Psicothema

Psicothema 2016, Vol. 28, No. 1, 89-95
doi: 10.7334/psicothema2015.92

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

Psychometric properties of the reading comprehension test ECOMPLEC.Sec

Ricardo Olmos Albacete[1], José Antonio León Cascón[1], Lorena Alicia Martín Arnal[1], José David Moreno Pérez[1], Inmaculada Escudero Domínguez[2] and Fernando Sánchez Sánchez[3]

[1] Universidad Autónoma de Madrid, [2] UNED and [3] TEA ediciones

Abstract

Background: ECOMPLEC.Sec is a reading comprehension test for secondary students, conceived from a multidimensional perspective in line with large-scale educational surveys such as PISA or PIRLS. The objective of this study was to validate the theoretical model of ECOMPLEC. Sec. A bifactor model that postulates the existence of a general reading comprehension factor and three specific factors provided a good fit to the data. **Method:** 1,912 adolescents (13-18 years) participated in this study. Data analysis included construct validity via confirmatory factor analysis, and factors were regressed onto observed covariates for the interpretation of the constructs. Reliability was calculated from a non-linear SEM in order to justify the interpretability of the observed scale and subscale scores. **Results:** The bifactor model exhibited a significantly better fit to the data than the second-order model. Furthermore, construct validity analysis suggests the existence of specific reading comprehension factors. Finally, the reliability study also supports the idea of using a total score to obtain a measure of reading comprehension. **Conclusions:** ECOMPLEC. Sec displays a valid parsimonious factor structure, as well as metric properties that make it a suitable tool to assess reading comprehension.

Keywords: psychometric properties, ECOMPLEC.Sec, reading comprehension test, bifactor model.

Resumen

Propiedades psicométricas de la prueba de compresión lectora ECOMPLEC.SEC. **Antecedentes:** ECOMPLEC.Sec es una prueba de comprensión lectora para estudiantes de Secundaria concebido desde una perspectiva multidimensional en consonancia con las pruebas educativas de gran escala como PISA o PIRLS. El objetivo de este estudio fue la validación del modelo teórico de ECOMPLEC.Sec. Un modelo bifactorial que presupone la existencia de un factor general de comprensión lectora y tres factores específicos ajustó adecuadamente a los datos. **Método:** 1.912 adolescentes (edades entre 13-18 años) participaron en este estudio. Los análisis estadísticos incluyen un análisis factorial confirmatorio cuyos factores se predicen por cuatro covariables con el fin de aportar significado a los constructos. La fiabilidad se abordó desde un modelo no lineal SEM para ayudar en la interpretación de las puntuaciones observadas de las escalas y subescalas. **Resultados:** el modelo bifactorial exhibió un ajuste significativamente mejor que el modelo factorial de segundo orden. Las evidencias de validez de constructo apuntan a la existencia de factores específicos de comprensión lectora. **Conclusiones:** ECOMPLEC.Sec muestra una estructura factorial parsimoniosa junto con unas propiedades psicométricas que hacen de ella una prueba adecuada para evaluar la comprensión lectora.

Palabras clave: propiedades psicométricas, ECOMPLEC.Sec, test de comprensión lectora, modelo bifactorial.

There is no doubt that reading is a crucial activity to acquire knowledge, for educational success, to access culture, and participate in society. This involves the recognition that written materials continue to be the main organized system for the transmission of knowledge. This crucial role should make us aware of the great importance of reading in education, society, culture, and work, as well as in personal growth (Vizcarro & León, 1998), to such an extent, that many developed and developing countries are spending much time and effort on research into this matter. A recent effort was the one performed by the PISA (*Programme for International Student Assessment*) and PIRLS (*Progress in International Reading Literacy Study*) projects (OECD, 2013; Rijmen, 2011), which have been applied since 2000 in all the OECD countries, seeking common criteria for the evaluation of reading competence.

