Cognitive components of foreign word stress processing difficulty in speakers of a native language with non-contrastive stress

Short title: Cognitive factors of L2 word stress processing

Ferenc Honbolygó[1,2], Andrea Kóbor[1], Valéria Csépe[1,3]

[1]: Brain Imaging Centre, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar tudósok körútja 2., H-1117, Budapest, Hungary

[2]: Institute of Psychology, Eötvös Loránd University, Izabella u. 46., H-1064, Budapest, Hungary

[3]: Department of Cognitive Science, Budapest University of Technology and Economics, Egry József utca 1., H-1111 Budapest, Hungary

Corresponding author: Ferenc Honbolygó

E-mail: honbolygo.ferenc@ttk.mta.hu

Address: Brain Imaging Centre, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar tudósok körútja 2., H-1117, Budapest, Hungary

Phone: + 36-1-3826615

1

Abstract

Objectives

Stress "deafness" is a difficulty in the detection of stress pattern changes of L2 words.

This study investigated the influence of cognitive factors and L2 proficiency on the

processing of L2 stress.

Methodology

Fifty-four native speakers of Hungarian, a language with non-contrastive stress,

participated in the study, who were categorized as not speaking German or having a

proficiency at the intermediate or advanced level. They had to recall sequences with

increasing length consisting of German pseudowords that differed in either their

phonemes (phoneme task) or stress patterns (stress task). Cognitive factors measured

included working memory, phonological awareness, and inhibitory control.

Data and Analysis

The accuracy data obtained in the sequence recall task was analyzed with generalized

linear mixed modeling. Two separate analyses were performed to investigate the

presence of stress "deafness" and the effect of cognitive factors.

Findings

Results showed that the stress task led to lower performance than the phoneme task,

irrespective of L2 proficiency. Furthermore, the analysis showed different cognitive

factors contributing to the performance in the tasks: in the phoneme task, it was working

memory, phonological awareness, and inhibitory control, while in the stress task, it was only the working memory and phonological awareness but not the inhibitory control.

Originality

This is the first study to provide evidence about the cognitive background of the stress "deafness" effect, and to suggest the differential role of inhibitory control in phoneme and stress processing.

Implications

These findings demonstrate the robustness of the stress "deafness" effect in a language with non-contrastive stress, provide evidence of the effect being independent of L2 proficiency, and suggest that speakers of languages with non-contrastive stress do not have the necessary cognitive basis to form accurate L2 stress representations.

Keywords: inhibitory control, L2 learning, phonological awareness, stress "deafness", working memory

Acquiring a foreign language presents difficulties for learners on several levels of linguistic processing. Most notably, the acquisition of native-like phonology is problematic for many learners (Scovel, 1969). Difficulties concern not only the production and perception of speech sounds but also the prosodic features, including stress and intonation. Stress contributes to the segmentation of continuous speech into words; therefore, the correct perception of word stress in essential in foreign language acquisition. In the present study, we investigated the perception of foreign language word stress: Whether knowledge of the foreign language affects word stress processing difficulties, and which cognitive factors influence the processing of word stress information.

Languages use stress differently. One important cross-linguistic dissimilarity is how stress differentiates the meaning of words: in languages with contrastive stress (e.g., English, Dutch, German), words with different stress patterns but similar phonetic structure can have different meanings. In languages with non-contrastive stress (e.g., French, Finnish, Hungarian), the prosodic pattern of words does not have such a lexical role. In a series of experiments, Dupoux and colleagues demonstrated that the non-contrastive nature of stress influences how French native speakers perceive stress pattern, leading to a marked impairment in stress processing termed stress "deafness".

Dupoux et al. (2001) provided evidence that French speakers compared to Spanish speakers showed impaired performance in a sequence recall task (SRT), in

which they had to memorize patterns of randomly changing stress-minimal pairs (words differing only in stress pattern, e.g., *TUpu* vs. *tuPU*; capital letters henceforth indicate stress), while showed no such impairment when they had to remember phoneme-minimal pairs (words differing in a single phoneme, e.g., *tuku* vs. *tupu*). French speakers showed an especially low performance on the stress task when the elements of the sequence to be remembered were created from acoustically variable stimuli by changing the fundamental frequency (f0) of words or were uttered by different speakers. The interpretation of the results was that the acoustic variation required that participants encode the elements phonologically in their short-term memory buffer to recall the sequences. However, for French speakers, this representation was not available in the case of stress pairs, because of the non-contrastive nature of stress.

In a further study, Dupoux et al. (2008) demonstrated that the performance of native speakers of French in SRT remained low even if they had a relatively high proficiency in a language with contrastive stress (in this case Spanish). The authors showed that native French learners of Spanish at the beginner, intermediate, and advanced levels – as classified by the authors based on a questionnaire related to the linguistic background and current language usage of participants – showed similar performance on SRT. More importantly, their performance did not differ from that of French speakers without any knowledge of Spanish. At the same time, all four French speaker groups performed worse than Spanish speakers. Based on these results, Dupoux

et al. (2008) argued that the impaired stress processing of native speakers of French does not depend on metalinguistic knowledge (which must have been certainly acquired by experienced L2 learners), but it represents a persistent limitation in the processing of stress information.

Peperkamp et al. (2010) extended the above result by investigating the stress processing deficit in other languages (Finnish, Hungarian, and Polish). The authors found that similarly to French speakers, Finnish and Hungarian speakers also showed a strong stress "deafness" effect, while Polish speakers exhibited a smaller stress "deafness" effect compared to the other three. Polish differs from the other languages in that it contains some irregularly stressed (mostly loan) words, while in the other three languages, stress is completely predictable, and there are no exceptions.

