The Compound Stimuli Visual Information (CSVI) Task Revisited: Presentation Time, Probability Distributions, and Attentional Capacity Limits

Sergio Morra (morra@nous.unige.it)
Dipartimento di Scienze della Formazione, Corso A. Podestà, 2
Genova, Italy

Lorenzo Muscella (lorenzo.muscella@gmail.com)
Dipartimento di Scienze della Formazione, Corso A. Podestà, 2
Genova, Italy

Abstract

This study examined the validity of the Bose-Einstein (B-E) model of the Compound Stimuli Visual Information (CSVI) task, and its assumptions. Two experiments compared adults' performance on the CSVI task in standard (5 sec) presentation condition and with shorter presentation times. Individual participants' performance was analyzed with the B-E model, with different assumptions on the number of attending acts in each condition. Both experiments found that the capacity limit estimates found in both conditions were highly correlated with each other, and their means did not differ. The goodness of fit of B-E distributions to the data was also tested. It is concluded that the B-E model provides a valid estimate of attentional capacity limits in the CSVI.

Keywords: limited capacity; attention; working memory; Bose-Einstein distribution

Introduction

There is currently widespread agreement that (i) working memory capacity is a major predictor of intelligence, reasoning ability, and other complex skills, (ii) the development of working memory capacity has an important role in many aspects of cognitive development, and (iii) domain-general, capacity-limited attentional resources are a core component of working memory and a basic determinant of its capacity (e.g., Anderson & Lebiere, 1998; Barrouillet, Bernardin & Camos, 2004; Cowan, 2002, 2005; Engle, Kane & Tuholski, 1999; Gathercole & Alloway, 2007; Halford, Wilson & Phillips, 1998; Oberauer, 2002; Vergauwe, Devaele, Langerock & Barrouillet, 2012).

A pioneering article by Pascual-Leone (1970) anticipated long ago these three statements, arguing that the motor of cognitive development throughout Piagetian stages and sub-stages is the maturational increase of a general-purpose resource (called "central computing space" in that article). Pascual-Leone (1970) also suggested that the nature of that resource is a limited amount of attentional energy (or "mental energy", from which the term "M capacity" to indicate the amount of this attentional resource), and that M capacity is at the core of Spearman's general intelligence. Increase of M capacity with age would enable children to activate a larger number of Piagetian "schemes" and, therefore, to construct increasingly complex cognitive structures.

To test his model of M capacity and cognitive development, Pascual-Leone (1970) created the Compound Stimuli Visual Information (CSVI) task and administered it to groups of children of different age. In a training phase of the task, participants acquired a repertoire of specific schemes, by learning to respond to different features of figures (e.g., square, red, dashed contour…) with different gestures (e.g., nod head, raise arm, stand up). When a child had fully learned these S-R pairs or "artificial schemes", the testing phase started; figures with different numbers of relevant features were presented for 5 sec each, and the child's task was to respond appropriately to each feature she could detect. Pascual-Leone assumed that children allocate attention to the task features according to a probabilistic model. Participants were assumed to have a limited capacity k (where k is an integer, increasing with age) that can be used to simultaneously activate no more than k schemes. Participants can attend repeatedly to the stimulus (having k units of capacity available on each attending act); in particular, Pascual-Leone (1970) assumed that after each attending act a participant evaluates whether she observed the stimulus well enough, and after k attending acts the participant feels attentional saturation and stops exploring the figure.[1] Thus, with long stimulus presentation and unlimited response time, a participant will attend to each stimulus figure k times, each time with a capacity of k units, for a total of k^2 available units of capacity. Pascual-Leone (1970) also assumed that the probability distribution of the number x of correct responses to items with n relevant features is a Bose-Einstein (B-E) distribution.

The probability mass function of this distribution (see also Feller, 1968) is:

$$\mathrm{p}\,(x) = \binom{n}{n-x}\binom{r-1}{x-1} \Big/ \binom{n+r-1}{r}$$

and, as a more intuitive metaphor, one can think of r undistinguishable balls thrown to a set of n distinguishable

[1] The assumption that the participant attends to a stimulus k times is clearly a simplification; a capacity of k does not logically imply that also the number of attending acts is k. However, it seems at least plausible that, having in episodic short-term memory (on average) k evaluations of a stimulus, a participant may feel that, with that stimulus, all of the job is done.

boxes, with the random variable x representing the number of boxes that turn out to be occupied by at least one ball. Referring to the CSVI, n is the number of (distinct) relevant features in a stimulus, r is the number of (undistinguishable) units of attentional capacity allocated to the stimulus, which takes the value of k^2 for the reasons given above, and x is the random variable that expresses the number of features detected and responded to.

