulary has equally been recognized as a significant predictor of listening comprehension across the first six years of schooling, regardless of orthography depth (Hagtvet, 2003; Li et al., 2021; Verhoeven & van Leeuwe, 2008; Wolfgramm et al., 2016). Empirical results from some prior studies have also suggested that comprehension, whether this is assessed orally (listening comprehension) or using written material (reading comprehension), predicts vocabulary (Seigneuric & Ehrlich, 2005; Verhoeven et al., 2011; Verhoeven & van Leeuwe, 2008). Conceptually, vocabulary may exert a direct influence on listening comprehension and reading comprehension because the knowledge of word meanings simply provide the means to comprehend oral and written language (Nagy & Scott, 2000; Webb, 2021). In addition, students with better comprehension skills tend to be more able to use context to deduce meanings of new words or words

that assume different meanings depending on this same context, which contributes to the expansion and enrichment of their vocabulary (Mol & Bus, 2011).

Interrelations between Reading Comprehension and Metacognitive Skills

The efficient use of reading strategies develops as children get older and become skilled readers (Baker et al., 2015; Pintrich & Zusho, 2002; Roeschl-Heils et al., 2003). Most children begin to be able to use different reading strategies in order to solve breakdowns in comprehension between 8 and 10 years (McNamara et al., 2007; Veenman et al., 2006). Skilled readers use metacognitive strategies in order to identify when they are not understanding some part of the text and, consequently, mobilize appropriate cognitive strategies to overcome this difficulty (Graesser, 2007; Pereira-Laird & Deane, 1997; Perfetti et al., 2005). In fact, the knowledge and use of reading strategies have been appointed as significant predictors of reading comprehension (Law, 2009; Muijselaar & de Jong, 2015; Roeschl-Heils et al., 2003; Samuelstuen & Bråten, 2005; van Kraayenoord et al., 2012). However, contrarily to the reading and linguistic skills, the existence of reciprocal relations between metacognitive skills and reading comprehension has been scarcely explored. As far as we know, only one study conducted with a sample of Dutch students from the beginning of fourth grade to the end of fifth grade tested and found a reciprocal relation between knowledge of reading strategies and reading comprehension (Muijselaar et al., 2017). On the one side, students with better knowledge about and use of reading strategies tend to be more skilled in construction of the main ideas from texts. On the other side, it is also to be expected that reading comprehension contributes to development of the knowledge and use of reading strategies because students learn from texts (McMaster et al., 2014; Muijselaar et al., 2017; Verhoeven & Perfetti, 2008). Good comprehenders not only read more texts but also more challenging texts and, consequently, face comprehension gaps more frequently and have more opportunities for test the application of a wide range of reading strategies in order to restore a coherent representation of the text (Muijselaar et al., 2017). Hence, students may acquire more knowledge about reading strategies and use them more regularly as they advance from lower to upper grades of primary school and have to read texts of increasing difficulty (Clemens et al., 2017; McMaster et al., 2014).

The Present Study

Many empirical studies have investigated the directionality of the relations between oral reading fluency, listening comprehension, vocabulary, and reading comprehension across a wide range of

orthographies and grade levels, although studies in European Portuguese have focused only the initial years of reading acquisition (Santos et al., 2020). Research about reciprocal relations between reading comprehension and reading strategy use is particularly scarce. The goal of this study was to investigate the longitudinal relations between reading comprehension, oral reading fluency, listening comprehension, vocabulary, and reading strategy use reciprocal in European Portuguese from grades 4 to 6. Based on the literature review, our first hypothesis was that reading comprehension has a reciprocal relation with oral reading fluency, listening comprehension, and vocabulary across all the studied grade levels. Despite the existence of scarce empirical evidence, we also hypothesized a reciprocal relation between reading strategy use and reading comprehension, as observed in a study conducted with Dutch, an orthography of intermediate depth (Muijselaar et al., 2017).

Method

Participants

The initial sample was composed of 133 students assessed at the end of fourth grade. From fourth grade to fifth grade, 23 students dropped out of the study. Consequently, the final sample included 110 students who completed at least two assessment timepoints. The students attended public (n = 95, 86.4%) and private schools (n = 15, 13.6%) in northern Portugal. This distribution is representative of the population, given that, according to national data from the General Direction of Education and Science Statistics for the academic year 2019/2020, more than 85% of children from first to sixth grades attended public schools. The participants' mean age was 9.45 years (SD = 0.54, range 9–11) in fourth grade, 10.45 years (SD = 0.54, range 10–12) in fifth grade and 11.96 years (SD = 0.31, range 11–13) in sixth grade. More than half of the students were female (n = 58, 52.7 %). All students were fluent speakers of European Portuguese. Bilingual children and students who benefited from measures to support inclusion at secondary and/or tertiary level were not included in the study.

Regarding mothers' education levels, 35.5% had completed a university degree, 29.1% had completed high school, and 34.5% had a lower educational degree (missing information was obtained for 0.9% of the mothers). Because of the low income of their families, 39.1% of the students in fourth grade received social support from the government for the acquisition of meals, school supplies and study visits.

Measures

Test of Reading Comprehension of Narrative Texts (TRC-n; Rodrigues et al., 2020; Santos et al., 2016, 2017).

It is a norm-referenced test that includes five vertically scaled forms that assess reading comprehension of narrative texts in students from second to sixth grades. In the present study, the test forms for students in fourth, fifth and sixth grades were administered (TRC-n-4, TRC-n-5 and TRC-n-6). The student read silently the text passages that are followed by multiple-choice questions (three options) and mark the response on the answer sheet. The test is untimed and the responses are scored as 0 (incorrect) or 1 (correct). The total raw score of each test form is converted to a standardized score. The reliability coefficients ranged between .72 and .94 for the TRC-n-4, between .75 and .95 for the TRC-n-5 and between .72 and .97 for the TRC-n-6. Regarding validity, the TRC-n results were statistically correlated with scores in other tests of reading-related skills.

Test of Listening Comprehension of Narrative Texts (TLC-n; Rodrigues et al., 2020; Santos et al., 2015; Viana et al., 2015).

It is also a norm-referenced test that comprises six vertically scaled forms that measures listening comprehension of narrative texts across grades 1–6. In this study, the test form for fourth, fifth and sixth graders were administered (TLC-n-4, TLC-n-5, and TLC-n-6). The student listens to the text passages and to the multiple-choice questions (three options) and marks the chosen option in a worksheet. The test is untimed. The responses are scored as 0 (incorrect) and 1 (correct), and the total raw score of each test form is converted to a standardized score. The reliability coefficients ranged from .70 and .98 for the TLC-n-4, from .70 to .95 for the TLC-n-5 and from .70 to .97 for the TLC-n-6. With regard to validity evidence, statistically significant correlations were obtained between the TLC-n forms and external criteria measures.

Reading Strategy Use (RSU; Ribeiro et al., 2015).

It is a scale with 22 items, in which each one of them describes a reading strategy (cognitive or metacognitive), and the student marks the frequency of its use on a 7-point Likert scale ranging between 1 (*never*) and 7 (*always*). The adaptation and validation studies of the Portuguese RSU provided evidence for one-dimensional structure with a Cronbach's alpha of .85.

Fluency Assessment Test [TAF, *Teste de Avaliação da Fluência*] (Rodrigues et al., 2022).

The TAF evaluates oral reading fluency in students from grade 1 to grade 6. It is composed of specific test forms for each grade. Each one includes three unpublished texts (one narrative with dialogue, one narrative without dialogue and one expository). The students read aloud each text during

1 min. The number of words read correctly per minute in the three texts of each test form is averaged and then converted to an equated score. Regarding validity evidence, significant correlation coefficients were found between the TAF scores and the ones obtained in other reading tests, teachers' ratings, and school outcomes, as well as high test-retest reliability coefficients.

Vocabulary Subtest of the Wechsler Intelligence Scale for Children–III (WISC-III; Wechsler, 2003)

This test is composed of 30 items that demand that students provide orally the definition of a given word. The response to each item is scored with 0, 1 or 2 points. The test is ended after four consecutive failures. The raw score is computed by adding the points of each item and then is converted to a standardized score. Reliability coefficients for the Portuguese version of this subtest ranged from .69 to .89.

Procedure

Legal authorization for data collection was solicited from the ethics committee of the University of Minho and the Portuguese Ministry of Education, as well as from the school boards. Parents or legal tutors were informed about the study goals and signed an informed consent to allow the participation of the students in the study. The anonymity and confidentiality of the data was assured.

At each measurement timepoint, the TRC-n, the TLC-n and the RSU were administered collectively in the classroom and the remaining tests were applied to each student individually in a quiet room at schools. The tests administration was performed by trained psychologists at the end of the academic year for each grade level.

Statistical Analyses

Descriptive statistics (mean and standard deviation) and correlation coefficients between oral reading fluency (ORF), listening comprehension (LC), vocabulary (VOC), reading strategy use (RSU), and reading comprehension (RC) were calculated for each grade level.

A longitudinal cross-lagged panel model design was used to assess the longitudinal interrelations between the five variables. Four reciprocal-causation models with cross-lagged paths were tested using Mplus version 7 (Muthén & Muthén, 2012). The maximum likelihood estimator with robust standard errors (MLR) was applied.

Model 1 included ORF, RC, LC and VOC in order to examine the cross-lagged relations between ORF and RC, between RC and LC, and between LC and VOC. Given the poor fit of this first model, it

was decomposed into Model 2 and 3. Model 2 tested the reciprocal relations between ORF and RC, and between LC and RC. Model 3 was similar to model 2, but VOC replaced LC. Model 4 comprised ORF, RC and RSU, testing reciprocal relations between ORF and RC as well as between RC and RSU.

The following indexes were used to evaluate the model fit: chi-square value (χ^2), chi-square to degrees of freedom ratio (χ^2/df), comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). The four models were compared using the Akaike information criterion (AIC) and the Bayesian information criterion (BIC; Singer & Willet, 2003). Values of the ratio χ^2/df between 2.00 and 3.00 indicates a good model fit (Schreiber et al., 2006). The model fit is also considered acceptable when CFI and TLI values are higher than .90, and considered good when these indexes are higher than .95 (Byrne, 2011; Hu & Bentler, 1999). The RMSEA and the SRMR values below .08 and .10 (respectively) also indicate an acceptable model fit, whereas these values below .05 indicate a good fit (Browne & Cudeck, 1993; Schermelleh-Engel et al., 2003). Regarding the comparison indices, models with lower AIC and BIC values have better fit (Raftery, 1995). The significance level was 5% for all analysis.

Results

Table 10 displays the descriptive statistics and the correlation coefficients between all variables.

Statistically significant correlations were found between reading comprehension and the remaining variables at each timepoint, except for reading strategy use in grades 4 and 6 with reading comprehension in grades 5 and 6. The lowest correlation coefficients were the ones between reading strategy use and reading comprehension. The highest correlation coefficients were found between listening comprehension and reading comprehension, as well as between vocabulary and reading comprehension, in every grade. Table 11 presents the fit indexes for all cross-lagged models.