The above findings (Dupoux et al., 2001, 2008; Peperkamp et al., 2010) concerning the impaired stress processing of speakers of languages with non-contrastive stress left open some questions. First, the effect of learning a language with contrastive stress was investigated only with French participants; therefore, the generalizability of this effect is questionable. Second, the contribution of cognitive factors to the performance on the SRT has not been investigated before. The SRT involves working memory (WM) related cognitive demands, because as the length of sequence increases, participants have to rely more on WM. WM has been found to be an important component of language learning (Baddeley, Gathercole, & Papagno, 1998; Gathercole

& Baddeley, 1989, 1990) and also of second language acquisition (Dörnyei & Skehan, 2008; Miyake & Friedman, 1998) (for review see Baddeley, 2015). Therefore, individual differences in WM could have a confounding impact on task performance.

Besides WM, phonological awareness (PA) and inhibitory control (IC) functions could also affect task-solving as language specific (PA) and general cognitive (IC) abilities. PA is the ability to attend to, detect, and manipulate the phonological units of language, and it could be assumed to be involved in learning the words of native (M. Snowling, Goulandris, Bowlby, & Howell, 1986; M. J. Snowling, Chiat, & Hulme, 1991) and of foreign language (C.-F. Hu, 2003; X. Hu et al., 2013). Hu et al. (2013) found that the phonological coding ability (measured with a PA task) and other cognitive and personality factors but not the WM predicted L2 pronunciation aptitude in adult advanced learners of English. Kormos and Sáfár (2008) demonstrated that PA and WM contribute differently to the success of L2 acquisition. Recently, Li and Chen (2016) demonstrated the effectiveness of PA training on L2 word reading.

IC, the ability to successfully respond to a task-relevant dimension while inhibiting inappropriate automatic responses or suppressing interference due to a task-irrelevant dimension (Brydges et al., 2012) was found to be related to learning a second language (Bialystok, 2011; Bialystok, Craik, Green, & Gollan, 2009; Martin-Rhee & Bialystok, 2008). Specifically, Bialystok et al. (2009) recite a large number of evidence showing that bilingual children and adults demonstrate an advantage in the inhibition of

unwanted information and selection of relevant information. The proposed explanation for this bilingual advantage is that bilingual speakers have to continuously coordinate two potentially active languages and they need a strong inhibition of the nonrelevant language while maintaining the attentional set on the language actually used. Therefore, it might be supposed that IC ability contributes to solving L2 tasks. Empirical evidence, provided for example by Lev-Ari and Peperkamp (Lev-Ari & Peperkamp, 2013) showed that L2 speakers with poorer IC skills experienced greater co-activation of competing linguistic units, leading to a poorer performance than speakers with better IC skills. Therefore, it can be assumed that speakers with higher IC skills are more likely to activate the language-relevant exemplars. The study of Pelham et al. (2014) shows that becoming fluently bilingual in young adulthood can lead to the same cognitive effects as becoming fluently bilingual in childhood in terms of the advantages in executive function.

In the present study, we intended to directly examine the influence of cognitive factors (WM, PA, and IC) on the processing of L2 stress using the SRT. We investigated the stress processing difficulty of native Hungarian speakers learning or not learning German as a second language. We assumed that Hungarian speakers show the stress "deafness" effect as demonstrated earlier by Peperkamp et al. (2010). Moreover, we expected that more general cognitive factors might contribute to the performance on the SRT as well. Previous research focused on more proximal dependent measures of

L2 skills (i.e., vocabulary, pronunciation, etc.) in specific groups and circumstances of L2 acquisition, and we wanted to broaden these results by investigating the influence of WM, PA, and IC on SRT performance, which are less proximal indices of L2 skills. We also supposed that by studying the effect of cognitive factors on SRT performance, we could elucidate their unique contribution to phoneme and stress processing, which could shed light on how the representations of these separate linguistic units differ from each other in L2 acquisition.

Method

Participants

Three groups (controls, intermediate learners, and advanced learners) of adult participants ($N = 60$) took part in the experiment; all of them were native Hungarian speakers. Altogether six participants were excluded from the sample: two participants were dropped out because of technical problems, one due to a neurological problem, one because of minor hearing problems that hindered task-solving, and two participants because of having a bilingual background. The final sample size was the following: n(Control) = 19, n(Intermediate) = 17, n(Advanced) = 18.

Participants were recruited from different universities and language schools in Budapest through advertisement and personal communication. Descriptive characteristics of the final sample are presented in Table 1.

Participants in the control group had never learned German. During recruitment, the language proficiency of German language learners was unknown. All applicants describing their German language proficiency as intermediate or advanced – based on subjective impressions, language exams, or attendance at language schools – were invited to take part in the experiment. During the testing session, German language learners were examined in a progressive linguistic test consisting of 100 multiple-choice items (Koukidis, 2003). The test, based on the sum of correct answers, determined *language proficiency* according to the levels of the CEFR (Common European Framework of Reference for Languages: Learning, Teaching, Assessment). Participants achieving B1 or B2 levels were classified as intermediate learners of German and those achieving C1 or C2 levels were assigned to the group of advanced learners. Intermediate and advanced learners differed significantly in the total score of the linguistic test ($M = 62.3$ vs. $M = 84.3$, see Table 1).

The study was approved by the Ethical Review Committee for Research in Psychology and was conducted in accordance with the Declaration of Helsinki. Prior to testing, participants familiarized with the details of the different tasks, and they gave their informed consent. Participants received a book token or a voucher for German language courses as a compensation for taking part in the study.

Table 1. Descriptive characteristics of the control, the intermediate learner, and the advanced learner groups.

	Control (n = 19)	Intermediate (n = 17)	Advanced (n = 18)	t / χ²
Age [years]				
Mean [a]	23 (6)	28 (12)	29 (9)	8.12*
Range	18 – 46	18 – 57	20 – 57	-----
Gender [Male/Female] [b]	3/16	6/11	4/14	1.86
Education [undergraduate /graduate / no degree] [b]	17/2/0	11/5/1	11/7/0	6.16
Foreign languages spoken [a]	2.2 (0.8)	2.5 (0.6)	2.8 (0.8)	7.32*
Total score of linguistic test [correct answers]	-----	62.3 (10.2)	84.3 (6.8)	-7.48***

Note. *p* < .05; ***p* < .001. Values in parentheses denote standard deviation. When the assumption of homogeneity of variances was violated, a Welch's modified *t*-test was conducted. [a] = In case of violating the assumption of normality, Kruskal-Wallis tests were performed. [b] = If the assumptions for a chi-squared test (at least 80% of the expected counts are more than five and all expected counts exceed one) were not met, exact significance tests were selected for Pearson's chi-square.

