

RELIABILITY AND VALIDITY OF THE LATVIAN VERSION OF THE COMPUTERIZED EXECUTIVE FUNCTIONS TEST “EXAMINER” IN A STUDENTS SAMPLE: RESULTS OF THE PILOT STUDY

Liena Ivanova

Ilze Plauca

Sandra Sebre

Malgozata Rascevska

University of Latvia, Latvia

Abstract. *The aim of this pilot-study was to assess reliability and validity of the Latvian version of the computerized executive functions test EXAMINER (Executive Abilities: Measures and Instruments for Neurobehavioral Evaluation and Research) in a sample of high school students and university students. The adaptation procedure included translation of the instructions and technical configuration of the test’s computerized version. The Latvian version of the EXAMINER includes tasks that measure executive functions of inhibition, set shifting and working memory updating. Participants included two sub-groups: 22 (14 males, 8 female) students from high school and 39 students (11 males, 28 female) from university. Results showed statistically significant differences between both sub-groups on most of the EXAMINER tasks. The main tendency was that the high school students showed faster reaction time, but made more mistakes. Results also showed good reliability for most of the EXAMINER tasks and good criterion validity. Potential improvements of the test and its potential use in schools and clinics are discussed.*

Keywords: *EXAMINER test battery, executive functions, Latvian adaptation, young adulthood.*

Introduction

Executive functions are indispensable for handling both simple and complex situations of daily life. The assessment of executive functions is used to understand learning ability, cognitive or functional difficulty, and is carried out by clinical observation or by psychometric testing (Barkley, 2012). The EXAMINER test battery was designed for such assessment, and it has good reliability indicators (Kramer, 2014). There has been a dearth of executive function tests available in Latvian, and therefore, the EXAMINER test was adapted to Latvian language and applied to assess executive functions of late adolescents and young adults in the pilot-study reported in this article.

Executive functions are crucial for maintaining goal-directed behavior. They have been described variously, as a combination of several components such as self-regulation, consistent behavior, elasticity, inhibition of response, planning, organizing of behavior and other components (e.g., Barkley, 2012; Goldstein, Naglieri, Princiotta & Otero, 2014). Three components of executive functions have been described as the most important – *inhibition*, *set shifting* and *working memory updating* (e.g., Miyake et al., 2000). These are also most widely assessed in research, perhaps because they can be operationalized more accurately than, for example, planning, self-regulation or other executive components. In the age group of sixteen to eighteen year olds, the three factor model of executive functions has proven to have the best explanatory value (Friedman, Miyake, Young, DeFries, Corley & Hewitt, 2008).

There are different types of psychometric executive function measures – psychometric test batteries (e.g., Delis-Kaplan Executive Function System; Delis, Kaplan & Kramer, 2001); CEFI; Naglieri & Goldstein, 2014) and individual tasks (e.g., Stroop test, Wisconsin Card Sorting test). Also other methods are being used to evaluate executive functions such as self-reports, parent or teacher reports or clinical observations (e.g., Barkley, 2012; Naglieri & Goldstein, 2014). It is mostly up to researchers, their aims and the methods available to them to decide which of the tests will be used. There are arguments both in favor of psychometric tests and clinical observations (Barkley, 2012).

Currently, there is lack of executive function measurement methods in Latvian that could be used for a broad age range and administered by computer. Therefore, it was decided to adapt the executive function measurement test battery EXAMINER which was developed at the University of California, San Francisco, Memory and Aging Center. It has been used with both clinical and normal samples within different age groups (see Kramer, 2014; Possin, Lamarre, Wood, Mungas & Kramer, 2013). The original EXAMINER test version for adults in English comprises the following tasks: Dot Counting, N-back Task (1-back; and 2-back), Flanker Task, Continuous Performance Task, Anti-Saccades, Set Shifting, and Fluency: Phonemic and Category, Unstructured Task, Insight and Social Norms Questionnaire.

