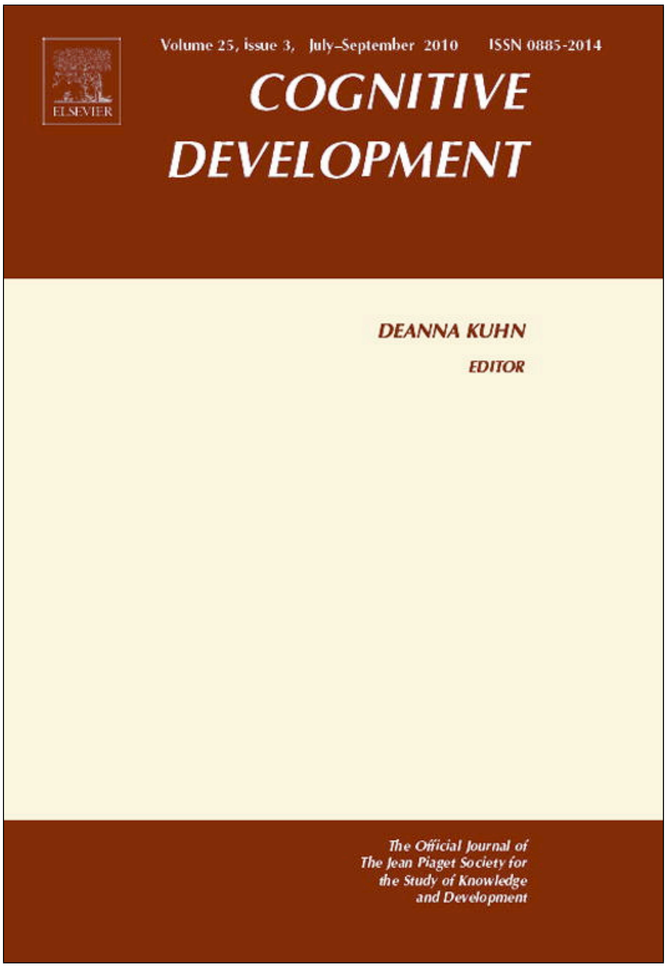
Volume 25, issue 3, July–September 2010 ISSN 0885-2014
COGNITIVE DEVELOPMENT
DEANNA KUHN
EDITOR
The Official Journal of The Jean Piaget Society for the Study of Knowledge and Development

Cognitive Development 25 (2010) 262–277

Contents lists available at ScienceDirect

Cognitive Development

Misleading cues improve developmental assessment of working memory capacity: The color matching tasks

Marie Arsalidou*, Juan Pascual-Leone, Janice Johnson

York University, Department of Psychology, Toronto, Canada

ARTICLE INFO

Keywords:
Working memory capacity
Cognitive development
Contextual influences
Misleading cues
Modified n-back

ABSTRACT

The theory of constructive operators was used as a framework to design two versions of a paradigm (color matching task, CMT) in which items are parametrically ordered in difficulty, and differ only contextually. Items in CMT-Balloon are facilitating, whereas items in CMT-Clown contain misleading cues. Participants of ages 7–14 years and adults ($N = 149$) were studied. We found significant model-predicted graded differences in performance between the facilitating and misleading tasks, across and within age groups, expressing age versus items' demand interactions. Younger children were differentially affected by contextual cues. Even though both task versions were highly correlated with a well-established developmental measure of attentional capacity, CMT-Clown, which contained misleading cues, was a better measure of working memory capacity. These results show a need to estimate degree of misleadingness whenever performance levels in working memory or mental attention tasks are compared and interpreted. Developmental profiles of both tasks are discussed in terms of contextual differences and neoPiagetian stages of development.

1. Introduction

Working memory refers to a set of processes responsible for actively storing and manipulating relevant information (e.g., Engle, 2001). Many researchers see developmental intelligence as expressing maturational change in the capacity of organisms to *actively* store and process information, which is done via schemes or other active organismic units (e.g., Case, 1998; Halford, Wilson, & Phillips, 1998; Pascual-Leone, 1970). Developmental intelligence corresponds to intellectual capacity assessed via

* Corresponding author at: Hospital for Sick Children, Diagnostic Imaging, 555 University Avenue, Toronto, Canada.
E-mail address: marie.arsalidou@sickkids.ca (M. Arsalidou).

0885-2014/$ – see front matter
doi:10.1016/j.cogdev.2010.07.001

problem-solving situations, which grows with chronological age in normal children – an approach pioneered by Binet, Piaget, and many others. There is a growing consensus that developmental intelligence has as main (but not only) organismic causal determinant, an endogenous (and age-bound) growth of *working memory*; more specifically, a growth of its maturational/capacity component sometimes called *mental attention* (Cowan, 2005; Pascual-Leone & Johnson, 2005). Using a neoPiagetian theoretical framework, we investigated how misleading contextual cues affect working memory capacity across development.

Previous studies have examined working memory capacity limitations across development using dual task paradigms (e.g., Cowan et al., 2005; Siegel, 1994; Towse & Hitch, 2008), measures of mental attentional capacity (e.g., Case, 1998; Johnson, Fabian, & Pascual-Leone, 1989; Morra, 2000; Pascual-Leone, 1970; Pascual-Leone & Baillargeon, 1994; Pascual-Leone & Johnson, 2005; de Ribaupierre & Bailleux, 1994), and relational complexity tasks (e.g., Andrews & Halford, 2002; Halford, 1993; Halford et al., 1998). However, developmental estimates of working memory capacity from different researchers are often mutually inconsistent, which suggests that fine *contextual aspects* in the testing situations may not be well understood. An important neglected characteristic of both developmental intelligence (Pascual-Leone & Johnson, 2005) and working memory (Engle & Kane, 2004) is that their assessment requires fine-tuning of contextual aspects, to ensure that their measures constitute *misleading situations* (Pascual-Leone, 1989). This leads to stable, quantitative estimates of working memory, or mental attention, invariant across types of tasks (Pascual-Leone & Johnson, 2005; Pascual-Leone, Johnson, Baskind, Dworsky, & Severtson, 2000).

Indeed, tasks containing irrelevant and conflicting features, or eliciting competition between different plans of action, may be more suitable for a valid, robust assessment of cognitive or intellectual capacity (e.g., Cowan & Morey, 2007; Engle, 2001; Engle & Kane, 2004; Pascual-Leone & Baillargeon, 1994; Pascual-Leone & Johnson, 2005; Pascual-Leone & Morra, 1991). Early individual difference findings on the cognitive effects of *integral* features and embedding contexts in tasks (Garner, 1974; Pascual-Leone, 1969, 1970, 1989; Wertheimer, 1959; Witkin, 1949; Witkin, Dyk, Faterson, Goodenough, & Karp, 1962) have highlighted the importance of conflicting or misleading contexts. Although Piaget in practice used this sort of context very early in his career, he took a long time to conceptualize this aspect. The later Piaget, however, clearly highlighted these conflicting/misleading aspects of tasks by emphasizing the role of dialectical contradictions/negations in development, and by studying elementary forms of dialectics that in children serve to resolve them (Pascual-Leone, in press; Piaget, 1974, 1980). Nonetheless a direct comparison between working memory capacity tasks that only differ in context has not been done in children. Contextual integral features that hinder performance are associated with the concept of misleadingness (Pascual-Leone, 1969, 1970, 1989). *Misleading situations* elicit schemes undesirable for the task at hand, lowering probability of success; they may contain integral albeit task irrelevant stimuli that elicit misleading schemes; or may elicit unwanted plans (executive schemes) that compete for application. In contrast, *facilitating and neutral situations* are minimally misleading if at all, and may elicit schemes relevant for the task that facilitate its solution (Pascual-Leone & Johnson, 1991, 2005).

Our present study investigated the role of misleading-versus-facilitating contexts in the design of valid mental attention tasks (to assess innate/developmental constituents of working memory). We pursued this goal by designing two tasks that share the same presentation protocol (a color matching task) and have same executive plan and relevant dimension of variation (i.e., working memory demand) within and across levels of items. The tasks differ, however, in that items of one task involve clearly misleading cues whereas those of the other do not, which justifies calling the tasks, vis-à-vis each other, misleading versus facilitating. We used an explicit method of mental-process analysis (metasubjective task analysis, Pascual-Leone & Johnson, 2005) to investigate differences between misleading versus facilitating (or neutral) tasks in working memory assessment.