Measures

Digit Span. The three Digit Span subtests (Forward, Backward, and Sequencing) of the Hungarian version of WAIS-IV (Rózsa & Kő, 2010; Wechsler, 2008) were administered for measuring WM. We derived the Digit Span scaled score and used it as the standardized equivalent of the total raw score.

Stroop Test. For measuring the interference suppression subcomponent of IC (Martin-Rhee & Bialystok, 2008), participants performed a computerized version of the

Golden Stroop Test (Golden, 1978), in which RT for each read or named item can be recorded. Participants were required to read or name as many items as they could in 45 seconds in each of the three conditions (word, color, and color-word). Only one condition was presented at a time on the screen and items (words, colors, and color-words, respectively) were listed in five columns and twenty rows. Participants proceeded from the top to the bottom. A rectangular frame indicated the item they actually read or named and they paced the frame to the next item by pressing the left mouse button. A normalized difference score of average RTs was used as an indicator of interference. This was calculated as follows: (RT measured for the color-word condition minus RT measured for the color condition) / RT measured for the color-word condition.

Phoneme Deletion. The phonological awareness (PA) of participants was examined by the Phoneme Deletion subtest of the comprehensive 3DM-H (Dyslexia Differential Diagnosis Maastricht, Blomert and Vaessen, 2009; Hungarian version: Tóth et al., 2014), which is a Hungarian adaptation of the originally Dutch computerized test battery for the assessment of developmental dyslexia. In this task, participants were asked to skip the first, last, or middle phoneme of auditory presented pseudowords, and to restore the remaining part of the letter strings. The whole task consisted of 27 items, and the derived variable indicated the speed of responding. Since there are no

standardized phonological awareness tests in Hungarian for adults, we used the Phoneme Deletion task of 3DM-H optimized for children.

Sequence Recall Task. We used a modified version of the Sequence Recall Task (SRT) described by Dupoux et al. (2001) and Dupoux et al. (2008). Two bisyllabic minimal pairs (CVCV pseudowords) following the phonotactic constraints of German language were constructed from the German verb *müde* (get tired) by using the Wuggy pseudoword generator (Keuleers & Brysbaert, 2010). In previous studies (Dupoux et al., 2001, 2008), stimuli having neutral or unknown stress related acoustical cues (Dutch) were used. In contrast with previous studies, we used stimuli with the actual acoustical cues of the foreign language, because we assumed that not only the stress pattern (i.e., the phonological features) but also its acoustical realization (i.e., the phonetic features) was an important factor of stress processing. We expected that using familiar acoustical features makes the task easier; therefore, the stress "deafness" effect would be smaller than in earlier studies (see Peperkamp et al., 2010), but we would still find evidence for the stress processing difficulty, which is a strong argument to its pervasiveness.

The phoneme contrast [*mäge – mäse*] differed only in one consonant; the stress contrast [*LIdu – liDU*], differed in the location of stress (i.e., initial vs. final). Stimuli were recorded ten times each in a carrier sentence produced by a female native speaker of German in a random order. Three tokens of each pseudoword with unambiguous

stress pattern were chosen. The duration of each token was equalized with the *lengthen overlap-add* formula of the Praat software (Boersma & Weenink, 2007) using the mean duration of the three tokens. Afterwards, the actual tokens were duplicated resulting in six tokens for each item. The study of Dupoux et al. (2001) suggests the joint contribution of memory load and phonetic variability to the stress "deafness" effect in the SRT. Therefore, in order to increase the acoustic variability, we shifted the overall f0 of the six tokens to 95, 97, 99, 101, 103, and 105 % of the original f0, without altering the f0 contours (for this method, see Dupoux et al., 2001). The word "OK" was presented following each sequence of pseudowords to act as a mask prohibiting the use of echoic memory when recalling the given sequence, and also to signal the end of the sequence (see Dupoux et al., 2001). This word was recorded by a male talker. The intensity level of all stimuli was equalized using root mean square normalization, and a 10 ms long rise and fall time was applied.

Mean durations of the pseudowords (*mäge, mäse, LIdu, liDU*) were 530, 590, 680, and 800 ms, respectively. Considering stress tokens, stressed syllables were longer than unstressed syllables both in case of *LIdu*, $t(5) = 3.09$, $p < .05$, and *liDU*, $t(5) = -7.48$, $p < .001$. The mean f0 of stressed syllables was higher compared to unstressed syllables; *LIdu*, $t(5) = 25.88$, $p < .001$, *liDU*, $t(5) = -9.5$, $p < .001$. Stressed syllables were also louder than unstressed ones, but this difference was significant only in the

case of initially stressed (*LIdu*) tokens, $t(5) = 7.61$, $p < .001$. Acoustic parameters of each pair of the stress contrast are displayed in Table 2.

The SRT consisted of two different parts: a learning and training phase and a testing phase. All participants were examined first on the phoneme contrast and then on the stress contrast; the whole procedure was the same in both contrasts. During the learning phase, they could listen to all tokens of each item by pressing the [1] and [2] buttons on the keyboard. In the phoneme / stress contrast, *mäge* / *LIdu* was mapped to button [1], and *mäse* / *liDU* was mapped to button [2]. The presentation of tokens followed one another in a pseudorandom order. Participants were allowed to keep listening the various tokens until they felt that they could discriminate them.

After learning the stimulus-response mappings, a short training phase started in which participants heard a token of one pseudoword, and they had to press the corresponding response key. Textual feedback ("Correct" / "Incorrect") was presented for 800 ms informing the participants whether they responded correctly. The maximum number of trials in the training phase was 30. After achieving 8 successful responses overall, they could proceed to the testing phase.