The developmental theory proposed by Pascual-Leone to account for Piagetian stages claimed that $k = 2$ in typical five-year-olds, and k would increase on average by 1 unit every second year, until a capacity of 7 units (reminiscent of the "magical number seven" of Miller, 1956) is reached during adolescence.

The results of that pioneering study were broadly in agreement with these hypotheses, and a number of other studies, with different methods, also supported this theory of capacity development in childhood and adolescence (see Morra, Gobbo, Marini & Sheese, 2008, for a review). Moreover, subsequent studies, using either the original CSVI task or a computerized version with a special keyboard for responses, yielded results consistent with the B-E model (e.g., Globerson, 1983; Johnson, Im-Bolter & Pascual-Leone, 2003); in our lab we obtained a mean estimate of $k = 6.21$ from an adult sample (Morra, 2015).

Nevertheless, some aspects of the CSVI and its B-E model could be questionable. First, a long exposure of stimuli could afford chunking or rehearsal strategies, which in turn would yield invalid (over-)estimates of capacity (Cowan, 2001). Second, one could wonder how plausible the assumption that, with long stimulus presentation, participants actually attend k times to the stimulus. Finally, great progress has been made in the last decades in the field of methods to assess the goodness of fit of expected to observed distributions; it would be desirable to assess the fit of the B-E distributions with more refined methods than those used at the time of the original study.

This paper aims to assess the validity of the B-E model and its assumptions. In particular, we assess whether brief presentation of the stimuli, followed by a mask, so that the participant can attend to the stimulus only once, yields capacity estimates equivalent to those obtained with the original 5 sec presentation under the assumption of k attending acts. Moreover, we shall evaluate the goodness of fit of B-E distributions to the distribution of correct responses observed in our participants.

Experiment 1

In this experiment we compared two conditions of the CSVI, one with stimuli presented for 5 sec as usual for this paradigm, and the other with a brief presentation of 80 msec followed by a mask. It was obviously expected that more features would be detected with long than brief presentation. The main hypotheses, however, were the following.

First, a parameter k representing the participant's limited attentional capacity (i.e., the number of units available on each attending act) can be estimated in both conditions, assuming that the stimulus is attended to k times in the long presentation condition, for a total of k^2 available units. In the short presentation condition, however, we assume that only one attending act is possible, so that the number of available units is equal to k. Although more features can be detected in the long presentation condition, because k^2 units are available in this condition and only k are available with short presentation, we hypothesize that the mean estimate of k is the same in the two conditions.

Second, if the estimate of k obtained from the CSVI is a valid measure of participants' capacity, then the individual participants' k measures obtained in both conditions should be highly correlated.

Method

Participants A total of 20 adults (18 women and 2 men), all with university education, with a mean age of 22.1 years (s.d. = 2.7) volunteered for this experiment.

Materials and Procedure The CSVI requires participants to respond to multiple features of a visual stimulus by pressing different keys on a special response box. The stimuli were presented on a 15-inch CRT monitor; the participant was comfortably sitting at a viewing distance of approximately 70 cm. The relevant features were square shape, red color, large size, dashed contour, presence of a frame around the figure, presence of an X in the centre, presence of an O in the centre, presence of a bar under the figure, and purple background. The response box had 12 keys, clearly distinguishable by shape and color. Nine keys were associated each to one of the 9 relevant features, two more keys were dummy fillers, and a larger red key was an "enter" key to be pressed by the participant to signal that s/he had finished responding to a trial.

The training stimuli were 72 figures, used to train participants on each of the 9 features; in each set of 8 figures, the intended feature was present in 4 and absent in the other 4. For each feature, the experimenter told the participant which key was associated to it and required the participant to respond. The practice stimuli were 50 figures, including 45 with one relevant feature (5 per each feature) and 5 with no relevant feature. They were used to allow the participant to practice correct responses to each feature, until a criterion of perfect performance was reached. Finally, there were 2 more practice stimuli with 3 features each; one was presented for 5 sec and the other for 80 msec for the participant's response.