Table 10

Descriptive Statistics and Correlations for Oral Reading Fluency (ORF), Listening Comprehension (LC), Reading Comprehension (RC), Vocabulary (VOC), and Reading Strategy Use (RSU) in Grades 4 (G4), 5 (G5) and 6 (G6)

Variable	*N*	*M* (*SD*)	ORF_{G4}	ORF_{G5}	ORF_{G6}	LC_{G4}	LC_{G5}	LC_{G6}	RC_{G4}	RC_{G5}	RC_{G6}	VOC_{G4}	VOC_{G5}	VOC_{G6}	RSU_{G4}	RSU_{G5}	RSU_{G6}
ORF_{G4}	110	116.26 (27.57)	–														
ORF_{G5}	110	136.60 (25.70)	.92***	–													
ORF_{G6}	75	147.26 (22.66)	.87***	.93***	–												
LC_{G4}	110	120.21 (9.86)	.13	.15	.18	–											
LC_{G5}	110	124.17 (10.23)	.03	.11	.15	.49***	–										
LC_{G6}	75	128.45 (11.00)	.21†	.26*	.29*	.63***	.45***	–									
RC_{G4}	110	113.14 (13.52)	.35***	.36***	.40***	.51***	.39***	.45***	–								
RC_{G5}	110	112.85 (11.36)	.36***	.37***	.35**	.54***	.45***	.60***	.60***	–							
RC_{G6}	75	117.73 (10.40)	.39***	.37***	.38***	.58***	.48***	.68***	.58***	.63***	–						
VOC_{G4}	110	11.41 (2.97)	.20*	.21*	.27*	.47***	.44***	.51***	.49***	.42***	.49***	–					
VOC_{G5}	110	10.52 (2.59)	.20*	.21*	.23*	.38***	.40***	.40***	.50***	.51***	.51***	.71***	–				
VOC_{G6}	75	9.49 (2.40)	.34**	.32**	.31**	.57***	.38***	.53***	.59***	.61***	.51***	.60***	.65***	–			
RSU_{G4}	110	109.07 (19.43)	.17†	.17†	.21†	.12	.03	-.01	.22*	.08	.16	-.02	-.04	.13	–		
RSU_{G5}	110	105.11 (19.02)	.31***	.31**	.37**	.33***	.21*	.25*	.40***	.33***	.26*	.22*	.20*	.43***	.38***	–	
RSU_{G6}	75	103.91 (17.97)	.26*	.24*	.25*	.11	.28*	.09	.23*	.16	.15	.06	-.03	.09	.30**	.50***	–

Note. M = Mean; *SD* = Standard Deviation.

†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 11

Fit Indexes for the Cross-lagged Models

Models	$\chi^2(df)$	χ^2/df	CFI	TLI	RMSEA [90% CI]	SRMR	AIC	BIC
Model 1	68.368(28)***	2.442	.944	0.880	.114 [.080, .149]	.110	7,983.224	8,150.654
Model 2	24.200(13)*	1.862	.979	0.947	.089 [.028, .143]	.062	6,707.506	6,818.226
Model 3	14.070(13)	1.082	.998	0.995	.027 [.000, .100]	.050	5,877.602	5,988.321
Model 4	19.278(13)	1.483	.987	0.967	.066 [.000, .125]	.058	7,157.475	7,268.195

Note. χ^2 = chi-square value; *df* = degrees of freedom; χ^2/df = chi-square to degrees of freedom ratio; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion.

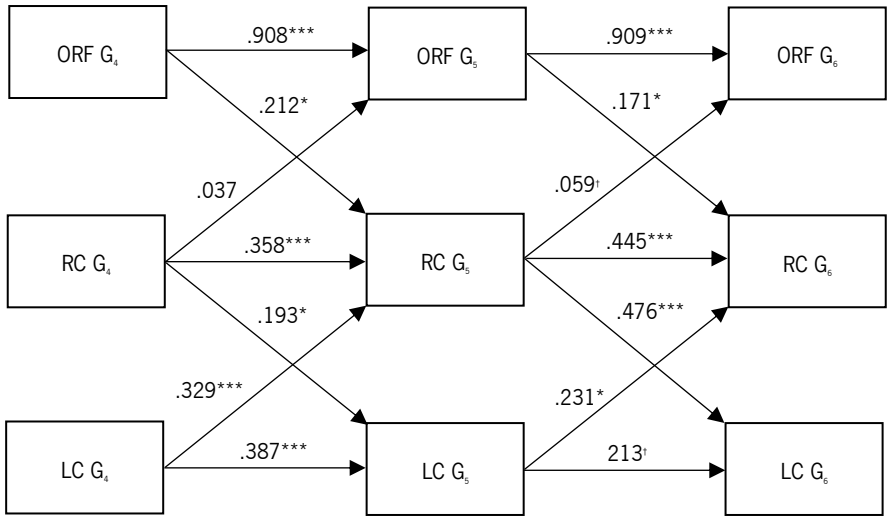
*$p < .05$. ***$p < .001$.

Model 1 presented a poor fit, given that only χ^2/df and CFI were within the acceptable range. Thus, this model was decomposed in the following two cross-lagged models.

The standardized regression paths for model 2 are depicted in Figure 4. This model presented a good fit. All fit indexes were within the acceptable range, with exception of RMSEA that was slightly above reference value. The relation between ORF and RC was found to be unidirectional at every measurement timepoint: RC was predicted by ORF, but the opposite was not confirmed. The cross-lagged paths between LC and RC indicate the existence of a reciprocal relation throughout grades 4–6.

Figure 4

Model 2: Test of the Reciprocal Relations between ORF and RC and between LC and RC



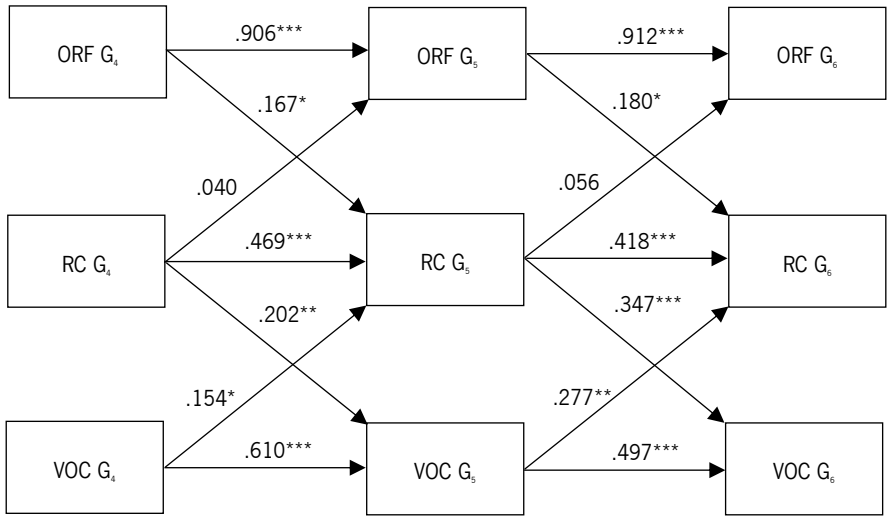
Note. The model included error covariances but they are not presented for a matter of parsimony. Standardized coefficients are depicted. G_4 = grade 4; G_5 = grade 5; G_6 = grade 6.

†$p < .10$. * $p < .05$. *** $p < .001$.

Model 3 fitted the data very well. The standardized regression paths for model 3 are depicted in Figure 5. The cross-lagged path between ORF and RC was similar to that obtained in the model 2. A reciprocal relation was found between VOC and RC in every grade.

Figure 5

Model 3: Test of the Reciprocal Relations between ORF and RC and between VOC and RC



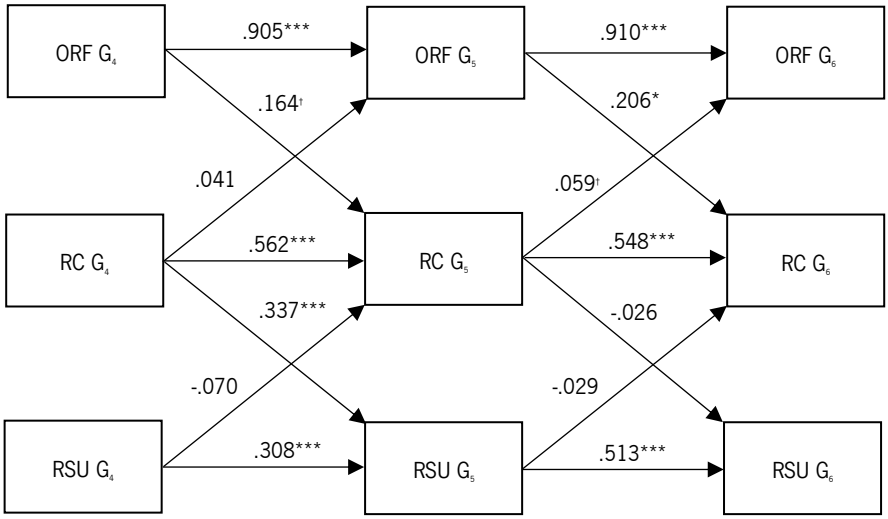
Note. The model included error covariances but they are not presented for a matter of parsimony. Standardized coefficients are depicted. G_4 = grade 4; G_5 = grade 5; G_6 = grade 6.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Model 4 presented a good fit. The standardized regression paths for model 4 are depicted in Figure 6. The relation between ORF and RC was similar to obtained in the previous models. Furthermore, RSU in grade 5 was predicted by RC in grade 4, but not the opposite.

Figure 6

Model 4: Test of the Reciprocal Relations between ORF and RC and between RSU and RC



Note. The model included error covariances but they are not presented for a matter of parsimony. Standardized coefficients are depicted. G_4 = grade 4; G_5 = grade 5; G_6 = grade 6.

†$p < .10$. *$p < .05$. ***$p < .001$.

Regarding model comparison, model 3 obtained the lowest AIC and BIC values, suggesting that it fits the data better than the remaining models.

Discussion

This study aimed to explore the longitudinal interrelations between reading comprehension, oral reading fluency, listening comprehension, vocabulary, and reading strategy use in European Portuguese from grades 4 to 6.

In order to test all our research hypotheses, we fitted four cross-lagged models. The first cross-lagged model tested the existence of reciprocal relations between reading comprehension, oral reading fluency, and listening comprehension, as well as between listening comprehension and vocabulary. Due to poor fit, this model was subsequently decomposed into two cross-lagged models, with oral reading fluency maintaining as predictor of reading comprehension across both models. Listening

comprehension and vocabulary were included as predictors of reading comprehension, in the second and third model, respectively. We hypothesized that reading comprehension would have a reciprocal relation with oral reading fluency, listening comprehension, and vocabulary. These reciprocal relations have been suggested in orthographies of varying depth (Berninger & Abbott, 2010; Klauda & Guthrie, 2008; Little et al., 2017; Santos et al., 2020; Seigneuric & Ehrlich, 2005; Verhoeven et al., 2011; Verhoeven & van Leeuwe, 2008, 2012; Wong, 2021). Finally, we fitted a fourth cross-lagged model with oral reading fluency, reading comprehension, and reading strategy use. Despite the existence of scarce empirical evidence, we hypothesized a reciprocal relation between reading strategy use and reading comprehension, as observed in a study conducted with Dutch, an orthography of intermediate depth (Muijselaar et al., 2017).

Congruent with results of previous research (e.g., Fernandes et al., 2017; Padeliadu & Antoniou, 2014; Yildirim, Rasinski, et al., 2019), oral reading fluency predicted reading comprehension at every measurement timepoint. As opposed to what we expected (e.g., Berninger et al., 2010; Little et al., 2017; Yildirim, Ates, et al., 2019), the opposite association was not observed. Research has suggested that the ability to extract meaning from text exerts an important influence in development of oral reading fluency, particularly when this last ability is not yet effortless and automatized (Jenkins et al., 2003). In this sense, it is plausible that this compensatory mechanism is no longer needed in the upper grades of primary school when students usually achieve a performance ceiling in oral reading fluency (Arnesen et al., 2017; Nese et al., 2012, 2013; Santos et al., 2020).