Design of the paradigm was guided by previous research, and by the neoPiagetian Theory of Constructive Operators (TCO; Pascual-Leone, 1970; Pascual-Leone & Johnson, 2005). According to TCO, mental attentional capacity grows on average by one symbolic unit every 2 years after the age of 3, reaching an average of seven units at 15–16 years and in adults (Pascual-Leone, 1970; Pascual-Leone & Johnson, 2005). Thus, on average, the maximum number of symbolic schemes (i.e., processing units) that age groups are expected to simultaneously retain with mental attention (here called *M*-capacity) is 1, 2, 3, 4, 5, 6, and 7, respectively, in years 3–4, 5–6, 7–8, 9–10, 11–12, 13–14, and 15-plus years. This is consistent with Miller's (1956) 7 ± 2 "magic number" for adults, and corresponds to Piaget's developmental stages, from which our *M*-capacity model was developed (Pascual-Leone, 1970).

Pascual-Leone and colleagues have developed a number of measures that assess mental attentional capacity across development; they are called *M*-tasks (for examples see Pascual-Leone & Johnson, 2005). We use *M*- to stand for *mental attention* in the context of related assessment constructs (e.g., *M*-capacity, *M*-demand, and *M*-scores). Performance scores on these tasks are called *M*-scores; they serve to estimate the capacity of participants' mental attention (i.e., *M*-capacity). Task characteristics required to appraise levels of *M*-capacity include the following: (a) parametric scaling of levels of complexity of items; (b) invariant executive demand across item levels; (c) minimal previous knowledge requirements – pre-training often is used to ensure that needed information is well acquired; (d) minimal language or conceptual requirements (Pascual-Leone et al., 2000); (e) preferred use of task items that constitute misleading situations, because in these situations contextual (often perceptual or learning) factors cause participants to commit errors unless they apply mental attention. In this study, some well-established developmental results of Pascual-Leone's *M*-capacity measurement (Pascual-Leone & Johnson, 2005, in press) are used to compare the two new *M*-tasks, one misleading and the other facilitating. We expected the misleading task to be a better *M*-measure because misleading contexts improve assessment of mental attention. In what follows we present the new tasks in detail. We then contrast behavioral results, within and across age groups, to investigate which sort of context leads to a better measurement of *M*-capacity.

1.1. *Paradigm design*

In the color matching tasks (CMT) we adapted an updating-task design known as 1-back task (e.g., Owen, McMillan, Laird, & Bullmore, 2005). In this timed paradigm participants see, one item by one, a series of items; and for each one they must indicate whether relevant features (or schemes) of the current item match those of the immediately preceding one. In our tasks we varied the number of features/schemes (the relevant colors) needed to determine a match. Thus, while keeping *executive demand* constant (i.e., number and kind of operations needed to solve the item), the *M-demand* was varied across classes of items by adjusting the number of relevant colors. Number of relevant colors, which varied from one to six (see Section 2 for details), determined the working memory complexity level of an item.

Participants indicated whether or not the current (i.e., *target*) item had the same set of *relevant* colors as the previous (i.e., *criterion*) item, irrespective of color location. Blue and green were *irrelevant* colors to be ignored, as were the facial features of the stimulus. Presented in different item blocks were classes of items that varied in difficulty level expressed in terms of *M*-demand. These levels spanned the expected *M*-capacity of our participants (7- to 14-year-olds and adults).

To examine contextual influences, we contrasted a task comprised of sets of balloons (CMT-Balloon) with one comprised of figures of a clown (CMT-Clown). CMT-Balloon is *facilitating*, because task relevant features (i.e., colors) of the stimuli are salient, being part of segregated shapes with relatively constant form and size, which are best identified by their color. The constant bundle of balloons was not salient as an integral whole. In contrast, CMT-Clown, a *misleading* task, contains an attractive *integral whole*, the clown itself, *which is very salient but irrelevant to the task – a "frame"* in which relevant colors are embedded. Further, costume parts (e.g., shoes, hat, patch) are attractive distractors, variable in shape and size, contiguous, and task irrelevant, making search for colors in the clown attentionally demanding. These contextual differences should influence the *M*-demand at every item level of the two CMT versions. We have estimated this demand using mental/metasubjective task analyses (MTA).

MTA, a method developed by Pascual-Leone and colleagues (Pascual-Leone & Baillargeon, 1994; Pascual-Leone & Goodman, 1979; Pascual-Leone & Johnson, 1991, 2005, in press), can be used to assess procedures in a task in terms of demand for *M*-capacity (i.e., *M*-demand). This method assumes a strategy for task solution, and assumes that relevant schemes are available to participants; such schemes relate to training provided and the participants' likely past experience. The task solution strategy we model here is not the only possible; but is simple and commonly adopted by participants. Our method of task analysis can model other strategies; but the chosen strategy had little executive demand and its expected results were consistent with other findings using misleading *M*-tasks (e.g., Pascual-Leone & Baillargeon, 1994; Pascual-Leone & Johnson, 2005). To perform the CMT (see Appendix A), participants carry out a sequence of mental steps, in each of which operative (i.e., procedural) schemes apply on figurative (representational – e.g., object, color) schemes to produce an outcome. Participants *scan* and *identify* colors in the current figure, in order to compare and *match* the *target* (i.e., currently relevant) color *set* with the *criterion* (i.e., previous-item) *set* of colors; and then *press* an appropriate key to signify match/mismatch. An overall executive strategy in the CMT embodies the executive goal and operations (both learned during training). Because these operations occur in sequence they are expected to account for one unit of *M*-demand during problem solving. This is both for CMT-Balloon and CMT-Clown.

Number and kind of irrelevant cues causes the lesser difficulty of CMT-Balloon versus CMT-Clown, but the relative difficulty of item classes in both tasks depends on number of relevant colors in the item. For both tasks irrelevant cues are colors blue and green, as well as location or repetition of relevant colors. For CMT-Clown additional irrelevant cues include the salient figure of a clown with its charming body parts, and colors in the clown's face. Because relevant colors are contextually embedded (integral features) in the clown figure, participants must *actively extract* one by one the relevant colors to check for a possible match with colors of the previous clown. This extraction process is not automatized, and therefore, it adds a second operative unit to the *M*-demand for solving CMT-Clown, making it one unit higher than that of CMT-Balloon. Other relevant schemes that require mental attention are the colors in the criterion set (from previous item). In CMT-Balloon (facilitating task) participants must boost with *M*-capacity one operative scheme plus the schemes for the *n* colors in the criterion set, resulting in an *M*-demand of $n+1$. In CMT-Clown (misleading task) they must boost two operative schemes, resulting in an *M*-demand of $n+2$.

In both CMT tasks executive schemes and main action schemes are constant across classes of items; what varies with each item class (each *M*-demand level) is the number *n* of relevant colors. Thus, the two CMT versions differ only in contextual characteristics, which in the CMT-Clown would require additional cognitive resources. Previous research in adults (Engle, 2001; Engle & Kane, 2004) and children (Pascual-Leone & Baillargeon, 1994; Pascual-Leone et al., 2000) demonstrated that interfering features in a task might improve assessment of working memory. If misleading cues have an integral function in the task, we predict that items with *n* (e.g., 3) relevant colors will have *M*-demand of $n+1$ (e.g., 4) in CMT-Balloon, but *M*-demand of $n+2$ (e.g., 5) in CMT-Clown. Details of the MTA procedure with an operator-logic formulation of key mental steps for each CMT version can be found in the Appendix A.