Reading comprehension is a complex process in which readers must generate multiple inferences, add previous information to what is being read, and, among other things, integrate the new information with prior knowledge (Ozuru, Dempsey, & McNamara, 2009). Comprehension models agree that comprehension takes place when the reader constructs one or several mental representations of a text (e.g., Graesser, 2007; Kintsch, 1988; León, 2004; Zwaan & Singer, 2003). Many students have difficulties in correctly understanding the information they read, and it is crucial for these students to adequately develop their comprehension skills to function well in school and later join the work market (van den Broek & Espin, 2012). Thus, it

Received: April 9, 2015 • Accepted: November 18, 2015
Corresponding author: José Antonio León Cascón
Facultad de Psicología
Universidad Autónoma de Madrid
28049 Madrid (Spain)
e-mail: joseantonio.leon@uam.es

is currently even more necessary to create instruments that can assess differences in reading comprehension from a more integrated, modern position within current theories of reading comprehension, in order to detect problems in students and be able to intervene correctly and early on.

ECOMPLEC.Sec is inspired by the PISA and PIRLS international reading comprehension tests (OECD, 2013), and distinguishes between reading as a leisure activity, an activity for the acquisition of models, and an activity to find relevant information in the informational noise of the current world. ECOMPLEC.Sec (León, Escudero, & Olmos, 2012) comprises three types of text: a narrative text, an expository text, and a discontinuous text, multiple-choice tests are used to measure comprehension in each of the texts. Based on the situational model (Kintsch, 1988), the proposed questions distinguish between two types of mental representation which have been the subject of in-depth research (Graesser, Singer, & Trabasso, 1994; Kintsch, 1988; León & Escudero, 2015), text base and situation model. The first type involves a type of comprehension that is explicitly informed in the text, whereas the situation model involves comprehension with a deeper level of inference, it requires more extensive information integration and previous knowledge from the reader. Moreover, the contents included in the questions contain different types of knowledge that are typically included in every type of text, such as a goal-oriented and empathetic knowledge in narrative texts, conceptual and scientific knowledge in expository texts, or spatial knowledge in discontinuous texts (Green, 1995).

For individual differences in reading comprehension, a variety of components are postulated such as working memory, inference, mind wandering, prior knowledge or word recognition (Cromley, Snyder-Hogan, & Luciw-Dubas, 2010; Unsworth & McMillan, 2013). For example, McVay & Kane (2012) measured reading comprehension using different tasks (e.g., inferences, short texts, essays, verbal SAT scores). They found that relationships between these different tasks were explained by a unique and general factor. Other studies show that domain-specific factors such as interest and motivation for the topic have strong impact on reading comprehension (Hidi & Harackiewicz, 2000; Unsworth & McMillan, 2013). Anmarkrud and Braten's (2009) and Unsworth & McMillan's (2013) studies found that motivation factors contribute unique variance to comprehension scores, over and above what is explained by domain-general factors (such as working memory or attention control). These studies lead us to suspect that a bifactor structure might adequately fit the data (for a description of a bifactor models, see Chen, Hayes, Carver, Laureceau, & Zhang, 2012; Reise, Moore, & Haviland, 2012). If a bifactor model accounts for the structure of the data, then a general construct would directly affect each of the 68 items of the ECOMPLEC. Sec (domain-general factor). Furthermore, there would be specific factors for each type of text explaining idiosyncratic features or aspects of each text (specific motivation with a particular text, or different familiarity with a specific text topic).

The present study aims to analyze the factor structure of the ECOMPLEC.Sec. To this end, three measurement models were tested, each of them nested within a less restricted model. An one-dimensional model was tested first (Model A in Figure 1). A unidimensional model only takes into account the existence of a general factor and is not the theoretical model on which ECOMPLEC.Sec is based. Its fit was compared to the higher-order model (Model B in Figure 1). This model takes into account text-dependent features and better approaches the theoretical model, but presupposes that a general factor model does not directly affect the items. For this reason, the higher-order model was compared to the bifactor model (Model C in Figure 1). This is the model that best fits the theory on which ECOMPLEC.Sec is based. Then, given the factor structure found, an analysis was performed with several covariates to provide an interpretation of the factor structure and scores from the instrument. Finally, a reliability study was analyzed which made it possible to assess the appropriateness of using certain scores yielded by the test.