Consistently with previous research (Berninger & Abbott, 2010; Santos et al., 2020; Verhoeven & van Leeuwe, 2008, 2012; Wong, 2021), the hypothesis of a reciprocal relation between listening and reading comprehension was fully verified in model 2, suggesting that higher levels of one skill are associated to higher levels of the other skill and vice versa. Similar standardized coefficients were obtained in other studies conducted in intermediate-depth orthographies that examined the relation between these skills using cross-lagged models (Santos et al., 2020; Verhoeven & van Leeuwe, 2012).

The relation between vocabulary and reading comprehension was also reciprocal as observed in model 3, suggesting that the knowledge of word meanings and comprehension of written texts influence each other from the end of grade 4 to the end of grade 6. Once again, this finding is consistent with results from prior studies (Seigneuric & Ehrlich, 2005; Verhoeven et al., 2011; Verhoeven & van Leeuwe, 2008).

Contrary to what we hypothesized, reading strategy use was not reciprocally related to reading comprehension at any measurement timepoint, as indicated by model 4. Reading comprehension in

grade 4 predicted reading strategy use in grade 5, but this relation was not found from grades 5 to 6. Reading strategy use also did not predict reading comprehension at any measurement timepoint. This result differs from the one obtained by Muijselaar et al. (2017) in a study with Dutch students. These differences may be related to the measurement procedures. In the study by Muijselaar et al. (2017), a questionnaire about knowledge of reading strategies composed of items focused on what to do before, during, and after reading a text was used, whereas in our study, a self-report measure of the frequency of use of a number of reading strategies was applied. Furthermore, the significant impact of reading comprehension at the end grade 4 on reading strategy use in grade 5 may be related to school transition that occurs in the Portuguese educational system. In Portugal, the first cycle of study spins from grade 1 through 4 and the second cycle of study includes grades 5 and 6. During the first four years of formal reading instruction (grades 1–4), it is expected that students learn to read and write as well as acquire knowledge in three central curricular contents (Portuguese, Mathematics and Sciences) under the guidance of only one teacher. When students go to grade 5, students begin to have lessons related to about 10 subject areas (such as history and geography of Portugal, natural sciences, mathematics, visual education, technological education) and each one is taught by a different teacher. In this grade level, they are expected to comprehend an increasing amounts of expository texts with specific terminologies and concepts, but also poetic texts which are considerably distinctive of narrative and expository texts in terms of the structure (e.g., syntax) and content (Sanacore & Palumbo, 2008). This new school reality may lead students to face more comprehension gaps, and this is particularly challenging for children who do not have high comprehension skills and do not have mastered the use of a wide and diverse range of reading strategies (Gregg & Sekeres, 2006). Therefore, during grade 5, students may more frequently feel the necessity of apply different reading strategies in order to solve their comprehension difficulties, but only children with higher levels of reading comprehension are able to do that frequently. Future studies should address this issue.

Models 2, 3 and 4 presented good fits, contrarily to model 1. Model 3 presented a better fit than models 2 and 4, showing that reading comprehension performance is better explained by oral reading fluency and vocabulary than by oral reading fluency and listening comprehension or by oral reading fluency and reading strategy use. This finding corroborates the results obtained in prior research, appointing that expressive measures of vocabulary depth may be more robust indicators of students' linguistic skill than measures of listening comprehension (Ouellette & Beers, 2010; Protopapas et al., 2013; Tilstra et al., 2009).

The main limitation of our study is related to the type of measure used to assess reading strategy use: a self-report measure. It implies a particular disadvantage: the effect of social desirability on students' responses, i.e., students may have reported that they often use certain reading strategy, when in fact they do not. Consequently, an overestimation of reading strategy use may have occurred and contributed to the inexistence of statistical significance for the majority of paths in the cross-lagged models that contained reading strategy use. Thus, further research should consider other measures of metacognitive skills, such as tests in which students are required to apply specific reading strategies.

Regardless of these weaknesses, the present work helps to fill relevant research gaps with clear theoretical and practical implications. This study contributes to expand the scientific knowledge about the longitudinal interrelations between reading, linguistic and metacognitive skills, namely oral reading fluency, listening comprehension, vocabulary, reading strategy use, and reading comprehension in an intermediate-depth orthography and in more advanced grades of primary school. In addition to improving the state of the art, our results also open paths for the exploration of other research hypotheses in future studies. The findings of this study make it clear two central ideas: (1) oral reading fluency, listening comprehension, and vocabulary are significant predictors of reading comprehension, even after the first school years, and (2) better levels of reading comprehension have a positive impact on listening comprehension and vocabulary. Thereby, at a theoretical level, our results seem to provide empirical support to the following premise: many of the skills and processes involved in reading occur simultaneously, facilitating each other (Hoover & Tunmer, 2020). At a practical level, these results highlighted that, even for children in the upper grades of primary school oral reading fluency is still central for reading comprehension skills. So, it can be concluded that continued attention to the accurate and fast reading of connected texts throughout the primary school years is fundamental to avoid comprehension difficulties. Moreover, the expansion and deepening of knowledge about word meanings and improvement of comprehension skills with oral and written texts should be privileged in the school contexts as effective strategies to foster an articulated promotion of these reading skills in grades 5–6.

References

Arnesen, A., Braeken, J., Baker, S. K., Meek-Hansen, W., Ogden, T., & Melby-Lervåg, M. (2017). Growth in oral reading fluency in a semitransparent orthography: Concurrent and predictive relations with reading proficiency in Norwegian, grades 2–5. *Reading Research Quarterly*, *52*(2), 177–201. https://doi.org/10.1002/rrq.159

Baker, D. L., Stoolmiller, M., Good, R. H., & Baker, S. K. (2011). Effect of reading comprehension on passage fluency in Spanish and English for second-grade English learners. *School Psychology Review*, *40*(3), 331–351. https://doi.org/10.1080/02796015.2011.12087702

Baker, L., DeWyngaert, L. U., & Zeliger-Kandasamy, A. (2015). Metacognition in comprehension instruction: New directions. In S. R. Parris & K. Headley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 72–87). The Guilford Press.

Benson, N. (2008). Cattell-Horn-Carroll cognitive abilities and reading achievement. *Journal of Psychoeducational Assessment*, *26*(1), 27–41. https://doi.org/10.1177/0734282907301424

Berninger, V. W., & Abbott, R. D. (2010). Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language systems in grades 1, 3, 5, and 7. *Journal of Educational Psychology*, *102*(3), 635–651. https://doi.org/10.1037/a0019319

Berninger, V. W., Abbott, R. D., Trivedi, P., Olson, E., Gould, L., Hiramatsu, S., Holsinger, M., McShane, M., Murphy, H., Norton, J., Boyd, A. S., & Westhaggen, S. Y. (2010). Applying the multiple dimensions of reading fluency to assessment and instruction. *Journal of Psychoeducational Assessment*, *28*(1), 3–18. https://doi.org/10.1177/0734282909336083

Botsas, G. (2017). Differences in strategy use in the reading comprehension of narrative and science texts among students with and without learning disabilities. *Learning Disabilities: A Contemporary Journal*, *15*(1), 139–162. https://doi.org/10.4324/9780203876428

Braze, D., Katz, L., Magnuson, J. S., Mencl, W. E., Tabor, W., Van Dyke, J. A., Gong, T., Johns, C. L., & Shankweiler, D. P. (2016). Vocabulary does not complicate the simple view of reading. *Reading and Writing*, *29*(3), 435–451. https://doi.org/10.1007/s11145-015-9608-6

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

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

Cadime, I., Rodrigues, B., Santos, S., Viana, F. L., Chaves-Sousa, S., Cosme, M. C., & Ribeiro, I. (2017). The role of word recognition, oral reading fluency and listening comprehension in the simple view of reading: A study in an intermediate depth orthography. *Reading and Writing*, *30*(3), 591–611. https://doi.org/10.1007/s11145-016-9691-3

Clemens, N. H., Simmons, D., Simmons, L. E., Wang, H., & Kwok, O. (2017). The prevalence of reading fluency and vocabulary difficulties among adolescents struggling with reading comprehension. *Journal of Psychoeducational Assessment*, *35*(8), 785–798. https://doi.org/10.1177/0734282916662120

Denton, C. A., Barth, A. E., Fletcher, J. M., Wexler, J., Vaughn, S., Cirino, P. T., Romain, M., & Francis, D. J. (2011). The relations among oral and silent reading fluency and comprehension in middle school: Implications for identification and instruction of students with reading difficulties. *Scientific Studies of Reading*, *15*(2), 109–135. https://doi.org/10.1080/10888431003623546

Fernandes, S., Querido, L., Verhaeghe, A., Marques, C., & Araújo, L. (2017). Reading development in European Portuguese: Relationships between oral reading fluency, vocabulary and reading comprehension. *Reading and Writing*, *30*(9), 1987–2007. https://doi.org/10.1007/s11145-017-9763-z

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. https://doi.org/10.1007/s10648-011-9175-6

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 239–256. https://doi.org/10.1207/S1532799XSSR0503_3

Gebauer, S. K., Zaunbauer, A. C. M., & Möller, J. (2013). Cross-language transfer in English immersion programs in Germany: Reading comprehension and reading fluency. *Contemporary Educational Psychology*, *38*(1), 64–74. https://doi.org/10.1016/j.cedpsych.2012.09.002

Goff, D. A., Pratt, C., & Ong, B. (2005). The relations between children's reading comprehension, working memory, language skills and components of reading decoding in a normal sample. *Reading and Writing*, *18*(7), 583–616. https://doi.org/10.1007/s11145-004-7109-0

Graesser, A. C. (2007). An introduction to strategic reading comprehension. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 3–26). Erlbaum.