1.2. Hypotheses

Our first prediction regarding context effects was that CMT-Clown will be more difficult than CMT-Balloon, due to the misleading factors that function as moderator variables of the *M*-demand effect. We have also asserted that misleading contexts are better for measuring *M*-capacity or working memory, in contrast to facilitating contexts. We evaluated this claim explicitly by comparing developmental patterns of performance on CMT-Balloon and CMT-Clown against theoretical predictions, as well as against another established *M*-task (FIT, figural intersections task; Pascual-Leone & Baillargeon, 1994). Individuals should not be able to reliably pass an item unless their *M*-capacity is at least equal to the item's *M*-demand (Pascual-Leone, 1970; Pascual-Leone & Baillargeon, 1994; Pascual-Leone & Johnson, 2005). This trade-off between participants' *M*-capacity and item *M*-demand leads to prediction of a "stepwise" developmental growth pattern in performance. Pascual-Leone and Baillargeon (1994) explain the model-theoretical importance of this "stepwise" developmental pattern. Such a pattern should be found only in misleading context tasks. Thus, if the predicted pattern is found in CMT-Clown, but not CMT-Balloon, it would quantitatively show the importance of misleading situations for the design of working memory or mental attention tasks. We anticipated that CMT-Clown would yield a developmental performance pattern consistent with predictions, regarding growth of *M*-capacity over the age-range tested (Pascual-Leone, 1970); and also that performance on CMT-Clown will correspond closely to that on FIT, a very different sort of *M*-task, whereas performance on CMT-Balloon will diverge from theoretical predictions and from FIT performance.

2. Methods

2.1. Participants

Data were collected from children in an urban public school in a middle income neighbourhood in Toronto (Hulchanksi, 2007). Ninety-eight percent of the school's population had been living in Canada for more than 2 years. Age groups studied were five, ranging from 7 years 3 months to 8 years 3 months ($n = 26$), 9 years 3 months to 10 years 2 months ($n = 31$), 11 years 3 months to 12 years 3 months ($n = 28$), and 13 years 4 months to 14 years 3 months ($n = 27$); and adults: 18 years 10 months–23 years 2 months ($n = 37$). Adult volunteers were undergraduate university students receiving partial course credit for participation.

2.2. Materials

2.2.1. Figural intersection task (FIT)

The FIT contains geometric shapes (2–8), presented separately on the right side of each test page and overlapping on the left side (Pascual-Leone & Baillargeon, 1994). Participants were asked to attend to every shape on the right, and then to locate the total intersection of these shapes in the compound figure on the left. In an item the number of *relevant shapes* (i.e., matching those found in the right field) corresponds to its *M*-demand. Seven levels of *M*-demand are presented in 42 randomly ordered items. Two of the six items at each level contained, on the left, three irrelevant shapes to be ignored, and for these items *M*-demand is one unit higher. For example, an item with just three relevant shapes has a demand of three; but if in addition three irrelevant shapes are present, the *M*-demand increases to four because these three shapes must be disregarded together. Training is provided to teach participants appropriate strategies for solving FIT items, such as how to identify similar shapes irrespective of size and rotation in a series of examples. For instance, the experimenter shows children two upright acute triangles which are rotated and overlapping on the other side of the page, and explains that even though the shapes are rotated, they are still the same shapes.

2.2.2. Color matching tasks (CMT)

The CMT has two versions: CMT-Balloon, which uses as template figure a set of balloons, and CMT-Clown, which uses a clown figure. Both figures (approximately 11.2 cm width × 12.5 cm height) have different parts colored-in. Each difficulty level was presented in blocks of eight sequentially given stimuli. Participants were asked to indicate whether the current figure contained the same set of relevant colors as those in the previous figure, irrespective of location. In the case of the clown, they also were told to ignore the face. Location of colors changed from criterion to target items. Relevant colors included red, yellow, purple, grey, brown, pink, and orange; irrelevant colors were blue and green. Without exception, all children and adults were able to identify and name these colors. There was a 50% probability of the correct response to be 'same'. In the case when the correct answer was 'different' the color combinations changed by one color (92% of changes) or two colors (8% of changes) in both versions. Irrelevant colors blue and green also were equally and randomly distributed in both tasks.

A total of 192 stimuli (168 trials + 24 baseline) were presented in two successive runs of 12 32s blocks, each block containing 8 stimuli. Difficulty levels followed a pseudo-random order within each run. Blocks of baseline stimuli ($N = 24$, 2s each) were figures colored only blue or green (task irrelevant colors) interleaved with task blocks. Participants had 3s to view a figure and make a response, followed by 1s inter-stimulus interval, during which a + sign was presented. They completed the CMT on a notebook personal computer (Toshiba, model: PTM30C-K0101E) and had to press (.) or (/) on the keyboard, to indicate whether or not colors matched. Stimuli were controlled and responses recorded using the software Presentation (v. 10.1.09.26.06, Neurobehavioral Systems Inc.).

During training, participants were seated approximately 60 cm from the screen and were introduced to the colors and their relevance, followed by a demonstration of appropriate strategies, like ignoring spatial location of the colors and the face of the clown. Each CMT version was about 14 min long, with a break in between. We presented CMT-Balloon first to serve as practice for the harder

CMT-Clown. Because both tasks are short it was unlikely that children would under-perform on CMT-Clown due to fatigue. Accuracy and latencies were recorded on both versions, however here we only report accuracy scores.

2.3. Procedure

Informed consent was acquired from all participants or the parents. Adults received FIT followed by CMT in a single session. Children received two sessions over a 6-week period. In the first session they were trained and completed CMT individually on the computer. After all children were tested with session one, on a different day they completed FIT in their classroom.

3. Results

3.1. Data scoring and screening

Items were scored as correct or incorrect. An empirical raw proportion score was calculated, using the total sum correct divided by the total number of items. The *M*-score corresponds to the *M*-demand of the highest level passed reliably (i.e., 70% correct for CMT and 66% for FIT), given similarly reliable performance on lower levels, allowing for only one lower level to fall below criterion.

In the CMT versions trials were identified as correct response, incorrect response and time-out (when a response was not given within 3s). Fig. 1 shows the proportion of trials for each category. Significant differences from 0.5 were examined for each age group at each *M*-demand level; circles signify performance at chance level (Fig. 1). Notice that in the CMT-Clown there are seven data points inside the circles, whereas in the CMT-Balloon there is only one. Items with high *M*-demand are clearly more difficult for younger children in the former than in the latter task.

Prior to statistical analysis, data were screened for outliers on all measures. Outliers were defined as scores 2 or more *SD*s away from the mean of each age group for each task. Outliers for each task were removed when the task was included in the analyses (the retained data covered 95% of the scores of each age group). This was done to ensure that results express central tendencies in these data – since here data serve to appraise effects of misleading versus facilitating cues in performance relative to theoretical predictions. For *M*-scores, there were five outliers associated with CMT-Balloon, five with CMT-Clown, and seven with FIT. A total of 16 participants exhibited *M*-score performance that

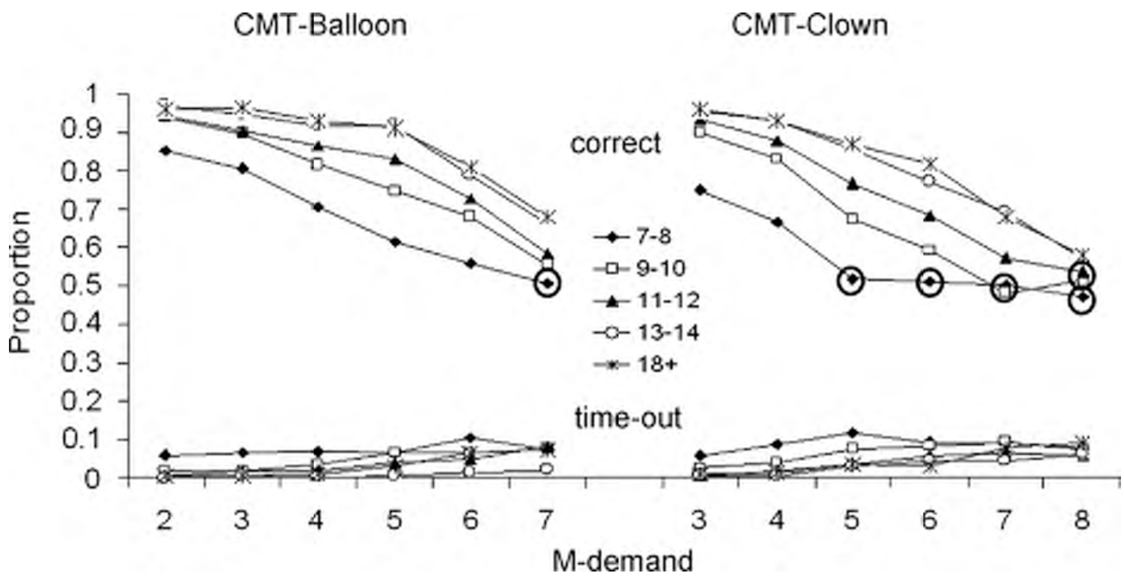

Fig. 1. Mean performance for proportion correct responses and time-outs. Correct responses occupy the higher proportions of the graph, whereas time-outs (i.e., response not given within 3s) are at the bottom showing less variability across *M*-demand levels. Circled values represent mean scores that were not significantly different ($p < 0.05$) from 0.50 (i.e. chance responding). No outliers were removed.