Method

Participants

In the present study, 1,942 adolescents (45.4% males) aged from 13 to 18 years were included. They represent all the ECOMPLEC administrations gathered during 2013 and 2014. Regarding educational level, 1,410 students were from the 2nd year and 530 were students from the 4th year of Secondary Education. From the entire sample, 30 students were discarded because they completed only one test. Regarding type of school, 34.6% of the participants came from concerted schools, 30.1% from private schools, 24.7% from public schools, and 10.6% from health/clinical centers. Concerning reading comprehension, 86% of the students were classified as "normal" students; 14% were classified as "suspected of reading comprehension problems". They came from different parts of Spain: Valencia (22.3%), Asturias (16.6%), Madrid (10.9%), Navarre (4.3%), Murcia (3.4%), Catalonia (2.8%), Cantabria (1.8%), and Basque Country (1.5%); and also from different countries: Guatemala (22.7%), Mexico (2.9%), Colombia (2.8%), Chile (2.4%), Argentina (1.8%), and Peru (1.5%), and other locations) less than 1%).

Instrument

ECOMPLEC.Sec is a reading comprehension test that includes three types of text — narrative, expository, and discontinuous — each of which comprises of the three main activities involved in reading: leisure, acquisition of knowledge, and search for information. The narrative text is by Julio Cortázar (1956), *Continuidad de los parques* (541 words) and included 25 question, multiple-choice test with three possible answers for each, as well as two metacognitive questions about the perceived difficulty of the text (difficult to understand, suitable, or easy). The expository text, *Los árboles estranguladores* (500 words), taken from an academic textbook, included a 23 question multiple-choice test as well as the same two metacognitive questions. Finally, the discontinuous text, *Ocio*, is a text about INJUVE data which displays graphs and figures as well as text pertaining to young Spaniards' leisure habits. It included 20 question multiple-choice text as well as the two metacognitive questions.

Procedure

Participants were administered ECOMPLEC.Sec in a class room or in the health/clinical center. They were first administered the narrative text (read the text and answer the questions), followed by the expository text and finished reading the discontinuous text. Each text takes approximately 20 minutes (León et al., 2012).