There were 56 test stimuli, i.e., 8 trials for each level from 2 to 8, where a "level" is defined as the number of relevant features present in a stimulus. The stimuli were arranged in a pseudo-random fixed sequence, the same for all participants. This sequence was divided into four blocks, each of which included two trials of each level. Each stimulus in the first and third block was presented for 80 msec, followed by a mask, and those in the second and fourth block for 5 sec each; the sets of stimuli presented in each condition were counterbalanced over participants.

Results and discussion Each trial was scored for number of correct responses, i.e., number of correctly detected features. Each participant's total number of correct responses (maximum possible = 140) was computed in both short and long presentation conditions.

The participants' total number of correct responses ranged from 52 to 115 (mean = 85.10, s.d. = 17.56) with short presentation, and from 114 to 138 (mean = 125.40, s.d. = 6.50) with long presentation. The effect of presentation time was significant, $t(19) = 13.59$, $p < .001$. This outcome was expected as quite obvious, and this analysis was carried out merely as a manipulation check, to ensure that short presentation actually reduced participants' ability to detect the relevant features.

The point of actual interest was the comparison of the capacity estimates obtained in both conditions. A participant's vector of mean number of correct responses on levels 2 to 8 was compared with all vectors of expected means generated by the Bose-Einstein model for each value of k from 2 to 9. In the B-E distributions, n was the number of features in each level, x was the number of correct responses ($1 \leq x \leq n$), and r was set as k in the short condition and k^2 in the long condition. The k value that yielded the smallest chi-square was selected as the best fitting estimate, in that condition, of the measure k of the participant's capacity.

As explained above, we assumed k attending acts with long presentation, but only 1 with short presentation. Under these assumptions, the mean estimated value of k was 6.55 (s.d. = 1.50) with long presentation, and 6.90 (s.d. = 2.25) with short presentation. The difference between these means was nonsignificant, $t(19) = -1.02$, $p > .32$.

The implication of this finding seems clear; although fewer relevant features were detected correctly with short presentation, the B-E estimates of attentional capacity in the two conditions were equivalent, provided that adequate assumptions were made on the number of attending acts in each condition. In other words, the different number of correct responses in the two presentation conditions was due to the different number of attending acts in each condition, but the participants' capacity limit remained the same across conditions.

The correlation between the estimates of k obtained with long and short presentation was highly significant, $r(18) = .73$, $p < .001$. This high correlation, together with the nonsignificant difference between the means, strongly suggests that the k estimates obtained in the two conditions measure the same construct. [2]

Experiment 2

This experiment was identical to the previous, with only one change. We replaced the short presentation (80 msec + mask) condition with a condition in which the participant would see the stimulus for three times,[3] each of them with a presentation of 80 msec, followed each time by a different mask. We assumed that, in this condition, the participant would attend to each stimulus exactly three times. Therefore, the long condition (in which we assume, according to Pascual-Leone's task analysis, k attending acts for a total of k^2 units of attentional capacity) can be compared with a triple-short condition, in which we assume 3 attending acts for a total of $3k$ available units.

Method

Participants A total of 20 adults (11 women and 9 men), all with university education, with a mean age of 21.7 years (s.d. = 2.4) volunteered for this experiment.

Materials and Procedure Everything was identical to the previous experiment, except that the short presentation condition was replaced by a triple-short condition, in which the stimulus was presented for three times in a row, each time for 80 msec, and each time followed by a different mask.

Results and discussion The way of scoring and analyzing the data was the same as in the previous experiment, except that, to estimate the amount k of the participant's capacity, in the B-E distributions r was set as k^2 in the long condition and $3k$ in the triple-short condition.

The participants' total number of correct responses ranged from 87 to 129 (mean = 108.55, s.d. = 11.86) with short presentation, and from 115 to 137 (mean = 123.80, s.d. = 5.96) with long presentation. The effect of presentation time was significant, $t(19) = 6.69$, $p < .001$. This shows that, compared with long presentation, also the triple-short presentation actually reduced participants' ability to detect the relevant features.