In the Latvian version, the following EXAMINER test tasks are included: *Flanker Task*, *Set Shifting*, *N-back Task: 1-back* and *N-back Task: 2-back*. These tasks assess *inhibition*, *set shifting* and *working memory updating*. The other tasks of the original version of the EXAMINER were excluded from adaptation for the following reasons: 1) this adaptation follows the Miyake (2000) model of three main executive functions; 2) some of the original tasks are not computer-based thus making the administration more complicated; 3) some of the original tasks (e.g., Social Norms Questionnaire) are culture-specific and would need to be modified; 4) for practical reasons the test had to be shortened for use in

school settings (max 35-40 minutes). For adaptation all test materials were double-translated from English to Latvian, and technical adjustments of the computerized test version were made, in accordance with test adaptation guidelines (International Test Commission, 2005).

Students in high school and university have different educational settings and slightly different age, thus it can be assumed that executive functions (among other cognitive abilities) should be higher for university students, as they have successfully finished high-school, accomplished some goals and entered university. Such difference would also support criterion validity of executive functions measurement. Previous research show that executive functions are genetically determined to a great extent (Friedman, Miyake, Young, DeFries, Corley & Hewitt, 2008), but their development is influenced by environmental factors as well. Research suggests that inhibition can mature already at early adolescence, before shifting and updating (Anderson, 2002; Ikeda, Okuzumi & Kokubun, 2013) that both fully mature later, in young adulthood. Results from longitudinal research on executive function maturation in early and late adolescence show that performance on inhibition tasks shows very small improvement from 11 to 19 years of age, thus indicating that inhibition has almost reached maturation in early adolescence. More improvement is found in the performance of working memory and attention switching tasks, thus indicating that these functions are still maturing during adolescence (Boelema et al., 2014). Shifting continues to develop in late adolescence, but working memory is still developing in young adulthood (Huizinga, Dolan & van der Molen, 2006). Young adults show superior performance in working memory tasks compared to adolescents, especially when working memory load is increased (Thomason et al., 2009).

The aim of this study was to assess reliability and validity of the Latvian version of the computerized executive functions test EXAMINER in a sample of high school students and university students in the pilot study.

Method

Measurement

The following tasks were included in the Latvian adaptation of the EXAMINER test battery: *Flanker Task*, *Set Shifting*, *Continuous Performance Test*, and *N-back Task: 1-back* and *N-back Task: 2-back*.

Flanker Task (FT) is designed to test inhibition – the ability to stop one's actions which are prone to automatic response. In this task several arrows are shown on the screen pointing to left or right. The participant is asked to indicate whether the arrow in the middle is pointing to left or right, and press the

appropriate arrow on the keyboard. The position of the middle arrow varies across trials and it can be congruent or incongruent with the other arrows.

Set Shifting (SS) is a task that requires the ability to switch attention between two dimensions that a figure has – color and shape – and respond appropriately. The participant sees a figure (triangle or rectangle) in the middle of the screen and has to compare it to two figures at the bottom of the screen. Each time a different cue appears indicating whether figures have to be compared by color or shape.

N-back Task (NT) is a test for working memory updating. A participant has to continuously remember stimuli that were shown one (*N-back Task: 1-back*) or two (*N-back Task: 2-back*) screens earlier. In the 1-back condition, the participant is asked to remember the position of a square and compare it to the position of the next square. But before the next square appears, the participant is asked to say out loud a number which appears in the middle of the screen. In the 2-back condition, the participant is asked to compare the position of each square to the one shown two screens before, thus making the task more complicated.

For high-school pupils also mean grades were gathered for subjects such as Mathematics, Physics, Languages and others. These grades of academic performance were used to assess criterion validity of the EXAMINER test.