Table 1
Proportion of participants passing CMT items, as a function of age group (Theoretical *M*-capacity) and item difficulty (Predicted *M*-demand).

CMT-Balloon	# of colors →	1	2	3	4	5	6
Age	*M*-demand →	2	3	4	5	6	7
7 - 8	*M*-capacity = 3	0.920	0.920	**0.600**	0.240	0.080	0.000
9 - 10	*M*-capacity = 4	1.000	1.000	0.839	**0.710**	0.484	0.129
11 - 12	*M*-capacity = 5	1.000	0.962	1.000	0.923	**0.731**	0.231
13 -14	*M*-capacity = 6	1.000	1.000	0.963	0.963	**0.815**	0.222
18+	*M*-capacity = 7	1.000	1.000	1.000	0.971	**0.943**	0.514
CMT-Clown	# of colors →	1	2	3	4	5	6
Age	*M*-demand →	3	4	5	6	7	8
7 - 8	*M*-capacity = 3	**0.692**	0.462	0.077	0.038	0.038	0.000
9 - 10	*M*-capacity = 4	0.966	**0.793**	0.310	0.103	0.034	0.069
11 - 12	*M*-capacity = 5	1.000	0.929	**0.679**	0.429	0.071	0.107
13 -14	*M*-capacity = 6	0.926	0.926	0.889	**0.741**	0.519	0.111
18+	*M*-capacity = 7	0.973	0.973	0.892	**0.865**	0.432	0.216

Note: Proportions indicate participants in a given age group who passed at least 70% of items in a given item level. # of colors is the number of relevant colors in an item level. *M*-capacity is the theoretical capacity of the age group. Bolded proportions highlight the highest *M*-demand level passed by at least 60% of participants. The highlighted staircase represents the theoretically predicted cut-off for each age group. Outliers were removed for each task.

exceeded 2 *SD*s on at least one of the tasks. Similar criteria were used for screening raw proportion scores, leading to identification of 13 outliers. In addition, six 13–14 year olds passed items of higher difficulty levels on FIT but failed lower ones (which violates theoretical assumptions); on the CMT versions their *M*-scores ranged between 5 and 7. The six participants were dropped from analyses that included FIT *M*-scores and raw proportion scores. Thus, 19 FIT scores for 13–14 year olds remained.

3.2. Stage-wise patterns in Clown and Balloon CMT

Items in CMT can be classified into levels using their predicted *M*-demand, and so can age groups in terms of their predicted *M*-capacity. Participants should not be able to reliably pass item levels beyond their capacity. Table 1 shows the proportion of participants in each age group who passed at least 70% of items at each CMT item level (i.e., proportion of students satisfying a 70% criterion). These data allow us to evaluate the predicted *M*-capacity/*M*-demand trade-off in performance, that is, whether success in an item occurs only when *M*-capacity is equal to or larger than the item's *M*-demand, as has been empirically demonstrated with FIT (Pascual-Leone & Baillargeon, 1994). The bolded proportions indicate the highest *M*-demand level passed by at least 60% of participants in an age group (a proportion that represents the majority of students). The bolded numbers generally fall one level above the predicted cut-off for CMT-Balloon, but (with one exception) within the cut-off for CMT-Clown. These data show that CMT-Clown behaves as predicted for *M*-measures and that

CMT-Balloon was easier than theoretically predicted by the model of CMT-Clown (see Appendix A), particularly for the two younger age groups. For instance, the majority of 7–8 year olds could only solve clown items with one color, however they could effectively solve balloon items with up to three colors. Although both tasks show an age-related pattern, the *M*-capacity/*M*-demand trade-off pattern is clearer with CMT-Clown.

3.3. Analysis of M-scores and raw proportion scores

For each task, participants were assigned an *M*-score corresponding to the *M*-demand of the highest item level they passed reliably, and a raw score corresponding to proportion of items passed. Mean scores are plotted in Fig. 2 and show age-related changes in performance. These observations are consistent with the results reported in Table 1. Because *M*-scores are rescaled scores taking into account predicted item *M*-demands, contrasting performance in CMT-Clown with CMT-Balloon illustrates the effect of contextual (misleading versus facilitating) features. Indeed, raw proportion score performance was higher on CMT-Balloon at all age levels. We conducted statistical analysis to confirm these observations.

First, a 5 (age) × 3 (task: CMT-Balloon, CMT-Clown, FIT) ANOVA with repeated measures on the task factor was used to examine differences among the *M*-score means as a function of age group and task. Mauchly's test of sphericity was significant (chi-square = 32.49, $p < 0.05$), and so degrees of freedom were corrected using Huynh–Feldt estimates of sphericity (epsilon = 0.85). Results showed a significant interaction of Age × Task ($F_{(6.76, 206.33)} = 7.22$, $MSE = 1.11$, $p < 0.001$, *partial* $\eta^2 = 0.19$), a within subjects effect of task ($F_{(1.69, 206.33)} = 6.29$, $MSE = 1.11$, $p = 0.004$, *partial* $\eta^2 = 0.049$), and a between subjects effect of age group ($F_{(4,122)} = 95.71$, $MSE = 1.57$, $p < 0.001$, *partial* $\eta^2 = 0.76$). The source of variability among tasks and within each age group was followed up with multiple pairwise comparisons corrected using the Bonferroni method. We refer to Fig. 2a for descriptive data. CMT-Clown did not differ from FIT in any age group, supporting the close correspondence between *M*-scores obtained on these two tasks: the difference $M_{\text{Clown}} - M_{\text{FIT}}$ equalled 0.33, −0.14, 0.70, 0.21, and −0.50 for children (7–8, 9–10, 11–12, 13–14 year olds) and adults, respectively. The pattern for the Balloon task was much less consistent. Compared to CMT-Clown, performance on CMT-Balloon was significantly higher for 7–8 and 9–10, and significantly lower for 13–14 year olds. CMT-Clown scores did not differ from CMT-Balloon scores for 11–12 year olds and adults. Compared to FIT, performance on CMT-Balloon was significantly higher for 7–8, 9–10, and 11–12 year olds, and significantly lower for adults. FIT and CMT-Balloon scores did not differ for 13–14 year olds. Given the established validity of FIT, these results suggest that CMT-Clown is a valid *M*-capacity task, whereas the CMT-Balloon may not be. Thus, differences between CMT-Clown and CMT-Balloon support our predictions on contextual influences, particularly evident for younger age groups.

To examine predicted growth of *M*-capacity with age, mean differences (*MD*) in performance were compared among age groups within each task, using Bonferroni correction. For the FIT, each age group differed significantly from the others (*MD* ranged from −1.36 to −4.30), with the exception of 9–10s versus 11–12s ($MD = -0.59$), and 13–14s versus adults ($MD = -0.75$). For CMT-Clown, each age group differed significantly from the others (*MD* ranged from −0.89 to −3.47), with the exception of 13–14s versus adults ($MD = -0.03$). CMT-Balloon yielded a less distinctive developmental growth pattern: 7–8s performed lower than all other age groups, and 9–10s scored lower than adults; but the other age comparisons were non-significant.