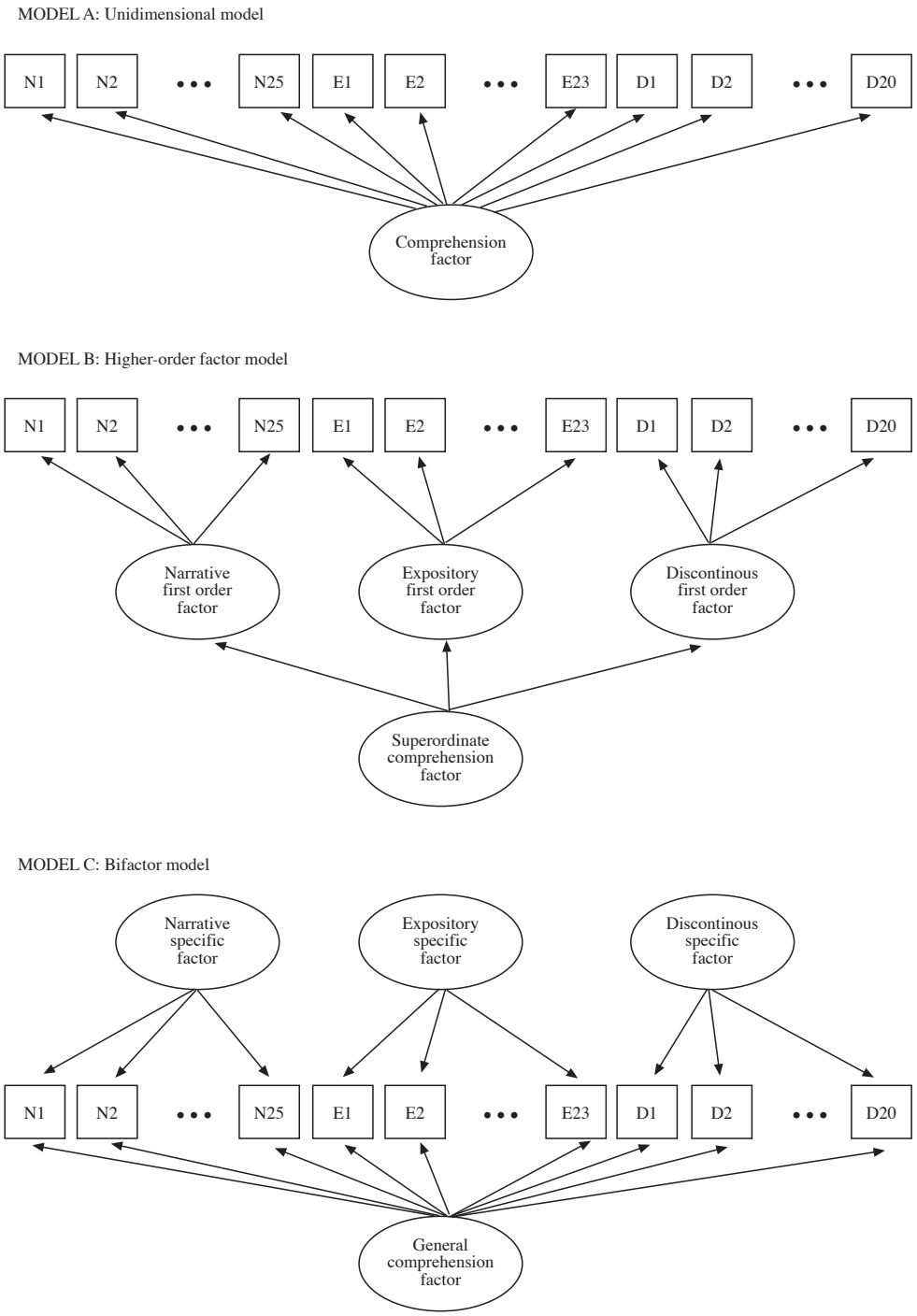

Figure 1. *Unidimensional, higher-order, and bifactor models*

Data analysis

The data was analyzed using the latent variable software Mplus 6.12 (Muthén & Muthén, 1998, 2010). Given that the observed indicators were categorical (an achievement test with binary data), a tetrachoric correlation matrix was the input matrix to perform all the factor models. The estimation parameter method was robust weighted least squares (WLSMV) (see Abad, Olea, Ponsoda, & García, 2011; Brown, 2006; Muthen & Muthen, 1998-2010). The goodness of fit indices used were the χ^2 test, the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI; Bentler, 1990) and the Tucker-Lewis Index (TLI; Tucker & Lewis, 1973). To compare statistically rival models (i.e. the nested model versus the parent model) we used the rescaled χ^2 difference test. The metacognitive questions, academic year, gender and normal/problem student was used as covariates in the factor model to better understand the general and specific factors.

Results

Factor structure of the instrument

It was fitted the three models presented in Figure 1 and compared them with rescaled chi-square different test ($\Delta\chi^2$). The results are shown in Table 1. For identification purposes, all factor variances were fixed to one. The unidimensional model degrades significantly the fit of the model with respect to higher-order model ($\Delta\chi^2$ (3) = 366.59, p<.001). The completely standardized loadings of the higher-order factor were .872 on narrative group factor, .926 on expository and .856 on discontinuous. However, the bifactor model represents a significant improvement over the higher-order model ($\Delta\chi^2$ (65) = 418.96, p<.001; ΔTLI > .01) and thus emerges as the preferred model. On the basis in all goodness of fit indices, it seems substantive (practical) the difference in fit between the two models (Gignac, 2007). In view of the data, the imposed restrictions on the higher-order model seemed inappropriate, and, consequently, the bifactor model was chosen. The rest of the validation study was conducted under the bifactor model.