The testing phase consisted of 2 warm-up trials followed by 8 test trials at each sequence length (i.e., experimental block). A sequence was defined as a repetition of pseudowords. The SRT consisted of 5 experimental blocks, each containing 2, 3, 4, 5, and 6 item long sequences. Sequence length increased successively; therefore, all

participants listened to the length 2 sequence at first, and length 6 sequence in the end. Their task was to reproduce each sequence by typing the corresponding keys in the correct order followed by the Enter button. Feedback ("Correct" / "Incorrect") was displayed after each warm-up trial, but participants were only informed about their performance after finishing the 8 test trials for a given sequence (total number of correct responses). Altogether 40 test trials were presented in the phoneme contrast and 40 in the stress contrast. Only perfectly correct reproduction of the sequences was coded as correct (see Dupoux et al., 2001, pp. 1608–1609). The construction of specific pseudoword sequences for each sequence length and each contrast was the same as in the study of Dupoux et al. (2001). Warm-up trials contained the maximum number of transitions for each sequence (e.g., 1212) resulting in regular patterns (Dupoux et al., 2001, p. 1608).

The trials were self-paced: a fixation cross was displayed until participants pressed Space to start the actual trial. The first token was presented after 400 ms elapsed following keypress. The silent period between two successive pseudowords was 50 ms (Dupoux et al., 2008), and the "OK" sound (see above) followed each sequence after 300 ms. Pseudoword sequences were presented in a random order. After the sequence and the "OK" sound, a screen ("Reply now:") was immediately displayed allowing to type in the response, and it remained visible until participants gave their answer. Before starting the testing phase, participants were warned to check each of their reply

afterwards and correct the input string if necessary before finalizing the response by pressing the Enter key.

Table 2. Mean acoustic parameters of the tokens of stress contrast.

	M_Duration (SD)		M_f0 (SD)		M_Intensity (SD)	
	li	du	li	du	li	du
LIdu	278 (10)	253 (10)	204 (3.8)	151 (6.8)	55 (1.1)	51 (0.7)
liDU	236 (19)	354 (19)	175 (5.5)	210 (13.9)	53 (0.5)	54 (0.9)
Diff(LIdu)	$p < .05$		$p < .001$		$p < .001$	
Diff(liDU)	$p < .001$		$p < .001$		$p = .141$	

Note. Duration [ms], f0 [Hz], Intensity [dB]. Capital letters denote stressed syllables.

Procedure

All measures were administered in a predefined fixed order as follows: Digit Span, SRT, Stroop Test, Phoneme Deletion, and the linguistic test. Individual testing sessions took place in a quiet room and lasted approximately 1 hour and 40 minutes. Participants were allowed to have short breaks between each measure. Computerized tasks (SRT, Stroop Test, Phoneme Deletion) were presented by using Presentation software (v. 16.3; Neurobehavioral Systems).

Data analysis

First, in order to test whether stress "deafness" is present in our sample and whether language proficiency modulates this effect, we performed generalized linear mixed modeling (GLMM) with binomial distribution on accuracy data obtained in the SRT. The advantage of GLMM is that it could account for the nonindependence of observations nested within participants (Barr, Levy, Scheepers, & Tily, 2013; Bolker et al., 2009), and it is more reliable than analysis of variance (ANOVA) for binary outcomes such as the correctness of responses (Jaeger, 2008).

For modeling, we used the *glmer* function implemented in the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) of R (R Core Team, 2014). According to the recommendations of Barr et al. (2013), we followed a design-driven approach in specifying the model. In this model, the dependent variable represented the *correctness of responses* (0 = incorrect, 1 = correct). We entered the factors *Contrast* (phoneme contrast, stress contrast) and *Sequence length* (2, 3, 4, 5, 6 items), and the interactions of *Contrast* * *Sequence length* and *Group* (controls, intermediate learners, advanced learners) * *Contrast* as fixed effects in the model. Thus, according to the main question of our study, we modeled whether the distinct effect of language proficiency (*Group*) on the performance difference between phoneme and stress contrast (*Contrast*) contributes to explaining the variance in SRT accuracy data. We handled all fixed factors as categorical variables. The number of *Trials* was modeled as random effect (random intercepts) to account for by-item variation. In the case of participants, we estimated

random intercepts and random slopes for all experimental fixed effects (*Contrast* and *Sequence length*) and their interaction. The model was fitted with maximum likelihood parameter estimates and the *bobyqa* optimizer algorithm was used to avoid convergence failures. As we used the treatment contrast, the reference level of a given factor (i.e., Phoneme Contrast and Sequence length 2) served as a baseline to estimate the other levels. Pairwise comparisons were performed by the *glht* function.

Second, to investigate the effect of cognitive factors on SRT performance, GLMM with binomial distribution on accuracy data was performed again. Among the fixed effects, we entered the *Digit Span* scaled score, the *Stroop* normalized difference score of mean RTs, and the mean speed of *Phoneme Deletion* as measures of WM, IC, and PA, respectively. The values of these cognitive factors were centered and scaled before their inclusion in the model. The factor *Contrast* (phoneme contrast, stress contrast) and the interactions of *Contrast * Digit Span*, *Contrast * Stroop*, and *Contrast * Phoneme Deletion* were also entered as we aimed to test whether cognitive factors influence performance in the same degree across phoneme and stress contrasts. In addition, consistently with the previous model, we entered *Sequence length* and the *Contrast * Sequence length* interaction to estimate the contribution of the increasing length of the sequences to SRT performance across the different contrasts. In the second model, in the case of the Participants variable, the estimation of random slopes similarly to the first model was not possible due to the non-convergence of the model; therefore,

we decided to model only the random intercept. *Trials* were modeled as random effects

(random intercept), similarly to the first model. We handled the *cognitive factors* as

numerical variables. This second analysis was conducted with the same settings as

described above.