Unlike CMT-Balloon, the developmental performance pattern on CMT-Clown showed "stage-wise" increase with age (Table 1, Fig. 2a). This is consistent with our prediction that passing performance will not be reached in items of a certain *M*-demand unless the age-group's *M*-capacity is equal to or larger than this demand. With effect sizes greater than Cohen's $d = 0.9$ for all significant comparisons between consecutive groups of children, age-related differences in performance are highly detectable in CMT-Clown.

Second, the same 5 (age) × 3 (task: CMT-Balloon, CMT-Clown, FIT) ANOVA was used to examine differences among the *raw proportion score* means with repeated measures on the task factor. Mauchly's test indicated that the assumption of sphericity was violated (chi-square = 91.64, $p < 0.05$); therefore, degrees of freedom were corrected using Huynh–Feldt estimates of sphericity (epsilon = 0.68).

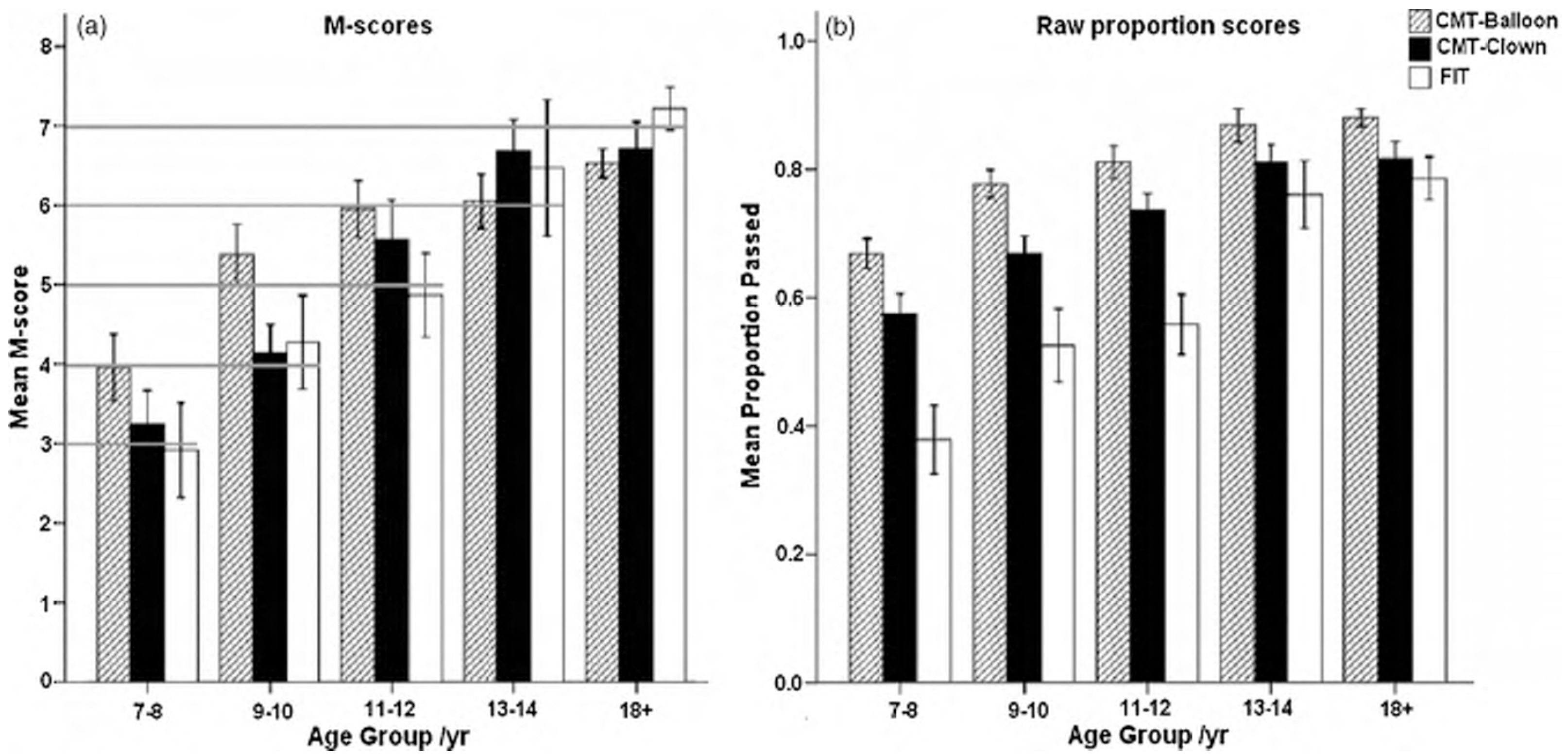

Fig. 2. a) Mean *M*-score and ±2 standard error bars as a function of task and age group (grey horizontal lines indicate the theoretical *M*-score for the different age groups). b) Mean raw proportion score and ±2 standard error bars as a function of task and age group. All outliers were removed.

Table 2
Correlations between all tasks.

Task	CMT-Balloon	CMT-Clown	FIT	Age
	(a) *M*-scores			
CMT-Balloon	–	0.72**	0.58**	0.63**
CMT-Clown	0.50**	–	0.65**	0.72**
FIT	0.24*	0.27*	–	0.72**
	(b) Raw proportion scores			
CMT-Balloon	–	0.86**	0.70**	0.70**
CMT-Clown	0.73**	–	0.69**	0.68**
FIT	0.40**	0.40**	–	0.72**

Note: First order Pearson *r* correlations appear above the diagonal (*M*-scores $N = 127$; Raw proportion scores $N = 130$). Partial correlations with age variance removed appear below the diagonal. All outliers were removed.
** $p < 0.001$
* $p < 0.01$.

Results showed a significant interaction of Age × Task ($F_{(5.458,\ 170.575)} = 11.47$, $MSE = 0.008$, $p < 0.001$, partial $\eta^2 = 0.27$), a within subjects effect of task ($F_{(1.365,\ 170.575)} = 232.90$, $MSE = 0.008$, $p < 0.001$, partial $\eta^2 = 0.65$), and a between subjects effect of age group ($F_{(4,\ 125)} = 95.084$, $MSE = 0.011$, $p < 0.001$, partial $\eta^2 = 0.75$). The largest effect was for age, similar to the *M*-score ANOVA. Bonferroni correction was used when conducting multiple pairwise comparisons.

Follow-up tests within age groups showed that mean proportion scores differed between tasks (*MD* ranged from 0.058 to 0.29) with two exceptions (CMT-Clown and FIT scores were not significantly different for 13–14 year olds and adults). Comparisons among age groups showed that, for both FIT and CMT-Balloon, 7–8 year olds performed significantly lower than 9–10 year olds (*MD* = 0.11), and 11–12 year olds lower than 13–14 years olds (*MD* = 0.058). On CMT-Clown, consecutive age groups scored differently from each other (*MD* ranged from 0.067 to 0.095), with the exception of 13–14 year olds and adults. Thus *M*-scores and raw proportion scores yielded theory-consistent developmental patterns for FIT and CMT-Clown, but not for CMT-Balloon.

Third, we examined correlations among *M*-scores (Table 2a) and among raw proportion scores (Table 2b). Partial correlations with age variance removed remained significant for both scoring methods. The three tasks were correlated with each other as well as with chronological age. As expected, the strongest correlation was between the two CMT versions, with 52% of shared variance for *M*-scores, and it was even higher for raw proportion scores. Consistent with already examined data using *M*-scores, FIT and CMT-Clown shared more variance (43%) than did FIT and CMT-Balloon (34%), although the latter is still substantial. Using *M*-scores, the correlation between Clown and FIT when Balloon was partialed out was 0.41 ($p < 0.001$); correlation between Balloon and FIT with Clown partialed out was 0.21 ($p = 0.016$); and correlation between Clown and Balloon partialing out FIT was 0.55 ($p < 0.0001$).

M-capacity grows with age, yet correlations remained significant when age variance was removed (Table 2). The significant partial correlations likely are due to the method variance of *M*-measures, as well as to related executive learning and individual differences in *M*-capacity.