Assuming k attending acts with long presentation and 3 with short presentation, the mean estimated value of k was 6.25 (s.d. = 1.59) with long presentation, and 6.30 (s.d. = 2.23) with triple-short presentation. The difference between these means was nonsignificant, $t(19) = -.12$, $p > .90$.

The correlation between the estimates of k obtained with long and short presentation was significant, $r(18) = .53$, $p < .02$. Once again, the significant correlation between the two estimates of k, together with the nonsignificant difference between their means, indicates that the k estimates obtained in the two conditions measure the same construct.[4] The

[2] The estimate of a participant's k is based on the distribution of correct responses in the whole task. As a proxy to reliability of measurement, we computed the correlation between the number of correct responses in the first and second half of the task, which was $r = .88$ for short presentation and $r = .68$ for long presentation.

[3] The authors are very grateful to Nelson Cowan for suggesting this experiment.

[4] In this experiment, the correlation between the number of correct responses in the first and second half of the task was $r = .73$ for short presentation and $r = .59$ for long presentation.

findings of the first experiment can be generalized to the comparison between the two conditions of the current one.

Fit of Probability Distributions

Both experiments reported above included conditions in which the manipulation of presentation times ensured that participants could attend to the stimulus only once or three times, respectively. The estimates of *k* obtained with long presentation were equivalent to those obtained, respectively, with short or triple-short presentation; therefore, we can conclude that the assumption of *k* attending acts in the long condition was supported. It can also be concluded that the capacity estimates obtained from the CSVI task with short or long presentation have a similar degree of validity.

However, it remains to examine whether the distribution of correct responses in the CSVI task actually approximates the B-E distribution. In the short and triple-short conditions we only have the data of 20 participants per condition – too few for assessing the form of their distributions. In the long condition, however, we can use the data from 60 people, i.e., 20 from each of the experiments reported above, and 20 more from another similar experiment (Morra & Patella, 2012, Exp.1). Because each participant performed 4 trials per level, with 60 participants we can rely on 240 data points for each of the distributions from level 2 to 8 in the long condition.

The participants were classified according to their *k*-value estimated in the long presentation condition. B-E distributions were generated for all values of n from 2 to 8 and all values of *k* from 4 to 9 (i.e., for the complete range of values found in the participants). Then, expected distributions for the total sample were obtained, for each n, as a weighted average of the distributions for each *k* value, with weights proportional to the number of participants who obtained that *k* value. The goodness of fit of these distributions to the data was assessed by mean of chi-square tests. For these tests, whenever the expected frequency of a value of *x* (i.e., for a certain number of correct responses) was < 1, both the expected and the observed frequencies for that *x* were collapsed with the following value of *x*.

The observed distributions and the distributions predicted from the B-E model are shown in Figure 1. Table 1 presents the goodness of fit of the B-E model (i.e., the chi-squares for the comparisons between observed and expected distributions, along with their probabilities) for each of these distributions.

As an alternative model, to be contrasted with the B-E model, we devised a binomial model. In this binomial model we assumed that, when a stimulus was presented, at least one feature would be detected, and the other features would be detected with a certain probability p, to be estimated from the data. The estimated value of p was .866. This alternative model has some face plausibility, because it makes simple assumptions on dichotomous events (each feature can either be detected or not), but it does not assume limited attentional capacity or indistinguishable units of attentional resources. The goodness of fit of the distributions predicted by this binomial model was tested in the same way as for those predicted by the B-E model.

The distributions predicted from the binomial model are also shown in Figure 1. Table 2 presents the goodness of fit of the B-E model for each of these distributions.