Participants

In total there were 61 participants comprising two subgroups: 1) 22 high-school students in their final year (14 males, 8 female) with mean age $M = 18.14$ ($SD = 0.35$) from regional State gymnasias; 2) 39 first year university students (11 males, 28 female) with mean age $M = 19.84$ ($SD = 0.79$) of University of Latvia. The age in both groups differed significantly ($t = -11.50$, $p = .00$). Students were offered to take part in the study as the part of the course. Participation in the study was voluntary and the participants were able to receive feedback about their results.

Procedure

Participants were each invited in a classroom equipped with computers. There was an individual session with each participant that took about 30 minutes. The test is fully computerized, all instructions can be seen on the screen and also the examiner explains them, if necessary. After each testing session a brief feedback was asked from each participant to find out how difficult each test seemed and to get other information about the test and testing settings that might be useful for the adaptation process.

The test was administered using free-access program for research *PsychoPy* (Peirce, 2007). All the data from each session were automatically recorded in MS Excel files and were further processed with statistical packages.

Results

First, the reliability and validity of the test and an analysis of the testing procedure is presented. All participants indicated that they understood the provided instructions and the tasks. Descriptive statistics of all the EXAMINER tasks can be seen in Table 1. To assess psychometric properties Cronbach's alpha was calculated for each task's accuracy rate and response time. For the FT test reliability was calculated for incongruent trials, SS for shifted trials, and NB for all trials (EXAMINER, 2011). Reliability of each test is presented separately for the high school and university student groups, as they are slightly different, and for both groups taken together (see Table 1). To assess criterion validity, results of both groups were compared, and for high-school group relation to grades were also examined.

Table 1 Means and standard deviations of EXAMINER scales of high school students and university students

Note. FT – Flanker Task, SS – Set Shifting, NB – N-Back Task.

EXAMINER task	High school students			University students			Reliability (both groups)			
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>t-test</i>	<i>p</i>	<i>N</i>	α
Inhibition										
FT precision	22	22.14	1.98	38	23.08	1.50	-2.07	.04	60	.91
FT reaction time	22	0.63	0.11	38	0.68	0.13	-2.33	.02	60	.96
Set shifting										
SS precision	22	60.50	2.43	36	58.53	5.55	1.57	.12	58	.82
SS reaction time	22	0.62	0.24	36	0.77	0.29	-2.64	.01	58	.97
Working memory updating										
NB1 precision	22	26.14	1.91	38	27.32	1.80	-2.39	.02	60	.31
NB1 reaction time	22	0.61	0.18	38	0.73	0.20	-3.72	.00	60	.82
NB2 precision	22	70.33	9.12	37	67.30	10.21	1.13	.26	59	.85
NB2 reaction time	22	0.89	0.41	37	0.98	0.38	-1.13	.27	59	.96

Reliability of the most EXAMINER tasks is high when both student groups are analyzed together, with the exception for NB1 precision measurement, which shows low reliability. For the high school group FT showed acceptable reliability for precision ($\alpha = .59$), and good reliability for reaction time ($\alpha = .85$); SS showed rather low reliability for precision ($\alpha = .45$) and very good reliability for reaction time ($\alpha = .94$). The low reliability for SS precision can be explained due to zero variance for several of the scale items and overall low variance for

items. This means that the high school group in the current study is rather homogeneous. NB1 showed very low reliability for precision ($\alpha = .21$) and very good reliability for reaction time ($\alpha = .83$). Low reliability for NB1 precision can be explained due to zero variance for several of the scale items. Contrary to NB1, NB2 showed very good reliability for precision ($\alpha = .84$). Although there was zero variance on some items, the total number of items is 90, thus making the overall reliability rather high. NB2 also show very good reliability for reaction time ($\alpha = .90$).

For the university student group Cronbach's alpha for the FT test was very good for both response time and accuracy ($\alpha = .97$); SS showed very good reliability for response time ($\alpha = .97$) and good for response accuracy ($\alpha = .85$). For NB1 and NB2 both had good reliability for response time ($\alpha = .91$ for NB1 and $\alpha = .96$ for NB2). For response accuracy NB2 showed good reliability ($\alpha = .86$), but very low for NB1 ($\alpha = .30$). Some task items showed small variance and this could be the reason for the low Cronbach's alpha.