Looking at completely standardized loadings (Table 2) from the bifactor model we observe the following: in the general comprehension the factor loadings of the 68 items were positive and statistically significant. It ranged from .217 (N15 = item 15 in narrative subscale) to .712 (E20 = item 20 in expository subscale). In the narrative specific factor all except two factor loadings were positive and statistically significant. In the expository specific factor the results were less clear. Four of the factor loadings were significant and negative. Five were not statistically significant. And 13 were positive and statistically significant, indicating both negative and positive conditional dependencies given the general comprehension factor. Finally, in the discontinuous specific factor, all factor loadings except one were positive and statistically significant.

Evidences of construct validity. Interpretability of the general and specific factors

One of the most important concerns with respect the bifactor analysis has to do with the interpretation of specific/group factors beyond the general factor (Chen et al., 2012; Reise et al., 2012). While the general factor has an immediate interpretation, that is, as a general reading comprehension factor, there is no obvious interpretation of the specific factors. Are they specific reading comprehension factors? Are they only nuisance (i.e. method) factors? To clarify this, the academic year (second and fourth grade), gender (male and female), metacognitive questions about perceived difficulty of each text (1 = difficult to understand; 2 = adequate; 3 = easy) and normal/suspect of comprehension problems were used as covariates to study possible differences in the latent means in the general and specific factors. The results showed that in the general reading comprehension factor the standardized difference between the second and fourth course was .727 (between a medium and large effect size in favor of fourth course). A small effect size was found (d = .191) between females and males (females had higher latent mean, p = .009). This result is consistent with Hyde's meta-analysis (1981) where he found a Cohen d = .24, revealing a small female advantage in verbal attitudes. The metacognitive question from the expository text showed a significant correlation (β = .288, p<.001) with the general factor (students who perceived the text as easy had higher general factor scores than those who perceived the text as difficult). Finally, students suspected of comprehension problems had lower factor scores than normal students (d = .376). In the specific narrative factor significant differences were found between fourth and second grades (d = .258, p = .012), between females and males (d = .252, p = .009), and between those suspected of

Table 1
Goodness of fit statistics for the three comparative models

Model	χ^2	df	$\Delta\chi^2$	Δdf	RMSEA [90% CI]	CFI	TLI
Undimensional model	4347.52 (p<.001)	2209			.023 [.022-.023]	.934	.932
Higher-order model	3633.84 (p<.001)	2206	366.59 (p<.001)	3	.018 [.017-.019]	.956	.955
Bifactor model	3194.59 (p<.001)	2141	418.96 (p<.001)	65	.016 [.015-.017]	.968	.966

Note: χ^2 = chi-square test, *df* = degrees of freedom, $\Delta\chi^2$ = chi-square difference test

Table 2
Factor solution of the incomplete bifactor model

	Narrative items		Expository items		Discontinuous items	
Item	StdF	StdN	StdF	StdE	StdF	StdD
1	.504	.270	.633	-.307	.467	.312
2	.516	.335	.659	-.110	.538	.308
3	.485	.161	.621	-.113	.439	.114
4	.387	.045	.520*	.473	.278	.200
5	.455	.235	.461	.519	.315	.158
6	.517	.167	.566	-.039	.309	.197
7	.351	.158	.308	.187	.572	.483
8	.480	.211	.405	.092	.326	.136
9	.484	.213	.625	-.070	.521	.443
10	.247	.201	.540	.088	.549	.221
11	.223	.377	.373	.197	.283	.245
12	.336	.307	.312	.092	.336	.180
13	.489	.108	.365	.212	.270	.235
14	.562	.443	.591*	.492	.419	.249
15	.217	.178	.551	-.011	.487	.437
16	.474	.256	.347	.396	.414	.267
17	.521	.067	.473	.178	.282	-.011
18	.489	.425	.318	.037	.527	.377
19	.377	.288	.528	-.081	.375	.170
20	.476	.216	.712	-.246	.565	.396
21	.604	.181	.506	.093		
22	.377	.357	.579	-.006		
23	.417	.324	.614	.375		
24	.308	.321				
25	.265	.294				