Gregg, M., & Sekeres, D. C. (2006). Supporting children's Reading of expository text in the geography classroom. *The Reading Teacher*, *60*(2), 102–110. https://doi.org/10.1598/rt.60.2.1

Hagtvet, B. E. (2003). Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic skills as well as phonological skills. *Reading and Writing*, *16*(6), 505–539. https://doi.org/10.1023/A:1025521722900

Hogan, T. P., Adlof, S. M., & Alonzo, C. N. (2014). On the importance of listening comprehension. *International Journal of Speech-Language Pathology*, *16*(3), 199–207. https://doi.org/10.3109/17549507.2014.904441

Hoover, W. A., & Tunmer, W. E. (2020). Summary of the Cognitive Foundations Framework. In *The cognitive foundations of reading and its acquisition: A framework with applications connecting teaching and learning* (pp. 85–88). Springer.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hudson, R. F., Torgesen, J. K., Lane, H. B., & Turner, S. J. (2012). Relations among reading skills and sub-skills and text-level reading proficiency in developing readers. *Reading and Writing*, *25*(2), 483–507. https://doi.org/10.1007/s11145-010-9283-6

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C. A., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology*, *95*(4), 719–729. https://doi.org/10.1037/0022-0663.95.4.719

Kim, Y.-S. G. (2015). Developmental, component-based model of reading fluency: An investigation of predictors of word-reading fluency, text-reading fluency, and reading comprehension. *Reading Research Quarterly*, *50*(4), 459–481. https://doi.org/10.1002/rrq.107

Kim, Y.-S. G., & Phillips, B. (2014). Cognitive correlates of listening comprehension. *Reading Research Quarterly*, *49*(3), 269–281. https://doi.org/10.1002/rrq.74

Kim, Y.-S. G., & Wagner, R. K. (2015). Text (oral) reading fluency as a construct in reading development: An investigation of its mediating role for children from grades 1 to 4. *Scientific Studies of Reading*, *19*(3), 224–242. https://doi.org/10.1080/10888438.2015.1007375

Klauda, S. L., & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology*, *100*(2), 310–321. https://doi.org/10.1037/0022-0663.100.2.310

Law, Y. K. (2009). The role of attribution beliefs, motivation and strategy use in Chinese fifth-graders' reading comprehension. *Educational Research*, *51*(1), 77–95.

https://doi.org/10.1080/00131880802704764
Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*, *89*(5), 1821–1838. https://doi.org/10.1111/cdev.12861
Li, L., Zhu, D., & Wu, X. (2021). The effects of vocabulary breadth and depth on reading comprehension in middle childhood: The mediator role of listening comprehension. *Reading and Writing Quarterly*, *37*(4), 336–347. https://doi.org/10.1080/10573569.2020.1809585
Little, C. W., Hart, S. A., Quinn, J. M., Tucker-Drob, E. M., Taylor, J., & Schatschneider, C. (2017). Exploring the co-development of reading fluency and reading comprehension: A twin study. *Child Development*, *88*(3), 934–945. https://doi.org/10.1111/cdev.12670
McMaster, K. L., Espin, C. A., & Van Den Broek, P. (2014). Making connections: Linking cognitive psychology and intervention research to improve comprehension of struggling readers. *Learning Disabilities Research and Practice*, *29*(1), 17–24. https://doi.org/10.1111/ldrp.12026
McNamara, D. S., Ozuru, Y., Best, R., & O'Reilly, T. (2007). The 4-pronged comprehension strategy framework. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 465–496). Erlbaum.
Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, *137*(2), 267–296. https://doi.org/10.1037/a0021890
Muijselaar, M. M. L., & de Jong, P. F. (2015). The effects of updating ability and knowledge of reading strategies on reading comprehension. *Learning and Individual Differences*, *43*, 111–117. https://doi.org/10.1016/j.lindif.2015.08.011
Muijselaar, M. M. L., Swart, N. M., Steenbeek-planting, E. G., Droop, M., Verhoeven, L., & de Jong, P. F. (2017). Developmental relations between reading comprehension and reading strategies. *Scientific Studies of Reading*, *21*(3), 194–209. https://doi.org/10.1080/10888438.2017.1278763
Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Muthén & Muthén.
Nagy, W. E., & Scott, J. (2000). Vocabulary processes. In M. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research, Vol. III* (pp. 338–355). Lawrence Erlbaum Associates, Inc.
Nation, K. (2005). Children's reading comprehension difficulties. In Margaret J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 248–265). Blackwell.
National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4769)*. U.S. Government Printing Office.
Nese, J. F. T., Biancarosa, G., Anderson, D., Lai, C. F., Alonzo, J., & Tindal, G. (2012). Within-year oral reading fluency with CBM: A comparison of models. *Reading and Writing*, *25*(4), 887–915. https://doi.org/10.1007/s11145-011-9304-0
Nese, J. F. T., Biancarosa, G., Cummings, K., Kennedy, P., Alonzo, J., & Tindal, G. (2013). In search of average growth: Describing within-year oral reading fluency growth across grades 1-8. *Journal of School Psychology*, *51*(5), 625–642. https://doi.org/10.1016/j.jsp.2013.05.006
Ouellette, G. P., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing*, *23*(2), 189–208. https://doi.org/10.1007/s11145-008-9159-1
Padeliadu, S., & Antoniou, F. (2014). The relationship between reading comprehension, decoding, and fluency in Greek: A cross-sectional study. *Reading & Writing Quarterly*, *30*(1), 1–31. https://doi.org/10.1080/10573569.2013.758932

Pereira-Laird, J. A., & Deane, F. P. (1997). Development and validation of a self-report measure of reading strategy use. *Reading Psychology*, *18*(3), 185–235. https://doi.org/10.1080/0270271970180301

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–214). John Benjamins Publishing Company.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Blackwell. https://doi.org/10.1002/9780470757642

Pintrich, P. R., & Zusho, A. (2002). The development of academic self-regulation: The role of cognitive and motivational factors. In A. Wigfield & J. Eccles (Eds.), *Development of Achievement Motivation* (pp. 249–284). Academic Press.

Protopapas, A., Mouzaki, A., Sideridis, G. D., Kotsolakou, A., & Simos, P. G. (2013). The role of vocabulary in the context of the simple view of reading. *Reading & Writing Quarterly, 29*, 168–202. https://doi.org/10.1080/10573569.2013.758569

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. https://doi.org/10.2307/ 271063

Ribeiro, I., Cadime, I., Freitas, T., & Viana, F. L. (2016). Beyond word recognition, fluency, and vocabulary: The influence of reasoning on reading comprehension. *Australian Journal of Psychology*, *68*(2), 107–115. https://doi.org/10.1111/ajpy.12095

Ribeiro, I., Dias, O., Oliveira, Í. M., Miranda, P., Ferreira, G., Saraiva, M., Paulo, R., & Cadime, I. (2015). Adaptação e Validação da Escala Reading Strategy use para a População Portuguesa. *Revista Iberoamericana de Diagnostico y Evaluacion Psicologica*, *1*(40), 25–36.

Rodrigues, B., Cadime, I., Freitas, T., Choupina, C., Baptista, A., Viana, F. L., & Ribeiro, I. (2022). Assessing oral reading fluency within and across grade levels: Development of equated test forms. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-022-01806-7

Rodrigues, B., Cadime, I., Viana, F. L., & Ribeiro, I. (2020). Developing and validating tests of reading and listening comprehension for fifth and sixth grade students in Portugal. *Frontiers in Psychology*, *11*, Article ID 610876. https://doi.org/10.3389/fpsyg.2020.610876

Roeschl-Heils, A., Schneider, W., & Van Kraayenoord, C. E. (2003). Reading, metacognition and motivation: A follow-up study of German students in grades 7 and 8. *European Journal of Psychology of Education*, *18*(1), 75–86. https://doi.org/10.1007/BF03173605

Samuelstuen, M. S., & Bråten, I. (2005). Decoding, knowledge, and strategies in comprehension of expository text. *Scandinavian Journal of Psychology*, *46*(2), 107–117. https://doi.org/10.1111/j.1467-9450.2005.00441.x

Sanacore, J., & Palumbo, A. (2008). Understanding the fourth-grade slump: Our point of view. *The Educational Forum*, *73*(1), 67–74. https://doi.org/10.1080/00131720802539648

Santos, S., Cadime, I., Viana, F. L., Chaves-Sousa, S., Gayo, E., Maia, J., & Ribeiro, I. (2017). Assessing reading comprehension with narrative and expository texts: Dimensionality and relationship with fluency, vocabulary and memory. *Scandinavian Journal of Psychology*, *58*(1), 1–8. https://doi.org/10.1111/sjop.12335

Santos, S., Cadime, I., Viana, F. L., Prieto, G., Chaves-Sousa, S., Spinillo, A. G., & Ribeiro, I. (2016). An application of the Rasch model to reading comprehension measurement. *Psicologia: Reflexão e Crítica*, *29*(38), 1–19. https://doi.org/10.1186/s41155-016-0044-6

Santos, S., Cadime, I., Viana, F. L., & Ribeiro, I. (2020). Cross-lagged relations among linguistic skills in European Portuguese: A longitudinal study. *Reading Research Quarterly*, *55*(2), 177–192. https://doi.org/10.1002/rrq.261

Santos, S., Viana, F. L., Ribeiro, I., Prieto, G., Brandão, S., & Cadime, I. (2015). Development of

listening comprehension tests with narrative and expository texts for Portuguese students. *Spanish Journal of Psychology*, *18*(e5), 1–7. https://doi.org/10.1017/sjp.2015.7

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74.

Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, *99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Seigneuric, A., & Ehrlich, M. F. (2005). Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. *Reading and Writing*, *18*(7), 617–656. https://doi.org/10.1007/s11145-005-2038-0

Shapiro, E. S., Solari, E., & Petscher, Y. (2008). Use of a measure of reading comprehension to enhance prediction on the state high stakes assessment. *Learning and Individual Differences*, *18*(3), 316–328. https://doi.org/10.1016/j.lindif.2008.03.002

Silberglitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools*, *43*(5), 527–535. https://doi.org/10.1002/pits.20175

Singer, J. D., & Willet, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* Oxford University Press.

Spear-Swerling, L. (2006). Children's reading comprehension and oral reading fluency in easy text. *Reading and Writing*, *19*(2), 199–220. https://doi.org/10.1007/s11145-005-4114-x

Swart, N. M., Muijselaar, M. M. L., Steenbeek-Planting, E. G., Droop, M., de Jong, P. F., & Verhoeven, L. (2017). Differential lexical predictors of reading comprehension in fourth graders. *Reading and Writing*, *30*(3), 489–507. https://doi.org/10.1007/s11145-016-9686-0

Tilstra, J., McMaster, K., Van den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: Components of the simple view of reading across grade levels. *Journal of Research in Reading*, *32*(4), 383–401. https://doi.org/10.1111/j.1467-9817.2009.01401.x

Tobia, V., & Bonifacci, P. (2015). The simple view of reading in a transparent orthography: The stronger role of oral comprehension. *Reading and Writing*, *28*(7), 939–957. https://doi.org/10.1007/s11145-015-9556-1

Torppa, M., Georgiou, G. K., Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., & Nurmi, J.-E. (2016). Examining the simple view of reading in a transparent orthography: A longitudinal study from kindergarten to grade 3. *International Journal of Special Education*, *30*(2), 70–84.

Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities*, *45*(5), 453–466. https://doi.org/10.1177/0022219411432685

Valencia, S. W., Smith, A. T., Reece, A. M., Li, M., Wixson, K. K., & Newman, H. (2010). Oral reading fluency assessment: Issues of construct, criterion, and consequential validity. *Reading Research Quarterly*, *45*(3), 270–291. https://doi.org/10.1598/rrq.45.3.1

van Kraayenoord, C. E., Beinicke, A., Schlagmüller, M., & Schneider, W. (2012). Word identification, metacognitive knowledge, motivation and reading comprehension: An Australian study of grade 3 and 4 pupils. *Australian Journal of Language and Literacy*, *35*(1), 51–68.

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, *1*(1), 3–14. https://doi.org/10.1007/s11409-006-6893-0

Verhoeven, L., & Perfetti, C. A. (2008). Advances in text comprehension: Model, process and development. *Applied Cognitive Psychology*, *22*, 293–301. https://doi.org/10.1002/acp.1417

Verhoeven, L., & van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A

longitudinal study. *Applied Cognitive Psychology*, *22*(3), 407–423. https://doi.org/10.1002/acp

Verhoeven, L., & van Leeuwe, J. (2012). The simple view of second language reading throughout the primary grades. *Reading and Writing*, *25*(8), 1805–1818. https://doi.org/10.1007/s11145-011-9346-3

Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading*, *15*(1), 8–25. https://doi.org/10.1080/10888438.2011.536125

Viana, F. L., Santos, S., Ribeiro, I., Chaves-Sousa, S., Brandão, S., Cadime, I., & Maia, J. (2015). Listening comprehension assessment: Validity studies of two vertically scaled tests for Portuguese students. *Universitas Psychologica*, *14*(1), 345–354. https://doi.org/10.11144/Javeriana.upsy14-1.lcav

Webb, S. (2021). Research investigating lexical coverage and lexical profiling: What we know, what we don't know, and what needs to be examined. *Reading in a Foreign Language*, *33*(2), 278–293. https://nflrc.hawaii.edu/rfl/

Wechsler, D. (2003). *Escala de Inteligência de Wechsler para Crianças – Terceira edição (WISC–III) [Wechsler Intelligence Scale for Children – Third Edition (WISC–III)]*. CEGOC-TEA.