To test the criterion validity of the Latvian version of the EXAMINER, the results of all tests were correlated with academic achievement results for Mathematics, Chemistry, Latvian language, Geography and others, gathered from high-school students ($n = 22$). The results showed positive correlation for most of the study subjects. Significant correlations were observed between Mathematics grade and SS shifted tasks ($r = .46$), Chemistry grade and FT response time which showed opposite correlation compared to other reaction time correlations ($r = .42$), FT accuracy rates ($r = .45$), and response time of 2-Back Task ($r = -.44$). SS shifted tasks also had significant correlations with both Latvian language and Literature grades ($r = .45$ and $r = .50$ respectively). English and Geography grades correlated with FT accuracy rates ($r = .46$ and $r = .64$), Geography grades correlated with response time of 2-Back Task ($r = -.49$). These correlations indicate that the criterion validity of the Latvian version of the EXAMINER is overall acceptable.

To test if there are any differences between executive functions in the two subgroups, t-test for independent samples was calculated for EXAMINER tasks, including assessment of Levene's test for equality of variances. Results (see Table 1) show that there are statistically significant differences between the two groups in reaction time of FT, SS and NB1, meaning that high school students had faster reaction. Differences were also found in precision of FT and NB1, with high school students being less precise than university students. No statistically significant differences were found in NB2 reaction time and precision. These results partly confirm criterion validity of the EXAMINER.

Discussion

The assessment of reliability and validity of the Latvian version of the computerized executive functions test EXAMINER was done in a sample of high school students and university students as a part of the pilot study for the recently adapted test.

The analyzed psychometric properties of the Latvian version of the EXAMINER test show that reliability indicators are good for most of the tests, indicating that the Latvian version of the EXAMINER can be used in further studies. However, some tests show lower reliability, mostly in regard to accuracy. This may be due to the small sample size ($n = 22$ and $n = 39$), high overall accuracy rates of the respondents and small variation in groups. The original version of the EXAMINER was designed for participants with a broad range of neurological conditions from different age groups (Kramer, 2014). This may result in very high accuracy rates in non-clinical samples as in the current study, whereby participants on the FT showed very high accuracy rate. In addition, the participants' feedback indicates that the test seemed too easy for them. For further use, it is recommended to test this task in another sample or modify the existing task to make it more complicated in order to show higher variation in a non-clinical sample, like students.

The original EXAMINER test results were correlated with several cognitive ability tests to assess criterion validity (EXAMINER, 2011). In the current pilot study of the Latvian version of the test academic grades of the high-school participants were correlated with the executive function test results. Previous studies have shown that results of executive functions tests most often are related to academic achievement (see Fuhs, Nesbitt, Farran & Dong, 2014). Also, the results from the current study show statistically significant correlations, confirming that criterion validity of the adapted test version is good.

Two student groups were compared regarding their executive functions, assessing also the criterion validity of the test. Results show that there are statistically significant differences between the two groups, and there is a tendency that high school students have faster reaction time, but are less precise, and university students are more precise, but react slower. The existence of such differences could be due to several reasons, as there are also several limitations to the study.

First, one limitation of the current pilot study is that the sample sizes were not large enough to make generalizations about the broader population, and these results might simply characterize the current sample. Second, there are some geographical differences – the high school students were living in a regional small town, but the university students came from both small and large

cities and were all currently living in a large city. Third, high-school students in the last year of school may be more anxious and may be hastier in doing such tasks. Fourth, contrary to previous research that suggests that inhibition may mature earlier in adolescence (Boelema et al., 2014), these results show higher results on inhibition for the university student. There are also some limitations of the testing procedure of the EXAMINER. The current version of this test can only be administered in individual sessions because some tasks (e.g., NB1) require oral responses from the participant. This is very suitable for clinical settings but not very convenient for large scale testing. This issue should be considered and a modified version of the task could be created for administration in group settings.