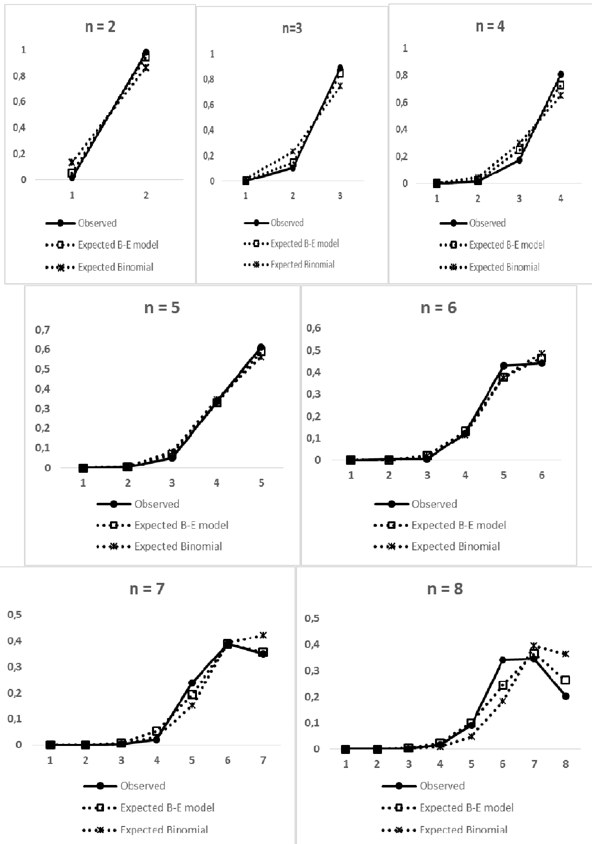

Figure 1: Observed distributions of correct responses (each from 240 data points) and expected distributions from the Bose-Einstein and the binomial models, for each stimulus level from 2 to 8.

Table 1: Goodness of fit of the Bose-Einstein model to the observed distributions.

n	χ^2	d.f.	*p*	
2	6.47	1	.011	*
3	3.39	2	.183	
4	8.34	2	.015	*
5	1.66	3	.647	
6	4.47	3	.215	
7	7.65	4	.105	
8	13.84	4	.008	**

(Note: * p<.05, ** *p*<.01 for the discrepancy between an observed and an expected distribution)

Table 2: Goodness of fit of the binomial model to the observed distributions.

n	χ^2	d.f.	*p*	
2	28.50	1	9 E-8	***
3	25.93	2	2 E-6	***
4	27.25	2	1 E-6	***
5	4.62	3	.202	
6	4.32	3	.229	
7	15.31	3	.002	**
8	63.18	4	6 E-13	***

(Note: * p<.05, ** p<.01, *** p<.001 for the discrepancy between an observed and an expected distribution)

Both Figure 1 and Table 1 indicate that the Bose-Einstein distributions fit the data reasonably well. Four out of seven distributions showed a good fit (p > .1) and in two other cases (n=2 and n=4) the discrepancies between expected and observed distributions, although significant, were actually very small. Only for n=8 there is some notable difference between observed and observed distributions, the observed scores being slightly lower than predicted by the model.

Figure 1 and Table 2 show that the binomial model, instead, did not fit well the data. Only two of the seven distributions fit the data well, and in the other five cases the discrepancies between observed and expected distributions were much larger. Also in the case (n=8) where the fit of the B-E model was least satisfactory, still the B-E model was much closer to the observed data than the binomial model was.

One could still wonder whether the good fit of the B-E model to the data was not an artifact, due to the calculation of a weighted average of six B-E distributions (for the six estimated values of k found in different participants). To check for this possibility we computed, in the same way as above, the goodness of fit of 42 B-E distributions (i.e., 7 values of n times 6 values of k), in order to detect any possible bias or interaction between k values and the fit of the distributions. We do not report here the details of this analysis, but we only mention that, out of 42 tests, only 4 showed a significant (p<.05) discrepancy between the observed and expected distributions. In particular, the participants with k=5 performed better than predicted on level 2 stimuli, and with smaller variance than predicted on level 7 stimuli; the participants with k=7 performed better than predicted on level 4 stimuli; and the participants with k=9 performed better than predicted on level 7 stimuli. No systematic bias or effect for different values of k could be detected, and therefore we can rule out the possibility that there was any artifact due to averaging B-E distributions for different groups of participants.

Conclusions

A detailed comparison between the predicted and observed distributions supported the validity of the B-E model, with parameters n and k^2, for the number of features that participants can detect in stimuli presented for 5 sec. The results of both experiments 1 and 2 showed that the estimates of k obtained for each participant from stimuli presented for 5 sec do not differ from, and correlate highly with, the estimates obtained in conditions of shorter stimulus presentation.

Therefore, all of the results of this study support the view that the B-E model, with the assumption of k attending acts to stimuli presented for 5 sec, with k units of attentional capacity available on each attending act, offers a valid and reliable estimate of the participants' attentional capacity.