The two main analyses (presence of stress "deafness" and effect of cognitive

factors) were performed separately, because previously, introducing the main effects of

cognitive factors together with all the other experimental and *Group* effects in the same

model required the estimation of a large number of parameters compared to the number

of observations. This yielded convergence failures and unreliable results; therefore, we

decided to run separate analysis for the two questions with essentially different focus.

Results

SRT testing phase

The summary of effects included in the model predicting SRT performance together

with model fit indices are presented in Table 3. Crucially, language proficiency did not

modulate performance on the SRT shown by the non-significant effects of *Group* ($|z$s$|$

≤ 0.77, ps $\geq .442$). Participants made far more errors in the stress contrast ($M = 25.42$ %,

$SD = 15.94$ %) than in the phoneme contrast ($M = 11.20$ %, $SD = 8.19$ %), indicated by

the significant simple effect of *Contrast* ($z = -2.71$, $p < .01$). Incorrect responses started

to increase at longer sequence lengths shown by the significant simple effects of

Sequence length 4 ($z = -2.21$, *p* < .05), *Sequence length 5* ($z = -2.83$, *p* < .01), and

Sequence length 6 ($z = -3.35$, *p* < .001). In addition, the interaction of *Contrast* and

Sequence length was also significant (*Contrast:Sequence length 5*: $z = 1.99$, *p* < .05;

Contrast:Sequence length 6: $z = 2.25$, *p* < .05). We set up a contrast matrix that defined

the comparison of *Contrast* effect (the difference of stress and phoneme contrasts)

between each levels of *Sequence length*. Pairwise tests showed that the probability of an

incorrect response in the stress contrast compared to the phoneme contrast was higher at

length 5 and length 6 than at the baseline length (length 2, see the test statistics above

and in Table 3). In addition, length 6 compared to length 4 yielded even more errors in

the stress contrast than in the phoneme contrast ($z = 2.00$; *p* < .05). Thus, the *Contrast* *

Sequence length interaction indicated that the degree of performance difference between

stress and phoneme contrast slightly increased with longer sequences (see Fig. 1 for

descriptive details), and stress "deafness" was persistent throughout the task.

Table 3. Summary of the GLMM on SRT performance.

Fixed effects	Estimate	SE	z-value	p-value
(Intercept)	8.19	2.13	3.85	<.001
Contrast	**-5.82**	**2.15**	**-2.71**	**.007**
Sequence length 3	-3.63	2.21	-1.65	.100
Sequence length 4	**-4.73**	**2.14**	**-2.21**	**.027**
Sequence length 5	**-6.02**	**2.13**	**-2.83**	**.005**
Sequence length 6	**-7.12**	**2.12**	**-3.35**	**<.001**
Group Int * Phoneme Contrast	-0.04	0.33	-0.12	.906
Group Adv * Phoneme Contrast	-0.25	0.33	-0.77	.442
Group Int * Stress Contrast	0.23	0.36	0.63	.529
Group Adv * Stress Contrast	0.00	0.37	0.01	.991
Contrast * Sequence length 3	3.53	2.25	1.57	.117
Contrast * Sequence length 4	4.00	2.17	1.84	.065
Contrast * Sequence length 5	**4.30**	**2.16**	**1.99**	**.046**
Contrast * Sequence length 6	**4.86**	**2.16**	**2.25**	**.024**
Random effects	Variance			
[Participants] (Intercept)	20.27			
Contrast	21.85			
Sequence length 3	17.16			
Sequence length 4	18.06			
Sequence length 5	18.70			
Sequence length 6	19.54			
Contrast * Sequence length 3	21.26			
Contrast * Sequence length 4	20.06			
Contrast * Sequence length 5	21.49			
Contrast * Sequence length 6	23.20			
[Trials] (Intercept)	0.01			
Fit statistics	AIC	BIC	-2LL	
	3347.8	3793.8	3207.8	

Note. SE = standard error; Int = intermediate learners; Adv = advanced learners; AIC = Akaike information criterion; BIC = Schwarz's Bayesian criterion; -2LL = -2 log-likelihood value. The two