Wolfgramm, C., Suter, N., & Göksel, E. (2016). Examining the role of concentration, vocabulary and self-concept in listening and reading comprehension. *International Journal of Listening*, *30*(1–2), 25–46. https://doi.org/10.1080/10904018.2015.1065746

Wong, Y. K. (2021). Developmental relations between listening and reading comprehension in young chinese language learners: A longitudinal study. *Journal of Psycholinguistic Research*, *50*(2), 261–273. https://doi.org/10.1007/s10936-018-9619-y

Yildirim, K., Ates, S., Çetinkaya, F. Ç., & Kaya Tosun, D. (2019). The relations between reading comprehension and reading fluency: Their reciprocal roles as an indicator and predictor. *Journal of Educational Sciences Research*, *9*(2), 67–81. https://doi.org/10.22521/jesr.2019.92.2

Yildirim, K., Rasinski, T., & Kaya, D. (2019). Fluency and comprehension of narrative texts in Turkish students in grades 4 through 8. *Education 3-13*, *47*(3), 348–357. https://doi.org/10.1080/03004279.2018.1449880

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice*, *24*(3), 4–12. https://doi.org/10.1111/j.1745-3992.2005.00014.x

CONCLUSION

Conclusion

The mastery of reading is undoubtedly a powerful tool for the development of each individual and, consequently, for the evolution of societies. For this reason, the assessment and promotion of reading skills throughout primary school supported by empirical evidence is systematically part of the (inter)national policy agenda. In this context, the present thesis aimed to contribute to fulfilling the need for formal tests of listening comprehension, reading comprehension, and oral reading fluency, that allow the direct comparison of intra and inter-individual differences (chapters 1 and 2). Moreover, the reciprocal relations between reading comprehension, listening comprehension, oral reading fluency, vocabulary, and metacognitive skills across grades 4 and 6 were analyzed (chapter 3).

In this conclusion, we will summarize the results obtained in the studies developed and described in each chapter, reflect on their limitations and point out guidelines for future studies. Considering that the articulation between research and practice is essential for the promotion of the instruction quality and students' learning, we will highlight the main contributions of this thesis for both fields. The approach to these issues is organized into two central topics – the development of reading tests and the interrelations between reading comprehension, listening comprehension, oral reading fluency, vocabulary, and metacognitive skills. Some final considerations will also be presented.

The Development of Reading Tests

The development of tests to assess reading skills (e.g., listening comprehension, reading comprehension, and oral reading fluency) implies specific challenges associated with assessing students at different time points, either within the same or across grade levels. Since it is expected that there is an increasing mastery of these skills over time (Arnesen et al., 2017; Oakhill et al., 2019), the use of standardized measures that allow a reliable interpretation of students' scores at different timepoints is crucial (Kolen & Brennan, 2014). These aspects sustain the main motivation for the construction of the listening and reading comprehension tests for fifth and sixth grade students (TLC-$n_{5/6}$ and TRC-$n_{5/6}$, respectively), as well as the fluency assessment test for students from first to sixth grades (TAF_{1-6}). Simultaneously, the recognized importance of listening comprehension and oral reading fluency for the reading comprehension performance in both the early and advanced school years (Cadime et al., 2017; Fernandes et al., 2017; Florit & Cain, 2011; Lervåg et al., 2018; Padeliadu & Antoniou, 2014; Tobia & Bonifacci, 2015) provided empirical support for the conceptual framework underlying the construction process of the aforementioned reading measures.

The construction and validation of the TLC-$n_{5/6}$ and TRC-$n_{5/6}$ aimed to follow up on the work developed within the scope of the Battery of Reading Assessment [BAL – *Bateria de Avaliação da Leitura*] (Ribeiro et al., 2014). Thus, we adopted the same methodological elements that had already been used in the development of the test forms for grades 1–4, namely: (a) the use of unpublished texts, (b) the use of multiple-choice items with three options, covering four comprehension levels: literal comprehension, inferential comprehension, critical comprehension and reorganization; and (c) the application of the Rasch model as a statistical procedure in order to perform the vertical equating of the test forms. The data from studies described in chapter 1 provided empirical evidence for the unidimensionality, local independence of the items, as well as for adequate reliability and validity of the listening comprehension and reading comprehension measures. Specifically, statistically significant correlations were found between both test forms of the TRC-n and the TLC-n, as well as between these test forms and vocabulary, working memory, reasoning skills, students' academic achievement and teacher ratings of students' reading skills. The TRC-n was significantly correlated with oral reading fluency in both grades but with use of reading strategies only in grade 5. Nevertheless, the collection of validity evidence should proceed in future studies. For example, it will be relevant to know the consequences of the administration of this type of tests in schools through the voice of different stakeholders (e.g., psychologists, students, teachers, parents, administrators and policymakers) and to analyze the classification accuracy indices of each test form in order to identify at-risk students for reading difficulties. Validity evidence may also be collected using the results obtained by students in the Portuguese national exams at the end of grade 8 and/or in the Portuguese end-of-cycle test at the end of grade 9 as external criteria to analyze the predictive properties of the TRC-$n_{5/6}$ and the TLC-$n_{5/6}$. Moreover, it would be pertinent to conduct additional differential item functioning analyses in order to understand to what extent the items are appropriate to assess minority groups. Future studies should also focus on the development of test forms using expository texts.

The development and validation process of the TAF_{1-6} followed a set of methodological and statistical recommendations anchored in previous empirical evidence, namely: (a) the use of distinct test forms, each consisting of three unpublished texts from the following typologies – narrative with dialogue, narrative without dialogue and expository (Briggs, 2011; Kolen & Brennan, 2014; Petscher & Kim, 2011); (b) the implementation of statistical procedures to perform horizontal and vertical equating to create comparable scores within each grade level and across different grade levels (Kolen & Brennan, 2014; Livingston, 2014). Although the texts were developed to be linguistically equivalent, the results described in chapter 2 revealed the existence of significant differences between the raw scores

obtained on the different TAF test forms within each grade level, suggesting that they could not be directly compared. It means that students' oral reading fluency scores depend on the test form that they read. To solve this issue, the following statistical equating procedures were applied: (1) equipercentile equating to make the scores, obtained from the test forms of each grade level, equivalent (horizontal equating); (2) chained equipercentile equating to place the scores, obtained from test forms of increasing difficulty across grade levels, on a common metric. Within-subjects difference tests with equated scores showed that the equating procedures ensured equivalence of scores between test forms within grade levels, so they could be used interchangeably. As expected with vertically linked test forms, between-subjects difference tests revealed that the mean of equated scores increases from first to sixth grades, indicating that the scores are able to capture the expected improvement of oral reading fluency as students advance from early to later grade levels. These results highlighted the potential advantages of the combined use of horizontal and vertical equating procedures in the development of oral reading fluency measures (Chaparro et al., 2018). The TAF test forms have shown good evidence of reliability and validity. Our finding that there was a very high test-retest correlation was found for all test forms of the TAF administered in two timepoints separated by two weeks. In addition, the results of the tests of difference showed that the means of the equated scores obtained in the TAF test forms at timepoint 2 were significantly higher than those obtained at timepoint 1. However, these differences were small-sized. Statistically significant correlations of the TAF test forms with other measures of oral reading fluency, word reading, and reading comprehension, as well as with teachers' ratings about decoding, oral reading fluency and reading comprehension skills of the students and their school outcomes, were found. Taken together, these results provide empirical support for the use of the TAF in the assessment of oral reading fluency in European Portuguese readers, from first to sixth grades. Further research should collect additional validity evidence, for example by testing to what extent the scores obtained in the TAF test forms predict students' results in the National Exams of Portuguese Language performed at different timepoints of school trajectory. Similar to what can be observed in other oral reading fluency tests (e.g., *DIBELS – Dynamic Indicators of Basic Early Literacy Skills;* University of Oregon, 2021), the standards definition for the interpretation of equated scores at different timepoints of each school year (for example, in December, April, and June that correspond to the last months of the first, second, and third term, respectively) may be a key aim in the subsequent studies. Specifically, the definition of four types of cut-off points (negligible risk, minimal risk, some risk, and at-risk) to identify students with different needs regarding oral reading fluency intervention (core support, strategic support, and intensive support) (University of Oregon, 2021) could be particularly

useful in terms of practical implications for the implementation of multilevel models in school settings. In addition, future research may also examine the classification accuracy of the TAF by performing ROC (Receiver Operating Characteristic) curve analysis, considering the area under the ROC curve and the sensitivity and specificity indices of this measure in each grade level.

In this thesis, oral reading fluency was assessed based on the number of words read correctly per minute. This is the most frequently used index in research as well as in practice (Beach & Traga Philippakos, 2021; Cruz et al., 2022; Fernandez & Arriondo, 2021; Kuhn & Schwanenflugel, 2019; Lee & Yoon, 2017; Washburn, 2022). To a large extent, this operationalization of the construct of oral reading fluency reflects the dominant forms of assessment and instruction of this skill conducted over the past 20 years (Kuhn & Schwanenflugel, 2019). Nonetheless, it should be noted that oral reading fluency is a multidimensional construct that also includes prosody which, in turn, is also related to comprehension (Calet et al., 2015; Spinillo et al., 2021; Zimmerman et al., 2019). In a meta-analysis conducted by Wolters et al. (2020), the overall average correlation between prosody and reading comprehension was .51 and no statistically significant differences in the magnitude of the correlations were found, as a function of grade level or orthographic depth. These results provide empirical evidence for the importance of assessing this dimension of oral reading fluency across different stages of reading development, both in opaque and transparent orthographies. Therefore, future studies may focus on the adaptation and validation of a measure to assess prosody in Portuguese students from first to grades.