3.4. Equivalence tests

Researchers in psychology often predict differences between means, and statistical tests usually check for significant differences. In contrast, tests of equivalence are more stringent: they examine whether means are statistically indistinguishable (Wellek, 2003). We anticipated that *M*-scores on CMT-Clown would be indistinguishable from both FIT *M*-scores and the theoretically predicted *M*-capacity values; and predicted that such close correspondence would not be found for CMT-Balloon. To examine these predictions, paired *t-tests for equivalence* were conducted among the tasks and the

Table 3
Equivalence tests among *M*-scores.

Age/yrs	CMT-Balloon^{BF}		FIT			CMT-Clown^{CF}
7–8	3.96	≠	2.92		=	3.25
9–10	5.38	≠	4.28		=	4.14
11–12	5.96	≠	4.87		≠	5.57
13–14	6.05	≠	6.47		=	6.68
18+	6.53	=	7.22		=	6.72
	CMT-Balloon^{BF}		M-Capacity			CMT-Clown^{CF}
7–8	3.96	≠	3		=	3.25
9–10	5.38	≠	4		=	4.14
11–12	5.96	≠	5		≠	5.57
13–14	6.05	=	6		≠	6.68
18+	6.53	=	7		=	6.72
	CMT-Balloon^{CB}		CMT-Clown	FIT^{CF}		M-Capacity
7–8	3.96	≠	3.25	2.92	=	3
9–10	5.38	≠	4.14	4.28	=	4
11–12	5.96	=	5.57	4.87	=	5
13–14	6.05	≠	6.68	6.47	≠	6
18+	6.53	=	6.72	7.22	=	7

Note: $p = 0.05$. Expected mean difference of 1 was used to calculate the equivalence interval values; = signifies equivalence; ≠ signifies non-equivalence. Values represent mean *M*-scores. Superscripts refer to the standard deviation difference used to calculate the obtained *t*-statistic (*t*-obt): BF stands for CMT-Balloon and FIT; CF stands for CMT-Clown and FIT; CB is for CMT-Clown and CMT-Balloon. If t-obt < t-crit (critical t-statistic), the means are statistically equivalent. All outliers were removed.

predicted *M*-capacity for each age group (Wellek, 2003). To our knowledge this is the first report utilizing paired t-tests of equivalence in psychological research.

M-scores attained by each age group represent an average that should have a standard deviation of at most one. We used a maximum expected mean difference of one to calculate equivalence intervals (Table 3). Overall, CMT-Clown demonstrated a strong equivalence across different age groups vis-à-vis both FIT and theoretically predicted *M*-capacity. As expected, CMT-Balloon failed to exhibit sustained equivalence across age groups when compared either to FIT or predicted *M*-capacity. In contrast FIT scores, as CMT-Clown scores, were statistically equivalent to the theoretical *M*-capacity of various age-group samples.

Because six FIT scores of 13–14 year olds had been dropped after being classified as outliers for theoretical reasons (i.e., *M*-score = 1, when age-group average is ∼6), we conducted additional equivalence tests for 13–14 year olds that included the theoretically defined outliers for this age group. This was done to appraise whether omitting them as outliers changed results. Under these conditions *M*-scores for CMT-Clown, CMT-Balloon and FIT were, respectively, 6.44, 6.08 and 5.15. Results with CMT-Balloon and FIT remained the same when compared with predicted *M*-capacity. Difference was observed relative to the data of Table 3 for 13–14 year olds only in two respects: (1) comparison of CMT-Clown and FIT showed non-equivalence – expected because six participants were under-performing in FIT; (2) CMT-Clown and CMT-Balloons were found to be equivalent, as we found with adults.

4. Discussion

We introduced the color matching task (CMT) in two versions to study contextual (facilitating versus misleading) cues as they affect assessment of mental attentional (working memory) capacity. Executive demand and action schemes were held constant across items within each CMT version; and they were very similar in both versions. We found that presence of misleading cues (in FIT and in CMT-Clown) improves considerably a task's ability to assess *M*-capacity across development. Indeed, a theoretical "stepwise" pattern of *M*-growth across development, long demonstrated with other *M*-measures (in the FIT – Pascual-Leone & Baillargeon, 1994; or in a compound stimuli visual information task – Johnson et al., 1989; Pascual-Leone, 1970) was recovered in our data with both FIT and CMT-Clown, but not with CMT-Balloon. In both FIT and CMT-Clown, performance expresses a trade-off between the estimated *M*-capacity of participants and the estimated *M*-demand of items. When this trade-off (i.e., *M*-capacity minus *M*-demand) results in a positive difference, participants tend to succeed; when the difference is negative, participants tend to fail – as mathematical models by Pascual-Leone (1970), Pascual-Leone and Baillargeon (1994), and Morra (2000) predict.

4.1. Facilitating versus misleading

This theoretical distinction (facilitating versus misleading situations) is empirically evident in performance scores, whether using *M*-scores or purely empirical raw proportion scores. CMT-Clown was generally harder than CMT-Balloon (Fig. 2). The task relevant colors in CMT-Clown are *embedded* into the figure of a clown – a highly salient irrelevant object with multiple irrelevant parts. These parts elicit various task-interfering schemes that strongly compete for activation. The lack of these irrelevant aspects and other facilitating aspects of CMT-Balloon (e.g., non-salience of balloon shapes) make easier its automatization with practice, at least at the lower-difficulty levels. In contrast, misleading aspects of CMT-Clown make it harder to gain task efficiency during testing, which causes “stepwise” performance as a function of age.

Although consistent with previous research (Pascual-Leone & Baillargeon, 1994; Pascual-Leone & Johnson, 2005; Pascual-Leone et al., 2000; Cowan et al., 2005; Engle, 2001; Engle & Kane, 2004), we have here shown for the first time, over five age group samples and using tasks that differ only in context, that suitably misleading contexts optimize assessment of attentional capacity or working memory. Indeed (see Fig. 1) the proportion of correct responses decreased as a function of *M*-demand for CMT-Balloon, but in all age groups, this mean performance was always significantly different from chance (i.e., .50), with one exception (i.e., 7–8 year olds for *M*-demand 7). Although CMT-Clown performance also decreased with *M*-demand we could observe that 7–8 year olds were performing at chance at *M*-demand levels 5–8; 9–10 year olds performed at chance at *M*-demand levels 7 and 8; and 11–12 year olds at *M*-demand 8. Thus even though number of colors to be processed remains the same (in the Clown and Balloon tasks) the context influences performance, and it does so dramatically in younger children. To investigate working memory in children, researchers should carefully control the task's contextual influences (facilitating or misleading), because they can affect outcome. Furthermore, tasks with a misleading context could provide more reliable means for assessing *M*-capacity or working memory.

In FIT, misleading features are created by overlapping lines in the test figures, which produce subsidiary irrelevant shapes, and constitute an “embedding context”. In these contexts, task relevant features are disguised by misleading perceptual habits or by the sort of gestalt perceptual processes often called field effects, which interfere with task solution (Pascual-Leone & Baillargeon, 1994). The key misleading aspect in CMT-Clown is the salient embedding context constituted by the clown figure as a whole. Participants' *M*-scores in FIT and CMT-Clown had (unlike CMT-Balloon) very similar quantitative values in each age group. Comparison of either raw proportion or *M*-scores in the two CMT versions shows clearly the effect of contextual features. When we compared the correlation of Clown with FIT, with Balloon partialed out (0.41, $p < 0.0001$), versus the correlation of Balloon with FIT, with Clown partialed out (0.21, $p = 0.016$), the Clown task performance appears to be closest to FIT.

Furthermore, when strict equivalence tests were conducted using *M*-scores, we found that CMT-Clown (*but not CMT-Balloon*) exhibited strong equivalence with both FIT and theoretical *M*-capacity values. These results demonstrated great similarity of CMT-Clown and FIT as *M*-capacity measures (despite their many perceptual and procedural differences). This is in contrast with the CMT-Balloon and CMT-Clown, which failed to show equivalence even though they share many procedural and perceptual characteristics. Because our *M*-capacity measures have classes of items designed to exhibit graded levels of mental-attention difficulty, when they are misleading these tasks elicit graded levels of effortful processing in different item classes. Such effortful processing causes graded changes in performance, which express limits of mental processing at different ages (or stages of development).