Note: StdF = completely standardized loading in the general comprehension factor; StdN = completely standardized loading in the specific narrative factor; StdE = completely standardized loading in the expository comprehension factor; StdD = completely standardized loading in the specific discontinuous factor. * Error variances corresponding to these indicators were correlated because de questions were the same (the alternatives were obviously different)

comprehension problems versus normal students (d = .389, p = .001). Also, it was a significant correlation with the metacognitive narrative question (β = .270, p<.001; those who perceived the text as easy scored higher than those who perceived the text as difficult in the specific narrative factor). In the specific discontinuous factor significant differences were found between second and fourth grades (d = .240, p<.032), between females and males (d = .229, p = .001) and between those suspected of comprehension problems versus normal students (d = .414, p = .001). It was a significant correlation with the metacognitive discontinuous question (β = .330, p<.001). Finally, in the specific expository factor significant differences were also found between fourth and second grades (d = .373, p<.001) and between females and males (d = .390, p<.001), but in this case in favor of males. No significant difference was found between students suspected of comprehension problems and the correlation with the metacognitive expository questions were not significant. The effect sizes and significant results found between the specific factors and the covariates suggest that the specific factors are idiosyncratic comprehension factors, instead of method or nuisance factors.

Given the exploratory nature of this analysis another bifactor analysis with course, sex, metacognitive questions and suspect/normal as covariates was run in a sample of 1,521 students from fourth and sixth grade of primary school (10 and 12 years old). ECOMPLEC.Pri was used (another version of ECOMPLEC to assess reading comprehension in younger students also using narrative, expository and discontinuous texts). The goodness of fit indices showed a good fit of the bifactor model to the data (χ^2(2137) = 2,902.42, p<.001; RMSEA = .015; CI 95% [.014 - .017]; CFI = .954; TLI = .951). As in the previous sample, significant differences and similar effect size in the general comprehension factor between sixth and fourth grades (d = .546, p<.001) and an effect size in favor of females (d = .212, p = .001) was found. The suspected of comprehension problems score lower in the general factor than normal students (d = .943, large effect). In the narrative and expository specific comprehension factors significant differences were found in favor of the sixth grade (d = .441 for narrative, d = .431 for expository). No differences were found in the discontinuous text between fourth and six grades. Males had higher latent means in the narrative (d = .218, p = .011), expository (d = .519, p<.001) and discontinuous (d = .196, p = .022) specific factors than females. The metacognitive narrative question showed a significant correlation with its own specific factor (β = .124, p = .006). The same result was found for the metacognitive expository question (β = .212, p<.001) and metacognitive discontinuous question (β = .252, p<.001). Finally, no differences were found between suspected of comprehension problem and normal students in the specific factors. Thus, results were partly replicated with this independent sample in younger students and this suggests that the specific factors are idiosyncratic comprehension factors. The main difference was that males scored higher in the three specific factors, while in secondary students females scored higher in the narrative and discontinuous specific factors.

Reliability

Cronbach's alpha coefficient was calculated for the entire test in the second (alpha = .894) and fourth (alpha = .858) grades. Thus, the internal consistency was satisfactory in both grades. Given the existence of three specific factors plus a general comprehension factor in the bifactor model, the reliability of each of the different factors was studied separately. For example, when a practitioner studies an examinee's score in the narrative subscale (the sum of the first 25 items), they might ask to what extent this score measures general reading comprehension, narrative specific comprehension and error measurement. From the bifactor model the different sources of reliability can easily be measured. It is worth noting that using Cronbach's alpha coefficient is not always the best reliability measure index (Bentler, 2009; Brown, 2006; Green & Yang, 2009; Raykov, 2001). Therefore, reliability was also calculated using the non-linear SEM model proposed in Green and Yang (2009), which yielded a reliability of .905 for the second grade and .884 for the fourth grade. Both reliabilities were calculated considering that the variance in the observed total scores was due to all sources of common variance (the general reading comprehension factor as well as the narrative, expository and discontinuous specific factors).