The estimated capacity of the participants in both experiments, averaged across experiments and conditions, was 6.5.

This study differed from the original (Pascual-Leone, 1970) because the participants were adults instead of children, the responses were given pressing different keys on a special keyboard, the stimuli were presented for either 5 sec or shorter times, and the testing technology and the statistical tools were more refined than they could be when the original study[5] was carried out. Despite all these differences and, in some cases, methodological refinements, all of the results supported the original B-E model. We can conclude that the CSVI task, analyzed according to the B-E model, can be used reliably to estimate the limits of attentional capacity.

It would be useful to compare the estimates of k derived from the CSVI with the capacity estimates obtained with other procedures, such as complex span or change detection. Complex span tasks (e.g., the counting span; Case, 1985) are often used as working memory measures; a comparison would require modelling the encoding and retrieval operations of the memory task, as well as the capacity demands of the interpolated task. Other researchers (e.g. Cowan, 2001) used a visual array task to derive capacity estimates that are generally lower than the ones proposed by Pascual-Leone, and obtained here. Morra and Patella (2012) suggested that this discrepancy could be explained, noting that Cowan's estimate only considers the declarative information involved in the visual array task, but if one also takes into account the attentional capacity allocated to the procedural information, then the two estimates come closer. Space limitations prevent extensive discussion here of this problem, which will be the topic of a subsequent paper.

References

Anderson, J. R., & Lebiere, C. (1998). The atomic components of thought. Mahwah, NJ: Erlbaum.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working

[5] Pascual-Leone (1970) averaged the distributions of correct responses across levels (i.e., values of n) into a single distribution, and evaluated visually the correspondence between the expected and observed distributions. In this study, we used current techniques to evaluate the goodness of fit of the expected distribution for each value of n.

memory spans. *Journal of Experimental Psychology: General, 133*, 83-100.
Case, R. (1985). *Intellectual development: Birth to adulthood.* Orlando: Academic Press.
Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-176.
Cowan, N. (2002). The search for what is fundamental in the development of working memory. In R. V. Kail, & H. W. Reese (Eds.), *Advances in child development and behavior* (pp. 1-49). Amsterdam: Elsevier.
Cowan, N. (2005). *Working memory capacity*. Hove, UK: Psychology Press.
Engle, R. W., Kane, W., & Tuholski, W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the Prefrontal cortex. In A. Miyake & P. Shah (Eds) *Models of Working Memory* (pp. 102-130). Cambridge: Cambridge University Press.
Feller W. (1968). An introduction to probability theory and its application, vol. I. New York: Wiley.
Gathercole, S. E., & Alloway, T. P. (2007). Working memory and classroom learning. In K. Thurman and K. Fiorello (Eds), *Cognitive Development in K-3 Classroom Learning: Research Applications.* New York: Erlbaum.
Globerson, T. (1983). Mental capacity and cognitive functioning: Developmental and social class differences. *Developmental Psychology, 19*, 225-230.
Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*, 803-831.
Johnson, J., Im-Bolter, N., & Pascual-Leone, J. (2003). Development of mental attention in gifted and mainstream children: The role of mental capacity, inhibition, and speed of processing. *Child Development, 74*, 1594-1614.
Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.
Morra, S. (2015). How do subvocal rehearsal and general attentional resources contribute to verbal short-term memory span? *Frontiers in Psychology,* **6**:145. doi: 10.3389/fpsyg.2015.00145
Morra, S., Gobbo, C., Marini, Z., & Sheese, R. (2008). *Cognitive development: Neo-Piagetian perspectives*. New York: Erlbaum.
Morra, S. & Patella, P. (2012). Modelling working memory capacity: Is the magical number four, seven, or does it depend or what you are counting? *Sixth European Working Memory Symposium (EWOMS-6)*, Fribourg (CH), September 3-5.
Oberauer K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory & Cognition, 28,* 411- 421.
Pascual-Leone J. (1970). A mathematical model for the transition rule in Piaget’s developmental stages. *Acta Psychologica, 32,* 301-345.
Vergauwe E., Devaele N., Langerock N., & Barrouillet P. (2012). Evidence for a central pool of general resources in working memory. *Journal of Cognitive Psychology, 24,* 359-366.