levels of *Contrast* factor were Phoneme and Stress Contrast, the five levels of *Sequence length* factor were 2, 3, 4, 5, and 6 items. As we used the treatment contrast, the baseline was the reference level of a given factor (i.e., Phoneme Contrast and Sequence length 2). Significant effects are in **bold**.
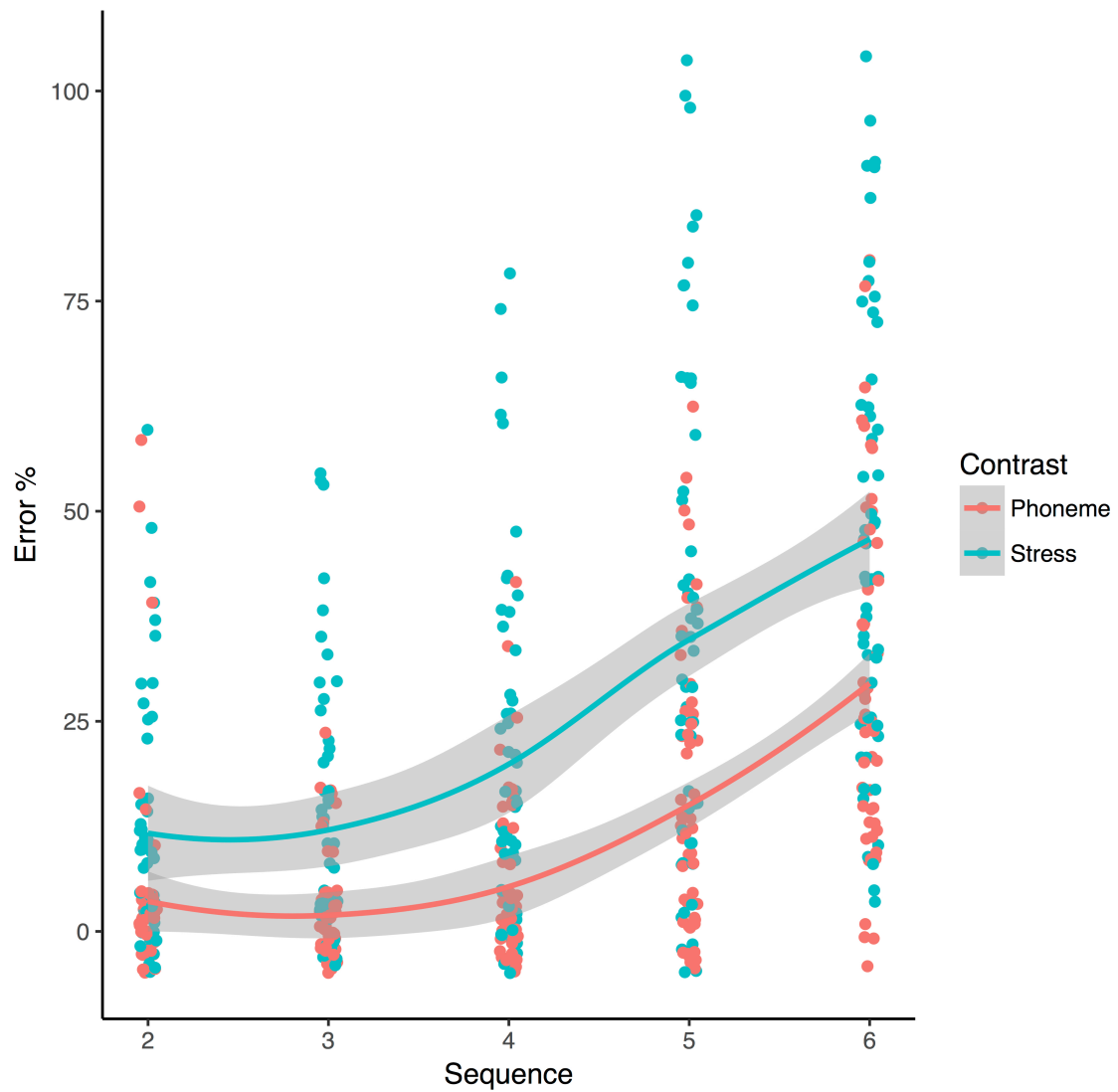

Figure 1. Interaction between *Contrast* and *Sequence length* for error percentage. Points denote individual data points averaged over trials. Note, that in contrast to the data used in the GLMM

analysis, the figure shows error percentages, i.e., higher values indicate more errors. Shaded areas represent 95% confidence intervals for means, depicted with continuous lines.

Cognitive predictors of performance on SRT

The summary of effects included in the model predicting SRT performance as a function of cognitive factors is presented in Table 4. In line with the main findings of the previous model on overall SRT performance without the influence of cognitive factors, analysis revealed significant simple effects of *Contrast* ($z = -4.33$, $p < .001$), *Sequence length 5* ($z = -5.77$, $p < .001$), and *Sequence length 6* ($z = -9.13$, $p < .001$), suggesting that stress contrast and greater sequence length predicted higher probability of errors. In contrast to the results of the previous model, the *Contrast * Sequence length* interaction was not significant ($|zs| \leq 1.53$, $ps \geq .126$) when the effect of the cognitive factors was accounted for. Considering the cognitive factors, a participant with lower *WM* ($z = 4.06$, $p < .001$), higher *susceptibility to interference* (i.e., weaker IC; $z = -2.43$, $p < .05$), and attenuated speed of *phonological processing* (i.e., greater values of PA; $z = -2.21$, $p < .05$) showed lower performance (higher error rate) on the SRT. More importantly, while the interactions of *Contrast * Digit Span* ($z = 0.49$, $p = .623$) and *Contrast * Phoneme Deletion* ($z = -0.52$, $p = .603$) were not significant, the interaction of *Contrast * Stroop* ($z = 2.34$, $p < .05$) was significant. The latter results indicate that while WM and PA similarly affected performance in the phoneme and

stress contrasts, *IC* differentially contributed to performance across the contrasts. As

follow-up analysis of the Contrast * Stroop significant interaction, we defined separate

models for the responses obtained in the *phoneme* contrast and *stress* contrast. These

models included the same variables as the original model with cognitive factors and

were conducted with the same settings described above, except that the *Contrast* factor

and its interactions were excluded. Here we only report statistics for the effect of IC.

While in the *phoneme* contrast of SRT, weaker IC was related to lower performance (z

$= -2.70$, $p < .01$), in the *stress* contrast, IC did not significantly contribute to

performance ($z = -0.49$, $p = .627$).

Table 4. Summary of the GLMMs on SRT performance involving cognitive factors.

Fixed effects	Estimate	SE	z-value	p-value
(Intercept)	3.66	0.29	12.82	<.001
Contrast	**-1.35**	**0.31**	**-4.33**	**<.001**
Sequence length 3	0.43	0.42	1.03	.304
Sequence length 4	-0.46	0.34	-1.35	.177
Sequence length 5	**-1.73**	**0.30**	**-5.77**	**<.001**
Sequence length 6	**-2.65**	**0.29**	**-9.13**	**<.001**
Digit Span	**0.49**	**0.12**	**4.06**	**<.001**
Stroop	**-0.28**	**0.12**	**-2.43**	**.015**
Phoneme Deletion	**-0.26**	**0.12**	**-2.21**	**.027**
Contrast * Sequence length 3	-0.52	0.47	-1.11	.265
Contrast * Sequence length 4	-0.24	0.40	-0.60	.546
Contrast * Sequence length 5	0.05	0.35	0.13	.898
Contrast * Sequence length 6	0.53	0.35	1.53	.126
Contrast * Digit Span	0.05	0.10	0.49	.623
Contrast * Stroop	**0.22**	**0.09**	**2.34**	**.019**
Contrast * Phoneme Deletion	-0.05	0.10	-0.52	.603
Random effects	Variance			
[Participants] (Intercept)	0.39			
[Trials] (Intercept)	0.01			
Fit statistics	AIC	BIC	-2LL	
	3296.1	3410.7	3260.1	

Note. Besides Contrast and Sequence length, indices of working memory (Digit Span), interference suppression (Stroop), and phonological awareness (Phoneme Deletion) were used as predictors. *SE* = standard error; AIC = Akaike information criterion; BIC = Schwarz's Bayesian criterion; -2LL = -2 log-likelihood value. The two levels of *Contrast* factor were Phoneme and Stress Contrast, the five levels of *Sequence length* factor were 2, 3, 4, 5, and 6 items. As we used the treatment contrast, the baseline was the reference level of a given factor (i.e., Phoneme Contrast and Sequence length 2). Significant effects are in **bold**.