4.2. "Stepwise" stages of development

When measured by *M*-scores, performance on CMT-Clown, but not CMT-Balloon, showed an across age growth pattern in graded levels predicted by Pascual-Leone's theory (Pascual-Leone, 1970; Pascual-Leone & Johnson, 2005). The graded trade-off between *M*-capacity and *M*-demand was expressed by different age-group performances within and across different levels of items. Such pattern (items passed only when participants' *M*-capacity is equal to or larger than items' *M*-demand) is clear in Table 1. The *M*-capacity model (Pascual-Leone & Johnson, 2005) can be seen as an endogenous, maturational component of working memory, which is interpretable as mental attentional "energy". As our data suggest, this energy can be measured via a mental "space" (i.e., *M*-capacity) generated when attentional effort is applied to maximally activate task relevant schemes. Cowan and Alloway (2009), who discuss our model, fail to see that this attentional resource is energy, although its measure appears as a mental space. Our results, which are generally congruent with other findings (e.g., Cowan et al., 2005; Cowan, Saults, & Morey, 2006; Logie & Pearson, 1997; Riggs, McTaggart, Simpson, & Freeman, 2006) show how to quantify age-related growth in working memory and thus explain transitions from one of Piaget's (or Case's) developmental substages (or stages) to the next, reaching a maximum capacity of about seven (Pascual-Leone, 1970; Pascual-Leone & Johnson, 2005).

This quantification remains controversial, and other researchers only do a qualitative analysis (e.g., Case, 1992), or follow other quantitative methods (e.g., Halford, Cowan, & Andrews, 2007; Halford et al., 1998) less grounded developmentally. Indeed, Halford et al. (2007) posit that 1, 1.5, 5, and 11-year-olds have, respectively, capacity to maintain in mental space 1, 2, 3, and 4 items active. These authors believe that 4 is the maximum adult level of mental capacity. To the extent that developmental stages formulated by Piaget or Case are considered acceptable, models with such narrow range of working memory growth must force researchers to explain some transitions between developmental stages as due solely to learning, whereas other transitions could be due to working memory growth and learning. Researchers who claim that a working memory of 4 is a maximum should investigate closely the possibility of these two sorts of stages, as it would support their view. Different views about working memory capacity growth can, and eventually will, be decided via developmental investigation.

5. Conclusions

Absence of misleading factors lowers *M*-demand of items, and does so *unevenly* for different age groups (due to *M*-demand/*M*-capacity trade-off). This is true particularly for items that can be overlearned with practice (usually facilitating items), even when they are well scaled and graded initially for attentional demand. The presence of misleading factors makes harder intra-task learning, and maintains more stable the task's mental demand, thus improving assessment of working memory across development (via mental attentional capacity).

Our misleading task, the CMT-Clown, exhibited two key characteristics of *M*-tasks: it empirically obeyed theory-predicted mental attentional scores for 7–8, 9–10, 11–12, 13–14 year olds, and adults (Pascual-Leone, 1970; Pascual-Leone & Johnson, 2005); and it strongly correlated with FIT, yielding performance *M*-scores at each age group that are very close to FIT scores. The "stepwise" developmental patterns (*M*-capacity/*M*-demand trade-off) found in this study (Table 1) replicate, within the CMT-Clown task, but not in CMT-Balloon, results obtained and statistically modeled by Pascual-Leone and Baillargeon (1994) within the FIT. This is important, theoretically and practically, because both CMT versions were created top–down (a priori) on the basis of theory-guided task analysis, and consequently both tasks share the same basic executive demand and action schemes. This evidence is strongly in support of our main assumption: *misleading contextual factors in a task or test situation improve considerably a test's ability to measure working memory in adults and across development*.

We should emphasize that the CMT paradigm is very suitable for use in fMRI studies, because it follows the protocol of a 1-back task (Owen et al., 2005). In CMT, parametric scaling of stimuli, fast presentation, the nature of responses (a binary-button press choice), and the presence of "baseline" items (non-response standard-figure items that separate response blocks), all are task characteristics intended for use with functional magnetic resonance imaging (fMRI). We designed the CMT tasks with neuroimaging compatibility in mind, to measure attentional capacity and investigate contextual influences on developmental stage patterns of performance in children. Indeed, we have successfully used CMT-Clown with fMRI in an adult sample and have shown that prefrontal activity increases linearly as a function of items' *M*-demand (Arsalidou, 2008; Arsalidou, Pascual-Leone, Johnson, & Taylor, 2008).

Appendix A.

A.1. Metasubjective task analysis (MTA)

MTA is a rationally based approach used to predict the *M*-demand of a task. It incorporates both relevant situational features and theory-predicted constraints posed by the knowledge the researcher has of the task and the type of participant. The TCO provides a framework in which MTA can be used to both identify an appropriate strategy for the task in question and model the operative and figurative schemes likely involved in the task solving process (Pascual-Leone & Johnson, 1991, 2005). Once validated with empirical data this analysis can be used to quantify the *M*-demand imposed by the specific situation. Identifying the *M*-demand that tasks impose on the individual can improve methodological choices in the design of new tasks, as illustrated in this paper by our success in testing the *a priori*-designed and predicted contrast between the two versions of the CMT.

We used metasubjective task analysis to estimate the *M*-demand of CMT items. The task analysis for the CMT-Clown is summarized in Formula 1, and it will help readers to consult the formula as they read the explanation below.

$$\begin{array}{l}
(1)\ M[\underline{\text{SCAN\&IDEN}}^{L1}\ (\{\#\text{IGN: f, irrC, locC, repC}\}_{L1}\ \underline{\#\text{Ex}} < \underline{\text{cc1}}, \underline{\text{cc2}}, \ldots \underline{\text{cc}n} > \text{tci})] \\
(2)\ M[\underline{\text{MATCH\&PRESS}}\ (<\underline{\text{cc1}}, \underline{\text{cc2}}, \ldots \underline{\text{cc}n} > \underline{\text{tci}})] \\
(3)\ M[\underline{\text{RESET}^{\underline{L3}}/\text{RECUR}}^{L2}\ (\{\#\text{set.c'} \leftarrow\}_{L3}\ \underline{\text{set.t}/\#\text{set.t}}:\{\text{tci'} \leftarrow \text{tci}\}_{L2})]
\end{array} \qquad (1)$$

The task has an overall executive procedure that consists of three successive steps. We symbolize *operative* (procedural) *schemes* with capital letters, and *figurative* (object, property or representation) *schemes* with lower case. *Parameters,* which prescribe conditions to the application of operatives, are identified by #. In the first step (1) a participant must *scan*, one by one, and *identify* (**SCAN&IDEN**) the relevant target colors found in the current Clown item. As the participant scans & identifies he/she has to *ignore* (**#IGN**) the face colors (**f**), irrelevant colors (**IrrC**), location of colors (**locC**), and repeated colors (**repC**). These, and the embedding context which is the Clown itself, are features making the task misleading. We assume that this injunction to ignore is already chunked with the operative **SCAN&IDEN**. We symbolize such chunking by placing these **#IGN** schemes inside curly brackets subscripting the letter ***L1*** to the second brace, and simultaneously placing a superscripted ***L1*** on **SCAN&IDEN**. This signifies that the latter is the operative scheme portion of a chunk that controls the former. The *L-boosting process*, symbolized by ***L1***, ***L2***, and ***L3*** in Formula 1, corresponds to multiple schemes that are so highly associated (structured together) that the chunk requires only one unit of *M*-energy to be hyperactivated.

Step (1) in the model shows a moment when the participant, scans the current set of target colors **(set.t) to actively extract (*#Ex*)** one of them (**tci**), **also keeping in mind** (i.e., *M*-boosting) the total set of criterion colors (**set.c**) from the previous item. The current target colors do not need to be *M*-boosted (voluntary attention) because they are perceptually given. Step 1 shows the schemes kept in mind. Participants must keep in mind *each* of the criterion colors separately, because they are no longer present, and next he/she will have to match them, one by one, with each of the target colors **(tci)**. This total set of criterion colors is symbolized in step (1) as **<cc1, cc2, . . ., cc*n*>**.