To conclude, the recently adapted Latvian version of executive functions test EXAMINER shows overall good reliability and validity and can be further used to assess inhibition, set shifting and working memory updating in Latvian samples. The current pilot-study in the samples of students from high-school and university also showed statistically significant differences between both groups on most of the EXAMINER tasks, and the main tendency was that the high school students had faster reaction time, but made more mistakes, whereas university students had slower reaction time, but were more accurate.

References

- Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 8 (2), 71-82. doi: 10.1076/chin.8.2.71.8724
- Barkley, R. (2012). *Executive functions: what they are, how they work, and why they evolved*. New York: The Guilford Press.
- Boelema, S. R., Harakeh, Z., Ormel, J., Hartman, C. A., Vollebergh, W. A. M., & van Zandvoort, M. J. E. (2014). Executive functioning shows differential maturation from early to late adolescence: Longitudinal findings from a TRAILS study. *Neuropsychology*, 28 (2), 177–187. <http://doi.org/10.1037/neu0000049>
- Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan executive function system*. San Antonio: The Psychological Corporation.
- EXAMINER. (2011). *Executive abilities: measures and instruments for neurobehavioral evaluation and research. User manual (EXAMINER)*. Memory and Aging Center, NIH Institute, University of California, San Francisco. (Electronic issue)
- Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology*, 137 (2), 201–225. doi:10.1037/0096-3445.137.2.201
- Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Developmental Psychology*, 50 (6), 1698–1709. <http://dx.doi.org/10.1037/a0036633>

- Goldstein, S., Naglieri, J. A., Princiotta, D., & Otero, T. M. (2014). Introduction: A history of executive functioning as a theoretical and clinical construct. In S. Goldstein & J. A. Naglieri (Eds.) *Handbook of executive functioning*. Springer: New York. (pp. 3-12).
- Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44* (11), 2017–2036. <http://doi.org/10.1016/j.neuropsychologia.2006.01.010>
- Ikeda, Y., Okuzumi, H., & Kokubun, M. (2013). Age-related trends of stroop-like interference in animal size tests in 5- to 12-year-old children and young adults. *Child Neuropsychology*, *19* (3), 276-291. doi:10.1080/09297049.2012.658364
- International Test Commission. (2005). International guidelines on test adaptation. [www.intestcom.org]
- Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: a question of construct validity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33* (3), 615–622. doi:10.1037/0278-7393.33.3.615
- Kramer, J. H. (2014). Special series introduction: NIH examiner and the assessment of executive functioning. *Journal of the International Neuropsychological Society*, *20*, 8-10. doi: 10.1017/S1355617713001185
- Miyake, A., Friedman, N., Emerson, M., Witzki, A., Howerter, A., & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobe tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49-100.
- Naglieri, J. A., & Goldstein, S. (2014). Using the comprehensive executive function inventory (cefi) to assess executive function: from theory to application. In S. Goldstein & J. A. Naglieri (Eds.) *Handbook of executive functioning*. Springer: New York. (pp. 159-170)
- Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162* (1-2), 8-13. doi:10.1016/j.jneumeth.2006.11.017
- Possin, K. L., Lamarre, A. K., Wood, K. A., Mungas, D. M., & Kramer, J. H. (2013). Ecological validity and neuroanatomical correlates of the NIH EXAMINER executive composite score. *Journal of the International Neuropsychological Society*, *19*, 1–9. doi:10.1017/S1355617713000611
- Thomason, M. E., Race, E., Burrows, B., Whitfield-Gabrieli, S., Glover, G. H., & Gabrieli, J. D. (2009). Development of spatial and verbal working memory capacity in the human brain. *Journal of Cognitive Neuroscience*, *21* (2), 316-332. doi:10.1162/jocn.2008.21028