The Interrelations between Reading Comprehension, Listening Comprehension, Oral Reading Fluency, Vocabulary, and Metacognitive Skills

Most empirical research on reading development has focused mainly on the early school grades, and there are fewer studies on the longitudinal predictors of reading skills in the more advanced grades (Psyridou et al., 2021). However, studies with samples of older students are crucial, both for research and practice, because reading development may not follow a linear development path across time. For example, the late emergence pattern (students whose reading learning difficulties were not noticeable in the first years of schooling, but are identified later, i.e., between fourth and eighth grades) and the late resolution pattern (students identified with reading learning difficulties in the early years of schooling, but who reverse this pattern of difficulties in later years) have been identified both in opaque and transparent orthographies (Catts et al., 2012; Psyridou et al., 2020; Torppa et al., 2015). This heterogeneity in the developmental trajectories justifies the need to analyze the reading comprehension

predictors over time with samples of students attending more advanced grades (e.g., 2nd and 3rd cycle of basic education). In the early stages of reading development, decoding (accuracy and speed of word reading) and reading comprehension are closely related skills: the comprehension of written texts can only occur when decoding is mastered (Florit & Cain, 2011). However, the strength of the relation between these two skills diminishes over time. This generally happens when students become sufficiently fluent and are then able to allocate more cognitive resources for the comprehension processes (García & Cain, 2014; Nation, 2019). As decoding skills gradually increase, linguistic skills such as listening comprehension and vocabulary become stronger predictors of reading comprehension (Torppa et al., 2016; Verhoeven & van Leeuwe, 2008). Likewise, students are also expected to become able to read and comprehend increasingly longer and harder texts (Clemens et al., 2017; Sanacore & Palumbo, 2008). Hence, the mastery and efficient use of a diverse range of reading strategies may have a significant impact on the resolution of comprehension gaps with which students are daily confronted in lessons (Duke & Cartwright, 2021). Therefore, the last goal of this work was to investigate the longitudinal interactions established between reading, linguistic, and metacognitive skills after the first school years (Berninger & Abbott, 2010; Kim, 2015b; Klauda & Guthrie, 2008; Little et al., 2017; Muijselaar et al., 2017; Santos et al., 2020; Seigneuric & Ehrlich, 2005; Verhoeven et al., 2011; Verhoeven & van Leeuwe, 2008, 2012). Specifically, the existence of reciprocal relations between reading comprehension, listening comprehension, oral reading fluency, vocabulary, and metacognitive skills in a semi-transparent orthography (European Portuguese) was analyzed, in a 3-year longitudinal study in which students were followed from fourth to sixth grades. Four reciprocal-causation models with cross-lagged paths were tested. The first cross-lagged model tested the reciprocal relations between reading comprehension and oral reading fluency, between reading comprehension and listening comprehension, as well as between listening comprehension and vocabulary. Due to its poor fit, this model was subsequently decomposed into two models, with oral reading fluency being maintained as a predictor of reading comprehension in both models. Listening comprehension and vocabulary were included as predictors of reading comprehension in the second and third models, respectively. Finally, a fourth model included oral reading, reading comprehension, and use of reading strategies, testing the existence of reciprocal relations between these last skills. From the analysis of the different cross-lagged models, it was possible to observe that model 3 was the one that had the best fit. These results suggest that, in European Portuguese, oral reading fluency and vocabulary seems to play a larger role in explaining individual differences in reading comprehension performance compared with listening comprehension and reading strategy use from fourth to sixth grades. Corroborating the

previous studies results (Berninger & Abbott, 2010; Santos et al., 2020; Seigneuric & Ehrlich, 2005; Verhoeven et al., 2011; Verhoeven & van Leeuwe, 2008, 2012; Wong, 2021), a reciprocal relation was found between listening comprehension and reading comprehension, as well as between vocabulary and reading comprehension. Contrary to what was hypothesized (Berninger et al., 2010; Gebauer et al., 2013; Hudson et al., 2012; Jenkins et al., 2003; Kim, 2015a; Klauda & Guthrie, 2008; Little et al., 2017; Muijselaar et al., 2017; Santos et al., 2020; Yildirim et al., 2019), oral reading fluency and reading strategy use were not reciprocally related to reading comprehension. Rather, oral reading fluency predicted reading comprehension in all measurement timepoints, but the opposite relation was not verified. Reading comprehension in fourth grade predicted reading strategy use in fifth grade, but the same was not found between fifth and sixth grades. The use of reading strategies also did not predict reading comprehension at either measurement timepoint.

These findings provide some empirical support for the assumptions underlying the Cognitive Foundations of Reading model (Hoover & Tunmer, 2020) and have clear implications for practice: (a) interventions to improve reading comprehension should be mainly targeted at promoting oral reading fluency, listening comprehension and vocabulary, and (b) improving reading comprehension may also positively influence the development of linguistic skills, namely listening comprehension and vocabulary. Therefore, the implementation of activities focused on training oral reading fluency, expanding semantic knowledge, and improving text comprehension skills in both modalities (oral and reading) should be privileged in school settings as an effective strategy to promote reading skills. Concerning promoting oral reading fluency, research has identified some effective strategies, such as (1) modeling students' reading through the teacher's reading; (2) practicing texts' reading through the method of assisted readings, readings in pairs, chorus readings, and repeated readings; (3) recording students' reading of texts and allowing them to listen to these recordings; (4) providing explicit instruction and immediate feedback in teaching decoding of unfamiliar and low-frequency words, (5) supplying materials appropriate to students' reading proficiency level for them to practice reading independently at school and/or at home, and (6) creating opportunities for the reading of progressively more difficult texts of different typologies (Beach & Traga Philippakos, 2021; Cruz et al., 2022; Gersten et al., 2020; Lee & Yoon, 2017; Martens et al., 2019; Zimmerman et al., 2019; Zimmermann et al., 2021). Since the vocabulary presented in textbooks may not be equally familiar to all students, it is recommended to develop activities before reading or listening of texts that include strategies such as activating prior knowledge and explicitly teaching the meaning of words in the text and concepts related to the topic (Foorman et al., 2016). It is also important to create opportunities for students to have contact with the

words in different situations, as well as to help them to use morphological and contextual clues to decipher word meanings (Christ & Wang, 2010; Cubukcu, 2008; Gomes-Koban et al., 2017; Wright & Cervetti, 2017). Another key aspect is to engage students in higher-level tasks since gains in comprehension are greater when students are confronted with cognitively challenging questions, e.g., questions from the inferential, critical, or reorganization comprehension levels (Cordeiro et al., 2021; Fuchs et al., 2008).

Social Impact

Research

During the last decades, one of the most common difficulties underlying reading research in Portugal has been the scarcity of validated tests with robust psychometric properties. This problem has been gradually been addressed through the development and validation of reading tests, but these were directed mainly at students attending the 1st cycle of basic education (Cadime et al., 2013; Carvalho, 2010; Ribeiro et al., 2014; Silva et al., 2020; Sucena & Castro, 2008, 2012). This thesis aims to contribute to fill this gap through the development of standardized tests to evaluate listening and reading comprehension in fifth and sixth grade students (TLC-$n_{5/6}$ and TRC-$n_{5/6}$) and oral reading fluency from first to sixth grades (TAF_{1-6}). The methodologies adopted in the construction of these tests allow the direct comparison of results obtained at different timepoints (both within the same grade levels and between different grade levels), opening new horizons for future research (cross-sectional studies but especially longitudinal studies) with samples of students from first to sixth grades.

Assessment and Intervention

The results of empirical research are unequivocal regarding the impact of the early identification of at-risk students and the consequent implementation of tailored interventions, in reducing the incidence or severity of students' difficulties (Catts et al., 2015). Multilevel models translate this preventive approach to practice (Catts et al., 2016). In Portugal, the Decree-Law no. 54/2018 of July 6 regulates the implementation of this approach and highlights not only the importance of universal screening for the signaling of at-risk students, but also monitoring of their learning over time. In the case of reading acquisition, despite the evident benefits of adopting this methodology in the early years, it is also important to consider the students results in more advanced grade levels, given that reading difficulties persist or can emerge after the 1st cycle of basic education (Rosa et al., 2020; Serrão et al., 2021). In this context, the tests developed in the scope of this thesis

may prove to be particularly promising tools for large-scale assessment. Both the TLC-$n_{5/6}$ and the TRC-$n_{5/6}$ may be administered collectively, allowing the assessment of a large number of students in a short period of time. Regarding the $TAF_{1-6,}$ although it has to be administered to students individually, it takes only about 4 min to administer to each student (1 min minute of reading per text, plus instructions). Due to the statistical procedures used in their construction, the TLC-$n_{5/6}$ and the TRC-$n_{5/6}$ (and the remaining test forms of the Battery of Reading Assessment) as well as the TAF_{1-6} also make it possible to compare results obtained at different time points along 1st and 2nd cycles of basic education. Based on the developmental perspective of reading proposed by Hoover and Tunmer (2020), the application of these tests allows the identification of students with different skill levels in listening comprehension, reading comprehension, and oral reading fluency. Thus, the TAF_{1-6}, the TLC-$n_{5/6}$, and the TRC-$n_{5/6}$ (as well as the TLC-n and TRC-n test forms of the Battery of Reading Assessment, developed for students from first/second to fourth grades) are a contribution to the differential diagnosis of reading learning difficulties throughout the first six grade levels.

The importance of assessing and monitoring student performance in oral reading fluency, listening, and reading comprehension is evident and goes beyond the diagnosis. The identification of specific reading learning difficulties should be the starting point for the definition of intervention plans within the framework of multilevel models. The adoption of these models is based on the assumption that each school should know the specific difficulties of its students and, consequently, allocate the necessary resources to reduce or eliminate those difficulties. Multilevel intervention presupposes the implementation of three types of measures – universal, selective, and additional – according to the needs of each student or group of students (Decree-Law no. 54/2018 of July 6, 2018). This conceptualization of the intervention embraces a paradigm shift since it alerts to the need of targeting educational support not only to students with significant difficulties but also to students who seem do not struggle with difficulties. Therefore, the identification of the above-mentioned profiles using the developed reading measures may support the choice of remedial strategies (selective or additional) for students with more difficulties, seeking to reverse the identified pattern of difficulties, but also of preventive strategies (universal) for students with intermediate performance in order to foster their reading achievement (Viana et al., 2019).

Final Considerations

As previously mentioned, the ability to read and comprehend is as powerful as it is complex, and its continuous assessment and promotion is considered imperative for individuals' success at

school and out of school (Cordeiro et al., 2021; Rios et al., 2020). The chapters that comprise this thesis provide not only robust measures for assessing listening comprehension, reading comprehension and oral reading fluency, but also a better understanding of the relations established between listening comprehension, oral reading fluency, vocabulary, metacognitive skills and reading comprehension in students between fourth and sixth grades. The developed tests allow an accurate assessment of three central learning skills, potentiating a better management of time and resources in schools, as well as the development of studies in the field of reading with samples of students attending 1st and 2nd cycles of basic education. Note that the TAF_{1-6} is a free access tool, so it may be easily used by researchers and practitioners. The study of the longitudinal and reciprocal relations between the aforementioned reading, linguistic, and metacognitive skills from fourth to sixth grades provides useful clues for the promotion of student learning in these school years, serving the main purpose of Applied Psychology research: the transference and the application of scientific knowledge into practice.

References

Arnesen, A., Braeken, J., Baker, S. K., Meek-Hansen, W., Ogden, T., & Melby-Lervåg, M. (2017). Growth in oral reading fluency in a semitransparent orthography: Concurrent and predictive relations with reading proficiency in Norwegian, grades 2–5. *Reading Research Quarterly*, *52*(2), 177–201. https://doi.org/10.1002/rrq.159

Beach, K. D., & Traga Philippakos, Z. A. (2021). Effects of a Summer reading intervention on the reading performance of elementary grade students from low-income families. *Reading and Writing Quarterly*, *37*(2), 169–189. https://doi.org/10.1080/10573569.2020.1760154

Berninger, V. W., & Abbott, R. D. (2010). Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language systems in grades 1, 3, 5, and 7. *Journal of Educational Psychology*, *102*(3), 635–651. https://doi.org/10.1037/a0019319

Berninger, V. W., Abbott, R. D., Trivedi, P., Olson, E., Gould, L., Hiramatsu, S., Holsinger, M., McShane, M., Murphy, H., Norton, J., Boyd, A. S., & Westhaggen, S. Y. (2010). Applying the multiple dimensions of reading fluency to assessment and instruction. *Journal of Psychoeducational Assessment*, *28*(1), 3–18. https://doi.org/10.1177/0734282909336083

Briggs, R. N. (2011). *Investigating variability in student performance on DIBELS Oral Reading Fluency third grade progress monitoring probes: Possible contributing factors (Publication No. 3466319) [Doctoral dissertation, University of Oregon]*. ProQuest Dissertations & Theses Global.