Why only interpreting the total score is recommended

As in the bifactor model it is easy to estimate the proportion of total observed variance that were due to each factor (the general and the specific factors), it is possible to judge the scale and subscale scores in terms of each factor contribution to the observed variance. For example, in fourth grade, the score from the narrative subscale (25 first items) gives a reliability of .781 (taking as common variance the general reading comprehension factor and the specific narrative factor variances), but only a reliability of .294 for narrative subscale after the general factor is controlled. Thus, it was found that a subscale score is not sufficiently reliable to interpret this score as specific comprehension factor because a subscale observed score is included in the general factor. Thus, given that the main objective of the instrument is to measure reading comprehension, we recommend that practitioners use the total observed score (the sum of the three text questions) only.

Discussion

The validation study of the ECOMPLEC.Sec yielded several important findings. First of all, the bifactor model fitted significantly and substantively better than the higher-order model. This is not only a trivial question of goodness of fit, but the evidences of construct validity and the reliability study are guided by this factor structure. The bifactor model suggests the existence of a general reading comprehension factor above and beyond all the idiosyncratic texts and questions that ECOMPLEC.Sec contains. As construct validity evidence, it was found that fourth grade had a higher latent mean (medium-large effect size) than second grade. Females also had a higher latent mean in this general comprehension factor (small effect size). Students with comprehension problems had lower latent mean than normal students (between small and medium effect size). Also, the validation study suggests the existence of three specific comprehension factors. The specific factors were regressed onto four observed covariates (academic level, gender, student suspected of comprehension problems and metacognitive questions). Significant and substantive effect sizes were found with these covariates (and the factor structure and regression results were partly replicated in an independent sample of 1,521 younger students). According to previous studies, these specific factors might be caused from several sources that are

text-dependent. The students can show different motivation and interest in each of the texts. There exists distinct structures and features in the three texts (e.g., the expository texts often have higher conceptual density, more technical concepts and more dependence on prior knowledge and vocabulary than narrative texts; Sáenz & Fuchs, 2002). Second, as Unsworth & McMillan (2013) have studied, domain-specific factors such as motivation or interest in the topic are independent from the general-factor domain (e.g., working memory) to predict reading comprehension. Hidi & Harackewicz (2000) also found that noncognitive factor that is text-dependent influence reading comprehension. Furthermore, ECOMPLEC.Sec showed to be a reliable instrument to measure reading comprehension. We recommend that practitioners use a total score from ECOMPLEC.Sec as a measure of reading comprehension.

Of course, these conclusions are only tentative. Future research is needed on the factor structure of ECOMPLEC.Sec in different samples in order to theoretically substantiate the existence of specific comprehension factors. It would be interesting to study if specific factors may predict beyond the general comprehension factor some external criteria (e.g., academic achievement problems, motivational problems, lower-level comprehension components, comprehension monitoring, etc). Also, metacognitive questions asking for the interest or motivation in the topic might help to understand the meaning of the specific factors. Although the bifactor model showed better fit than the higher-order model, this usually happens where there are slight miss-specifications in the model (e.g., ignoring the existence of cross-loadings or error covariances; see Murray & Johnson, 2013). It is not clear which of the two, bifactor or higher-order models, better describes human cognitive abilities, and this is not a question of statistical relative fit, but rather of substantiating theories. ECOMPLEC.Sec is not an exception. Thus, as this is the first approximation of ECOMPLEC. Sec validation, other studies will provide more evidence to improve this study.

Acknowledgements

This work has been funded by MINNECO PSI2013- 47219-P project.

References

Abad, F., Olea, J., Ponsoda, V., & García, C. (2011). *Medición en Ciencias Sociales y de la Salud* [Measurement in social and health sciences]. Madrid: Editorial Síntesis.

Anmarkrud, O., & Braten, I. (2009). Motivation for reading comprehension. *Learning and Individual Differences, 19*, 252-256.

Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin, 107,* 238-246.

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika, 74*(1), 137-143.