Discussion

In the present study, we investigated in speakers of a native language with non-contrastive stress how cognitive factors and L2 proficiency contribute to recalling sequences of increasing length consisting of L2 pseudowords with a phoneme or a stress change. We used generalized linear mixed modelling with binomial distribution on accuracy data obtained in the SRT, because GLMM is suggested to be a more reliable data analysis method than ANOVA for binary outcomes (Jaeger, 2008). This allowed us to investigate individual responses instead of aggregated performance rates in the different sequences.

The main finding of our study is that participants made more errors in the stress contrast than in the phoneme contrast, and with increasing sequence length the difference between the two contrasts slightly increased. This is a clear indication of a strong stress "deafness" effect, corroborating previous results of Peperkamp et al. (2010). Our data also show that the stress "deafness" effect is persistent in Hungarian speakers, similar to French speakers (Dupoux et al., 2008), because we did not find any significant difference in SRT performance between participants learning or not-learning German. All participants had a similar difficulty in recalling sequences with stress contrast regardless of their L2 proficiency, indicating that the impaired processing of stress information remains unchanged even if someone is highly familiar with a language in which the stress contrast is a significant phonological feature. Since we used

pseudowords with the actual acoustic characteristics of an L2 that are supposedly familiar to the participants, our results also show that familiarity with the phonetic properties of stress do not help to overcome the stress processing deficit of L2 learners.

To investigate cognitive components modulating SRT performance over L2 proficiency, we performed GLMM analyses involving measures of WM, PA, and IC. According to the results, higher WM and enhanced speed in phonological processing led to lower error rates in both the phoneme and stress tasks. Since SRT relies on learning and recalling sequences of pseudowords from working memory, the involvement of WM and PA is not unexpected. As we presented in the introduction, WM is known to be crucial in language acquisition and also in second language acquisition (Baddeley, 2015; Baddeley et al., 1998; Dörnyei & Skehan, 2008; Gathercole & Baddeley, 1989, 1990; Miyake & Friedman, 1998). Similarly, PA involves forming and retaining representations of speech stimuli, and it has been found to be involved in word learning ability in L1 and L2 (C.-F. Hu, 2003; X. Hu et al., 2013; M. Snowling et al., 1986; M. J. Snowling et al., 1991). Although both WM and PA contributed to task performance, based on the present results, it is not possible to decide whether they affect stress and phoneme perception in L2, or they are necessary components for solving the SRT.

The appearance of IC as a predictor only in the phoneme task but not in the stress task is a surprising finding. There could be several potential suggestions to

explain why participants with higher IC performed better on the phoneme task but this association was not observed on the stress task. One possible assumption, originating from studies on bilingualism (Bialystok, 2011; Bialystok et al., 2009; Martin-Rhee & Bialystok, 2008), is that participants with higher IC could more efficiently suppress competing L1 phoneme representations to appropriately encode the L2 stimulus sequence in short-term memory. The fact that no effect of IC emerged in the stress task might imply that in contrast to the phoneme task, there are no competing L1-L2 representations in the stress task, because L2 stress is not represented accurately due to the stress "deafness" effect. It might also be suggested that participants with a better IC could inhibit the competing elements of the sequence more successfully, which promoted task solving, but only in the case of phoneme representations. Dupoux et al. (2008) argue that the stress processing impairment of French participants can be explained by psycholinguistic models (Brown, 2000; Flege, Frieda, & Nozawa, 1997) that assume the influence of native abstract phonetic or phonological features in processing L2 features. These models predict that the background of perceptual deficits is that non-native phonetic / phonological contrasts cannot be parsed according to the contrastive features of the native language. This is what happens in the case of stress, which is not a contrastive feature in French or Hungarian.

Alternatively, the differential IC effect could be due to the phonological units to be processed: it might be assumed that the processing of phonemes is more dependent

on IC than the processing of stress. For example, Lev-Ari and Keysar (Lev-Ari & Keysar, 2014) and Lev-Ari and Peperkamp (Lev-Ari & Peperkamp, 2014) found that participants with poorer executive control experienced greater coactivation of lexical elements, suggesting that individual differences in IC can influence phonological representations in speech perception. Previous neurocognitive research showed that phonemes and word stress are processed along separable neural mechanisms (Luo & Poeppel, 2012; Poeppel, 2014); consequently, it might be hypothesized that the processing of phoneme and stress information relies on different cognitive components. In order to resolve this issue, cross-linguistic studies are needed in which the effect of IC is investigated on the phoneme vs. stress SRT performance of listeners not showing the stress "deafness" effect.

In sum, our results indicate that Hungarian speakers show the stress "deafness" effect in the SRT using German pseudowords, and this effect is not modulated by their proficiency in German. We demonstrated that not only performance but also the cognitive mechanisms underlying task solving are slightly different in the phoneme and in the stress contrast of the SRT. These findings may have important implications in language teaching, because they show that speakers of languages with non-contrastive stress might not have the necessary cognitive basis to form the L2 representations; therefore, more emphasis should be put and different methods should be applied in teaching suprasegmental than segmental phonology in the foreign language.

References

Baddeley, A. D. (2015). Working memory in second language learning. In Z. Wen, M. Mota, & A. McNeill (Eds.), *Working memory in second language acquistion and processing* (pp. 17–28).