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*(3), 384–389. https://doi.org/10.7334/psicothema2012.288

Cadime, I., Rodrigues, B., Santos, S., Viana, F. L., Chaves-Sousa, S., Cosme, M. C., & Ribeiro, I. (2017). The role of word recognition, oral reading fluency and listening comprehension in the simple view of reading: A study in an intermediate depth orthography. *Reading and Writing*, *30*(3), 591–611. https://doi.org/10.1007/s11145-016-9691-3

Calet, N., Gutiérrez-Palma, N., & Defior, S. (2015). A cross-sectional study of fluency and reading comprehension in Spanish primary school children. *Journal of Research in Reading*, *38*(3), 272–285. https://doi.org/10.1111/1467-9817.12019

Carvalho, A. (2010). *Teste de avaliação da fluência e precisão da leitura: O Rei [Test of reading fluency and accuracy: The King]*. Edipsico.

Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology*, *104*(1), 166–181. https://doi.org/10.1037/a0025323

Catts, H. W., Nielsen, D. C., Bridges, M. S., & Liu, Y. S. (2016). Early identification of reading comprehension difficulties. *Journal of Learning Disabilities*, *49*(5), 451–465. https://doi.org/10.1177/0022219414556121

Catts, H. W., Nielsen, D. C., Bridges, M. S., Liu, Y. S., & Bontempo, D. E. (2015). Early identification of reading disabilities within an RTI framework. *Journal of Learning Disabilities*, *48*(3), 281–297. https://doi.org/10.1177/0022219413498115

Chaparro, E. A., Stoolmiller, M., Park, Y., Baker, S. K., Basaraba, D., Fien, H., & Smith, J. L. M. (2018). Evaluating passage and order effects of oral reading fluency passages in second grade: A partial replication. *Assessment for Effective Intervention*, *44*(1), 3–16. https://doi.org/10.1177/1534508417741128

Christ, T., & Wang, X. C. (2010). Bridging the vocabulary gap: What the research tells us about vocabulary instruction in early childhood. *Young Children*, *65*(4), 84–91.

Clemens, N. H., Simmons, D., Simmons, L. E., Wang, H., & Kwok, O. (2017). The prevalence of

reading fluency and vocabulary difficulties among adolescents struggling with reading comprehension. *Journal of Psychoeducational Assessment*, *35*(8), 785–798. https://doi.org/10.1177/0734282916662120

Cordeiro, C., Nunes, A., Castro, S. L., & Limpo, T. (2021). Teaching reading comprehension in Portuguese primary and middle schools. *Technology, Knowledge and Learning*, *26*(2), 339–354. https://doi.org/10.1007/s10758-020-09486-y

Cruz, J., Mendes, S. A., Marques, S., Alves, D., & Cadime, I. (2022). Face-to-face versus remote: Effects of an intervention in reading fluency during COVID-19 pandemic. *Frontiers in Education*, *6*, Article ID 817711. https://doi.org/10.3389/feduc.2021.817711

Cubukcu, F. (2008). Enhancing vocabulary development and reading comprehension through metacognitive strategies. *Issues in Educational Research*, *18*(1), 1–11.

Decree-Law no. 54/2018 of July 6 (2018). *Diário Da República: I Série, n.º 129*, 2918–2928. Retrieved July 8, 2021, from https://dre.pt/dre/detalhe/decreto-lei/54-2018-115652961

Duff, F. J., & Clarke, P. J. (2011). Practitioner review: Reading disorders: What are the effective interventions and how should they be implemented and evaluated? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *52*(1), 3–12. https://doi.org/10.1111/j.1469-7610.2010.02310.x

Duke, N. K., & Cartwright, K. B. (2021). The science of reading progresses: Communicating advances beyond the simple view of reading. *Reading Research Quarterly*, *56*, S25–S44. https://doi.org/10.1002/rrq.411

Fernandes, S., Querido, L., Verhaeghe, A., Marques, C., & Araújo, L. (2017). Reading development in European Portuguese: Relationships between oral reading fluency, vocabulary and reading comprehension. *Reading and Writing*, *30*(9), 1987–2007. https://doi.org/10.1007/s11145-017-9763-z

Fernandez, A. L., & Arriondo, G. J. (2021). Reading fluency as a measure of educational level. *Dementia e Neuropsychologia*, *15*(3), 361–365. https://doi.org/10.1590/1980-57642021dn15-030008

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. https://doi.org/10.1007/s10648-011-9175-6

Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C. A., Dimino, J., Furgeson, J., Hayes, L., Henke, J., Justice, L., Keating, B., Lewis, W., Sattar, S., Streke, A., Wagner, R., & Wissel, S. (2016). *Foundational skills to support reading for understanding in kindergarten through 3rd grade (NCEE 2016–4008)*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Fuchs, D., Fuchs, L. S., & Vaughn, S. (2008). *Response to intervention: A framework for reading educators.* International Reading Association.

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, *84*(1), 74–111. https://doi.org/10.3102/0034654313499616

Gebauer, S. K., Zaunbauer, A. C. M., & Möller, J. (2013). Cross-language transfer in English immersion programs in Germany: Reading comprehension and reading fluency. *Contemporary Educational Psychology*, *38*(1), 64–74. https://doi.org/10.1016/j.cedpsych.2012.09.002

Gersten, R., Haymond, K., Newman-Gonchar, R., Dimino, J., & Jayanthi, M. (2020). Meta-analysis of the impact of reading interventions for students in the primary grades. *Journal of Research on Educational Effectiveness*, *13*(2), 401–427. https://doi.org/10.1080/19345747.2019.1689591

Gomes-Koban, C., Simpson, I. C., Valle, A., & Defior, S. (2017). Oral vocabulary training program for Spanish third-graders with low socio-economic status: A randomized controlled trial. *PLoS ONE*, *12*(11), Article e0188157. https://doi.org/10.1371/journal.pone.0188157

Hoover, W. A., & Tunmer, W. E. (2020). *The cognitive foundations of reading and its acquisition: A framework with applications connecting teaching and learning*. Springer. https://doi.org/10.1007/978-3-030-44195-1

Hudson, R. F., Torgesen, J. K., Lane, H. B., & Turner, S. J. (2012). Relations among reading skills and sub-skills and text-level reading proficiency in developing readers. *Reading and Writing*, *25*(2), 483–507. https://doi.org/10.1007/s11145-010-9283-6

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C. A., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology*, *95*(4), 719–729. https://doi.org/10.1037/0022-0663.95.4.719

Kim, Y.-S. G. (2015a). Developmental, component-based model of reading fluency: An investigation of predictors of word-reading fluency, text-reading fluency, and reading comprehension. *Reading Research Quarterly*, *50*(4), 459–481. https://doi.org/10.1002/rrq.107

Kim, Y.-S. G. (2015b). Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. *Child Development*, *86*(1), 128–144. https://doi.org/10.1111/cdev.12293

Klauda, S. L., & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology*, *100*(2), 310–321. https://doi.org/10.1037/0022-0663.100.2.310

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). Springer. https://doi.org/10.1007/978-1-4939-0317-7

Kuhn, M. R., & Schwanenflugel, P. J. (2019). Prosody, pacing, and situational fluency (or why fluency matters for older readers). *Journal of Adolescent and Adult Literacy*, *62*(4), 363–368. https://doi.org/10.1002/jaal.867

Lee, J., & Yoon, S. Y. (2017). The effects of repeated reading on reading fluency for students with reading disabilities: A meta-analysis. *Journal of Learning Disabilities*, *50*(2), 213–224. https://doi.org/10.1177/0022219415605194

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*, *89*(5), 1821–1838. https://doi.org/10.1111/cdev.12861

Little, C. W., Hart, S. A., Quinn, J. M., Tucker-Drob, E. M., Taylor, J., & Schatschneider, C. (2017). Exploring the co-development of reading fluency and reading comprehension: A twin study. *Child Development*, *88*(3), 934–945. https://doi.org/10.1111/cdev.12670

Livingston, S. A. (2014). *Equating test scores (without IRT)* (2nd ed.). Educational Testing Service.

Martens, B. K., Young, N. D., Mullane, M. P., Baxter, E. L., Sallade, S. J., Kellen, D., Long, S. J., Sullivan, W. E., Womack, A. J., & Underberg, J. (2019). Effects of word overlap on generalized gains from a repeated readings intervention. *Journal of School Psychology*, *74*, 1–9. https://doi.org/10.1016/j.jsp.2019.05.002

Muijselaar, M. M. L., Swart, N. M., Steenbeek-planting, E. G., Droop, M., Verhoeven, L., & de Jong, P. F. (2017). Developmental relations between reading comprehension and reading strategies. *Scientific Studies of Reading*, *21*(3), 194–209. https://doi.org/10.1080/10888438.2017.1278763

Nation, K. (2019). Children's reading difficulties, language, and reflections on the simple view of reading. *Australian Journal of Learning Difficulties*, *24*(1), 47–73. https://doi.org/10.1080/19404158.2019.1609272

Oakhill, J., Cain, K., & Elbro, C. (2019). Reading comprehension and reading comprehension

difficulties. In D. A. Kilpatrick, R. M. Joshi, & R. K. Wagner (Eds.), *Reading development and difficulties: Bridging the gap between research and practice* (pp. 83–115). Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-26550-2_5

Padeliadu, S., & Antoniou, F. (2014). The relationship between reading comprehension, decoding, and fluency in Greek: A cross-sectional study. *Reading & Writing Quarterly*, *30*(1), 1–31. https://doi.org/10.1080/10573569.2013.758932

Petscher, Y., & Kim, Y.-S. G. (2011). The utility and accuracy of oral reading fluency score types in predicting reading comprehension. *Journal of School Psychology*, *49*(1), 107–129. https://doi.org/10.1016/j.jsp.2010.09.004

Psyridou, M., Tolvanen, A., de Jong, P. F., Lerkkanen, M. K., Poikkeus, A. M., & Torppa, M. (2021). Developmental profiles of reading fluency and reading comprehension from grades 1 to 9 and their early identification. *Developmental Psychology*, *57*(11), 1840–1854. https://doi.org/10.1037/dev0000976

Psyridou, M., Tolvanen, A., Lerkkanen, M. K., Poikkeus, A. M., & Torppa, M. (2020). Longitudinal stability of reading difficulties: Examining the effects of measurement error, cut-offs, and buffer zones in identification. *Frontiers in Psychology*, *10*, Article ID 2841. https://doi.org/10.3389/fpsyg.2019.02841

Ribeiro, I., Viana, F. L., Santos, S., Cadime, I., Chaves-Sousa, S., Vale, A. P., & Spinillo, A. G. (2014). *BAL – Bateria de Avaliação da Leitura: Abordagens teóricas e opções metodológicas [BAL – Battery of Reading Assessment: Theoretical approaches and methodological options]*. CEGOC-TEA.