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

Chen, F. F., Hayes, A., Carver, C. S., Laureceau, J-P., & Zhang, Z (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality, 80*(1), 219-251.

Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2010). Reading comprehension of scientific text: A domain-specific test of the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 102*, 687-700.

Graesser, A. C. (2007). An introduction to strategic Reading comprehension. In D. S. McNamara (Ed.), *Reading comprehension strategies* (pp. 3-26). Mahwah, NJ: Lawrence Erlbaum Associates.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371-395.

Gignac, G. E. (2007). Multi-facto modelling in individual differences research: Some suggestions and recommendations. *Personality and Individual Differences, 42,* 37-48.

Green, B. A. (1995). Comprehension of expository text from an unfamiliar domain: Effects of instruction that provides either domain-specific or strategy knowledge. *Contemporary Educational Psychology, 2,* 313-319.

Green, S. B., & Yang, Y. (2009). Reliability of summed items scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika, 74*(1), 155-167.

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research, 70*, 151-179.

Hyde, J. (1981). How large are cognitive gender differences? A meta-analysis. *Psychological Bulletin, 104*, 53-69.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review, 95,* 163-182.

León, J. A. (2004). Un nuevo enfoque de la competencia lectora basado en diferentes tipos de comprensión [A new approach to reading skills based on different kinds of comprehension]. *Seminario de primavera 2004*. Fundación Santillana, 39-50.

León, J.A., & Escudero, I. (2015). Understanding causality in Science discourse for Middle and High School Students. Summary task as a Strategy for Improving Comprehension. In K.L. Santi and D. Reed (Eds.), *Improving Comprehension for Middle and High School Students* (pp. 75-98). Springer International Publishing Switzerland.

León, J. A., Escudero, I., & Olmos, R. (2012). *Ecomplec: una propuesta de evaluación de la comprensión lectora en Primaria y Secundaria* [A proposal for the assessment of reading comprehension in primary and secondary education]. Madrid: TEA Ediciones (ISBN 978-84-15262-45-9).

León, J. A., Escudero, I., Olmos, R., Sanz, M. M., Dávalos, T., & García, T. (2009). ECOMPLEC: un modelo de evaluación de la comprensión lectora en diversos tramos de la Educación Secundaria. *Psicología Educativa, 15*(2), 123-142.

McVey, J., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology, 141*(2), 302-320.

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence, 41,* 407-422.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus User's Guide*. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

OECD (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. PISA, OECD Publishing, Paris.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction, 19*(3), 228-242.

Raykov, T. (2001). Bias of coefficient afor fixed congeneric measures with correlated errors. *Applied Psychological Measuremalet, 25*(1), 69-76.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2012). Bifactor models and rotations: Exploring the extent to which multidimensional data Yield univocal Scale Scores. *Journal of Personality Assessment, 92, 6,* 544-559.

Rijmen, F. (2011). *Hierarchical factor item response theory models for PIRLS: Capturing clustering effects at multiple levels*. Research Report. ETS, Princeton, New Jersey.

Sáenz, L. M., & Fuchs, L. S. (2002). Examining the reading difficulty of secondary students with learning disabilities. Expository versus narrative text. *Remedial and Special Education, 23*(1), 31-41.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1-10.

Unsworth, N., & McMillan, B. D. (2013). Mind wandering and reading comprehension: Examining the roles of wording memory capacity, interest, motivation, and topic experience. *Journal of Experimental Psychology: Learning, Memory and Cognition, 39*(3), 832-842.

Van den Broek, P., &, Espin, C. A. (2012). Connecting cognitive theory and assessment: Measuring individual differences in reading comprehension. *School Psychology Review, 41*(3), 315-325.

Vizcarro, C., & León, J. A. (Coords.) (1998). *Nuevas tecnologías para el aprendizaje* [New technologies for learning]. Madrid: Pirámide.

Zwaan, R., & Singer, M. (2003). Text comprehension. In A. C. Graesser, M. A. Gernsbacher & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 83-121). Mahwah, NJ: Lawrence Erlbaum Associates.