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–173.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bialystok, E. (2011). Coordination of executive functions in monolingual and bilingual children. *Journal of Experimental Child Psychology*, *110*(3), 461–468. https://doi.org/10.1016/j.jecp.2011.05.005

Bialystok, E., Craik, F., Green, D., & Gollan, T. (2009). Bilingual Minds. *Psychological Science in the Public Interest*, *10*(3), 89–129. https://doi.org/10.1177/1529100610387084

Blomert, L., & Vaessen, A. (2009). *Differentiaal Diagnostiek van Dyslexie: Cognitieve*

analyse van lezen en spellen [Dyslexia Differential Diagnosis: Cognitive analysis of reading and spelling]. Amsterdam: Boom. Boom Test, Amsterdam.

Boersma, P., & Weenink, D. (2007). Praat: doing phonetics by computer (Version 4.5.)[Computer program]. *Retrieved from Http://www.praat.org/.*

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, *24*(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. *Second Language Acquisition and Linguistic Theory*, *1*, 4–64.

Brydges, C. R., Clunies-Ross, K., Clohessy, M., Lo, Z. L., Nguyen, A., Rousset, C., … Fox, A. M. (2012). Dissociable components of cognitive control: An event-related potential (ERP) study of response inhibition and interference suppression. *PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0034482

Dörnyei, Z., & Skehan, P. (2008). Individual Differences in Second Language Learning. In *The Handbook of Second Language Acquisition* (pp. 589–630). https://doi.org/10.1002/9780470756492.ch18

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness." *Journal of the Acoustical Society of America*, *110*(3), 1606–

1618. https://doi.org/10.1121/1.1380437

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent
stress "deafness": The case of French learners of Spanish. *Cognition*, *106*(2), 682–
706. https://doi.org/10.1016/j.cognition.2007.04.001

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use
affects the pronunciation of an L2. *Journal of Phonetics*, *25*, 169–186.
https://doi.org/10.1006/jpho.1996.0040

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological
STM in the development of vocabulary in children: A longitudinal study. *Journal
of Memory and Language*, *28*(2), 200–213. https://doi.org/10.1016/0749-
596X(89)90044-2

Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in
vocabulary acquisition: A study of young children learning new names. *British
Journal of Psychology*, *81*(4), 439–454. https://doi.org/10.1111/j.2044-
8295.1990.tb02371.x

Golden, C. J. (1978). *Stroop Color and Word Test*. Wood Dale, IL: Stoelting Co.

Hu, C.-F. (2003). Phonological memory, phonological awareness, and foreign language
word learning. *Language Learning*, *53*(3), 429–462.

Hu, X., Ackermann, H., Martin, J. A., Erb, M., Winkler, S., & Reiterer, S. M. (2013).
Language aptitude for pronunciation in advanced second language (L2) Learners:

Behavioural predictors and neural substrates. *Brain and Language*, *127*(3), 366–376. https://doi.org/http://dx.doi.org/10.1016/j.bandl.2012.11.006

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, *11*(2), 261–271. https://doi.org/10.1017/S1366728908003416

Koukidis, S. (2003). *Das grosse deutsche Übungsbuch*. Athen: Praxis Verlag.

Lev-Ari, S., & Keysar, B. (2014). Executive control influences linguistic representations. *Memory & Cognition*, *42*(2), 247–63. https://doi.org/10.3758/s13421-013-0352-3

Lev-Ari, S., & Peperkamp, S. (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics*, *41*(5), 320–331. https://doi.org/10.1016/j.wocn.2013.06.002

Lev-Ari, S., & Peperkamp, S. (2014). The influence of inhibitory skill on phonological representations in production and perception. *Journal of Phonetics*, *47*, 36–46.

https://doi.org/10.1016/j.wocn.2014.09.001

Li, Y. syuan, & Chen, S. hui. (2016). Relative effectiveness of phonological and morphological awareness training on L2 word reading in EFL children. *System*, *60*, 93–104. https://doi.org/10.1016/j.system.2016.06.005

Luo, H., & Poeppel, D. (2012). Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. *Frontiers in Psychology*, *3*, 170. https://doi.org/10.3389/fpsyg.2012.00170

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *11*(1), 81–93.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In *Foreign language learning Psycholinguistic studies on training and retention* (pp. 339–364).

Pelham, S. D., & Abrams, L. (2014). Cognitive advantages and disadvantages in early and late bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 313–325. https://doi.org/10.1037/a0035224

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, *38*(3), 422–430.

Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology*, *28*, 142–149.

https://doi.org/10.1016/j.conb.2014.07.005

R Core Team. (2014). R: A Language and Environment for Statistical Computing.

Vienna, Austria.

Rózsa, S., & Kő, N. (2010). A WAIS--IV hazai adaptációjának folyamata és a normák

kialakítása. In S. Rózsa & N. Kő (Eds.), *A WAIS--IV felnőtt intelligenciateszt*

magyar kézikönyve. Hazai tapasztalatok, vizsgálati eredmények és normák. (pp.

29–49). OS Hungary Tesztfejlesztő.

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance.

Language Learning, *19*(3–4), 245–253.

Snowling, M., Goulandris, N., Bowlby, M., & Howell, P. (1986). Segmentation and

speech perception in relation to reading skill: A developmental analysis. *Journal of*

Experimental Child Psychology, *41*(3), 489–507. https://doi.org/10.1016/0022-

0965(86)90006-8

Snowling, M. J., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological

processes. Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied*

Psycholinguistics, *12*, 369–373. https://doi.org/10.1017/S0142716400009279

Tóth, D., Csépe, V., Vaessen, A., & Blomert, L. (2014). *A diszlexia*

differenciáldiagnózisa. Az olvasás és helyesírás kognitív elemzése. Technikai

kézikönyv. [3DM-H: Dyslexia Differential Diagnosis. Cognitive analysis of

reading and spelling][Hungarian]. Kogentum: Nyíregyháza, Hungary.

Wechsler, D. (2008). *Wechsler adult intelligence scale--Fourth Edition (WAIS--IV)*. *San Antonio, TX: NCS Pearson*. San Antonio, TX: NCS Pearson.