Rios, J. A., Ling, G., Pugh, R., Becker, D., & Bacall, A. (2020). Identifying critical 21st-century skills for workplace success: A content analysis of job advertisements. *Educational Researcher*, *49*(2), 80–89. https://doi.org/10.3102/0013189X19890600

Rosa, V. A., Maia, J. S., Mascarenhas, D., & Teodoro, A. (2020). PISA, TIMSS e PIRLS em Portugal: Uma análise comparativa. *Revista Portuguesa de Educação*, *33*(1), 94–120. https://doi.org/10.21814/rpe.18380

Sanacore, J., & Palumbo, A. (2008). Understanding the fourth-grade slump: Our point of view. *The Educational Forum*, *73*(1), 67–74. https://doi.org/10.1080/00131720802539648

Santos, S., Cadime, I., Viana, F. L., & Ribeiro, I. (2020). Cross-lagged relations among linguistic skills in European Portuguese: A longitudinal study. *Reading Research Quarterly*, *55*(2), 177–192. https://doi.org/10.1002/rrq.261

Seigneuric, A., & Ehrlich, M. F. (2005). Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. *Reading and Writing*, *18*(7), 617–656. https://doi.org/10.1007/s11145-005-2038-0

Serrão, A., Simões, P., & Pires, R. (2021). *Estudo de aferição amostral do Ensino Básico 2021. Volume I – Resultados nacionais [Study of Basic Education 2021. Volume I – National results]*. Instituto de Avaliação Educativa, I.P. Retrieved December 28, 2021, from https://iave.pt/relatorios/

Silva, A. F., Marques, C., & Sucena, A. (2020). Validity evidence of the Reading Screening Test for Portuguese first graders. *Frontiers in Education*, *5*, Article ID 570639. https://doi.org/10.3389/feduc.2020.570639

Spencer, M., & Wagner, R. K. (2018). The comprehension problems of children with poor reading comprehension despite adequate decoding: A meta-analysis. *Review of Educational Research*, *88*(3), 366–400. https://doi.org/10.3102/0034654317749187

Spinillo, A. G., Paula, F. V., & Miller, M. T. A. B. (2021). Da relação entre prosódia e compreensão leitora: Considerações teóricas, metodológicas e controvérsias. *Psicologia USP*, *32*(e210047), 1–12. https://doi.org/10.1590/0103-6564e210047

Sucena, A., & Castro, S. L. (2008). *Aprender a ler e avaliar a leitura. O TIL: Teste de Idade de Leitura*

[Learning to read and assess reading. The TIL: Reading Age Test]. Almedina.

Sucena, A., & Castro, S. L. (2012). *ALEPE – Avaliação da leitura em Português Europeu [ALEPE – Reading assessment in European Portuguese]*. CEGOC-TEA.

Tobia, V., & Bonifacci, P. (2015). The simple view of reading in a transparent orthography: The stronger role of oral comprehension. *Reading and Writing*, *28*(7), 939–957. https://doi.org/10.1007/s11145-015-9556-1

Torppa, M., Eklund, K., van Bergen, E., & Lyytinen, H. (2015). Late-emerging and resolving dyslexia: A follow-up study from age 3 to 14. *Journal of Abnormal Child Psychology*, *43*(7), 1389–1401. https://doi.org/10.1007/s10802-015-0003-1

Torppa, M., Georgiou, G. K., Lerkkanen, M. K., Niemi, P., Poikkeus, A. M., & Nurmi, J. E. (2016). Examining the simple view of reading in a transparent orthography: A longitudinal study from kindergarten to grade 3. *Merrill-Palmer Quarterly*, *62*(2), 179–206. https://doi.org/10.13110/merrpalmquar1982.62.2.0179

University of Oregon (2021). *8th Edition of Dynamic Indicators of Basic Early Literacy Skills: Administration and scoring guide*. University of Oregon.

Verhoeven, L., & van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, *22*(3), 407–423. https://doi.org/10.1002/acp

Verhoeven, L., & van Leeuwe, J. (2012). The simple view of second language reading throughout the primary grades. *Reading and Writing*, *25*(8), 1805–1818. https://doi.org/10.1007/s11145-011-9346-3

Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading*, *15*(1), 8–25. https://doi.org/10.1080/10888438.2011.536125

Viana, F. L., Ribeiro, I., Santos, S., & Cadime, I. (2019). Ensinar a compreender. Perspetivas e propostas [Teaching to understand. Perspectives and proposals]. In S. R. K. Guimarães & F. V. Paula (Eds.), *Compreensão da Leitura: Processos cognitivos e estratégias de ensino* (1.ª edição, pp. 175–196). Vetor.

Washburn, J. (2022). Reviewing evidence on the relations between oral reading fluency and reading comprehension for adolescents. *Journal of Learning Disabilities*, *55*(1), 22–42. https://doi.org/10.1177/00222194211045122

Wolters, A. P., Kim, Y.-S. G., & Szura, J. W. (2020). Is reading prosody related to reading comprehension? A meta-analysis. *Scientific Studies of Reading*, *26*(1), 1–20. https://doi.org/10.1080/10888438.2020.1850733

Wong, Y. K. (2021). Developmental relations between listening and reading comprehension in young chinese language learners: A longitudinal study. *Journal of Psycholinguistic Research*, *50*(2), 261–273. https://doi.org/10.1007/s10936-018-9619-y

Wright, T. S., & Cervetti, G. N. (2017). A systematic review of the research on vocabulary instruction that limpacts text comprehension. *Reading Research Quarterly*, *52*(2), 203–226. https://doi.org/10.1002/rrq.163

Yildirim, K., Ates, S., Çetinkaya, F. Ç., & Kaya Tosun, D. (2019). The relations between reading comprehension and reading fluency: Their reciprocal roles as an indicator and predictor. *Journal of Educational Sciences Research*, *9*(2), 67–81. https://doi.org/10.22521/jesr.2019.92.2

Zimmerman, B. S., Rasinski, T. V., Was, C. A., Rawson, K. A., Dunlosky, J., Kruse, S. D., & Nikbakht, E. (2019). Enhancing outcomes for struggling readers: Empirical analysis of the fluency development lesson. *Reading Psychology*, *40*(1), 70–94. https://doi.org/10.1080/02702711.2018.1555365

Zimmermann, L. M., Reed, D. K., & Aloe, A. M. (2021). A meta-analysis of non-repetitive reading fluency interventions for students with reading difficulties. *Remedial and Special Education*, *42*(2), 78–93. https://doi.org/10.1177/0741932519855058

APPENDIXES

Appendix A. Notification of the Ethical Approval from the Portuguese General Direction of Education

De: mime-noreply@gepe.min-edu.pt [mime-noreply@gepe.min-edu.pt]

Enviado: 3 de Abril de 2018 16:13

Para: CIPSI; Maria Iolanda Ferreira Silva Ribeiro

Assunto: Monotorização de Inquéritos em Meio Escolar: Inquérito nº 0461200013

Exmo(a)s. Sr(a)s.

O pedido de autorização do inquérito n.º 0461200013, com a designação *Relações recíprocas entre a compreensão oral, a fluência de leitura, a metacompreensão, o vocabulário e a compreensão da leitura: Um estudo longitudinal com alunos do 4º ao 6º ano de escolaridade*, registado em 27-03-2018, foi aprovado.

Avaliação do inquérito:

> Exmo.(a) Senhor(a) Iolanda Ribeiro
>
> Venho por este meio informar que o pedido de realização de inquérito em meio escolar é autorizado uma vez que, submetido a análise, cumpre os requisitos, devendo atender-se às observações aduzidas.
>
> Com os melhores cumprimentos
>
> José Vítor Pedroso
>
> Diretor-Geral
>
> DGE

Observações:

a) A realização dos Inquéritos fica sujeita a autorização das Direções dos Agrupamentos de Escolas do ensino público a contactar para a realização do estudo. Merece especial atenção o modo, o momento e condições de aplicação dos instrumentos de recolha de dados em meio escolar, porque muito onerosos, devendo fazer-se em estreita articulação com as Direções dos Agrupamentos.

b) Deve considerar-se o disposto na Lei nº 67/98 em matéria de garantia de anonimato dos sujeitos, confidencialidade, proteção e segurança dos dados, (não identificar ou tornar identificável) sendo necessário solicitar o consentimento informado e esclarecido do titular dos dados, resultando obrigações que o responsável tem de cumprir. Destas deve dar

conhecimento a todos os que intervenham na recolha e tratamento de dados pessoais. No caso presente de recolha de dados de alunos menores (menos de 18 anos), inclusive de dados de avaliação escolar que constam do processo individual do aluno, este deverá ser atestado pelos seus representantes legais. As autorizações assinadas pelos Encarregados de Educação devem ficar em poder da Escola/Agrupamento ao qual pertencem os alunos. Não deve haver cruzamento ou associação de dados entre os que são recolhidos pelos instrumentos de inquirição e os constantes da declaração de consentimento informado.

c) Informa-se, ainda, que a DGE não é competente para autorizar a realização de estudos/aplicação de inquéritos ou outros instrumentos em estabelecimentos de ensino privados e para autorizar a realização de intervenções educativas/desenvolvimento de projetos e atividades/programas de intervenção/formação em meio escolar, dado ser competência da Escola/Agrupamento.

Pode consultar na Internet toda a informação referente a este pedido no endereço http://mime.gepe.min-edu.pt. Para tal terá de se autenticar fornecendo os dados de acesso da entidade.

Appendix B. Ethical Approval from the Ethics' Committee for the Research in Human and Social Sciences of the University of Minho

Universidade do Minho

Conselho de Ética

Conselho de Ética – Comissão de Ética para a Investigação em Ciências Sociais e Humanas

Identificação do documento: CEICSH 022/2019

Título do projeto: *Relações recíprocas entre a compreensão oral, a fluência de leitura, a metacompreensão, o vocabulário e a compreensão da leitura: um estudo longitudinal com alunos do 4.o ao 6.o ano de escolaridade*

Investigador(a) Responsável: Bruna Regina da Silva Rodrigues, aluna de Doutoramento em Psicologia Aplicada, Centro de Investigação em Psicologia, Escola de Psicologia da Universidade do Minho; Maria Iolanda Ferreira da Silva Ribeiro (Orientadora), Professora auxiliar na Escola de Psicologia da Universidade do Minho; Irene Maria Dias Cadime (Orientadora), Professora auxiliar convidada na Faculdade de Filosofa e Ciências Sociais da Universidade Católica Portuguesa, Membro integrado no Centro de Investigação em Estudos da Criança, Instituto de Educação, Universidade do Minho

PARECER

A Comissão de Ética para a Investigação em Ciências Sociais e Humanas (CEICSH) analisou o processo relativo ao projeto de investigação acima identificado, intitulado *Relações recíprocas entre a compreensão oral, a fluência de leitura, a metacompreensão, o vocabulário e a compreensão da leitura: um estudo longitudinal com alunos do 4.o ao 6.o ano de escolaridade.*

Os documentos apresentados revelam que o projeto obedece aos requisitos exigidos para as boas práticas na investigação com humanos, em conformidade com as normas nacionais e internacionais que regulam a investigação em Ciências Sociais e Humanas.

Face ao exposto, a CEICSH nada tem a opor à realização do projeto, emitindo o seu parecer favorável.

Braga, 3 de maio de 2019.

O Presidente da CEICSH

SIGN HERE

Assinado por : **ACÍLIO DA SILVA ESTANQUEIRO ROCHA**

Num. de Identificação Civil: BI042754054

Data: 2019.05.10 11:10:10 Hora de Verão de GMT

Anexo: Formulário de identificação e caracterização do projeto

CARTÃO DE CIDADÃO