In step (2) participants, pursuing the analysis begun in step 1, have to match each and every target color **tci** with the total set of criterion colors (to check whether the color in question is among the criterion colors). This is done by keeping in mind (boosted with *M*-capacity) the operative scheme (**MATCH&PRESS**) of step 2, the target color to be matched (**tci**), *and* the ***n*** criterion-color schemes of step 1. Finally, step (3) plays a dual function, indicated by the incompatibility sign (/) that separates operatives **RESET** and **RECUR**; this logical connective symbolizes that one or the other operative, *but not both*, will be applied at suitable moments of the task process. **RESET** applies whenever a new item is introduced to ensure that the set of target colors (**set.t**) of the just finished item is retained with its function changed (←), now becoming the new set (**set.c'**) of criterion colors (**#set.c'←set.t**). This change is chunked (subscripted ***L3***) with **RESET**. In contrast, **RECUR** applies within the processing of each item. The **RECUR** operative changes *M*-centration within the set of target colors from one target color matched to another target color (**tci'**) not yet matched (**#set.t: tci'←tci**). This change process is chunked (subscripted ***L2***) with **RECUR**.

We have underlined in Formula 1 the schemes that in each step must be held online using *M*-capacity, because they are not salient and are not a subordinate part of an activated chunk. Counting these underlined schemes for each step we find that in step (1) the number of schemes to be held within *M*-capacity is equal to $2+n$, where n is the number of criterion colors in the trial item. Participants must use *M*-capacity to boost the operative **SCAN&IDEN**, each of the colors **cc1** to **cc*n*** of the criterion set, and the operative process ***#Ex*** that extracts a current target-set color (**tci**). Each of these extracted colors will, one by one during the recurring Step (2), be matched against the criterion set of colors. In step (3) this *M*-demand number is equal to 2: One of the two mutually incompatible (/) operatives, and the corresponding scheme parameter (#), which is not chunked (remains outside braces of ***L2*** or of ***L3***, because a set of distinct colors is not perceptually given – only individual colors). Thus, participants should be able to solve a given CMT-Clown item, if they have necessary schemes, when their *M*-capacity) is equal to $2+n$. Because classes of items differ in the number of relevant colors presented, i.e., n varies with the item class from 1 to 6, *M*-demand also varies from 3 (i.e., $2+1$) to 8 (i.e., $2+6$).

Using the same Formula 1 of CMT-Clown we can obtain the task analysis for CMT-Balloon. To do so it is necessary to recall that neither the face to be ignored (i.e., **#IGN: f**), nor the embedding figure of a Clown, exists in CMT-Balloon. This circumstance makes the Balloon task contextually facilitating. As a consequence current target colors (**tci**) do not need to be actively extracted and operative process **#Ex** is not necessary: target colors can easily be read from the current perceptual context. The consequence of this difference in terms of *M*-demand for Step (1) is in reducing it by 1 unit: $1+n$, where n is the number of relevant colors, and *M*-demand for the task over classes of items varies between 2 (i.e., $1+1$) to 7 (i.e., $1+6$).

References

Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology*, *45*, 153–219.

Arsalidou, M. (2008). *Context, complexity, and the developmental assessment of attentional capacity*. Unpublished doctoral dissertation, York University, Toronto, ON, Canada.

Arsalidou, M., Pascual-Leone, J., Johnson, J., & Taylor, M. J. (2008). Neural responses to a visuospatial task with six levels of mental demand. *Canadian Journal of Experimental Psychology*, *62*, 262.

Case, R. (1992). *The mind's staircase: Exploring the central conceptual underpinnings of children's thought and knowledge*. Hillsdale, N.J: Erlbaum Press.

Case, R. (1998). The development of conceptual structures. In W. Damon (Series Ed.), D. Kuhn, & R. S. Siegler (Vol. Eds.), Handbook of child psychology: Vol. 2. *Cognition, perception, and language* (5th ed., pp. 745–800). New York: Wiley.

Cowan, N. (2005). *Working memory capacity*. New York: Psychology Press.

Cowan, N., & Alloway, T. (2009). Development of working memory in childhood. In M. L. Courage, & N. Cowan (Eds.), *The development of memory in infancy and childhood* (pp. 303–342). Hove, UK: Psychology Press.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., et al. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*, 42–100.

Cowan, N., & Morey, C. C. (2007). How can dual-task working memory retention limits be investigated? *Psychological Science*, *18*, 686–688.

Cowan, N., Saults, J. S., & Morey, C. C. (2006). Development of working memory for verbal-spatial associations. *Journal of Memory and Language*, *55*, 274–289.

de Ribaupierre, A., & Bailleux, C. (1994). Developmental change in a spatial task of attentional capacity: An essay towards an integration of two working memory models. *International Journal of Behavioral Development*, *17*, 5–35.

Engle, R. W. (2001). What is working-memory capacity? In H. L. Roediger III, & J. S. Nairne (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 297–314). Washington, DC: American Psychological Association.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 145–199). NY: Elsevier.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

Halford, G. S. (1993). *Children's understanding: The development of mental models*. Hillsdale, NJ: Erlbaum.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Science*, *11*, 236–241.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Brain and Behavioural Sciences*, *21*, 803–864.

Hulchanksi, J. D. (2007). The three cities within Toronto: income polarization among Toronto's neighbourhoods, 1970–2000. *Research Bulletin 41*. December. Toronto: Centre for Urban and Community Studies.

Johnson, J., Fabian, V., & Pascual-Leone, J. (1989). Quantitative hardware-stages that constrain language development. *Human Development*, *32*, 245–271.

Logie, R. H., & Pearson, D. G. (1997). The inner eye and the inner scribe of visuo-spatial working memory: Evidence from developmental fractionation. *European Journal of Cognitive Psychology*, *11*, 241–257.

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. (2000). A new model of verbal short-term memory. *Journal of Experimental Child Psychology*, *75*, 191–227.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*, 46–59.

Pascual-Leone, J. (1969). *Cognitive development and cognitive style: A general psychological integration*. Unpublished doctoral dissertation. Switzerland: University of Geneva.

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica*, *63*, 301–345.

Pascual-Leone, J. (1989). An organismic process model of Witkin's field-dependence-independence. In T. Globerson, & T. Zelniker (Eds.), *Cognitive style and cognitive development* (pp. 36–70). Norwood, NJ: Ablex.

Pascual-Leone, J. Piaget as a pioneer of dialectical constructivism: seeking dynamic processes for human science. In E. Marti & C. Rodriguez (Eds.), *Piaget after Piaget*. Piscataway, NJ: Transactional Publishers, in press.

Pascual-Leone, J., & Baillargeon, R. (1994). Developmental measure of mental attention. *International Journal of Behavioral Development*, *17*, 161–200.

Pascual-Leone, J., & Goodman, D. (1979). Intelligence and experience: A neoPiagetian approach. *Instructional Science*, *8*, 301–367.

Pascual-Leone, J., & Johnson, J. (1991). The psychological unit and its role in task analysis: A reinterpretation of object permanence. In M. Chandler, & M. Chapman (Eds.), *Criteria for competence: Controversies in the conceptualization and assessment of children's abilities* (pp. 153–187). Hillsdale, NJ: Erlbaum.

Pascual-Leone, J., & Johnson, J. (2005). A dialectical constructivist view of developmental intelligence. In O. Wilhelm, & R. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 177–201). Thousand Oaks, CA: Sage.

Pascual-Leone, J. & Johnson, J. A developmental theory of mental attention: Its applications to measurement and task analysis. In P. Barrouillet & V. Gaillard (Eds.), "*Cognitive development and working memory: from neo-Piagetian to cognitive approaches*". New York: Psychology Press, in press.

Pascual-Leone, J., Johnson, J., Baskind, S., Dworsky, S., & Severtson, E. (2000). Culture-fair assessment and the processes of mental attention. In S. Kozulin, & Y. Rand (Eds.), *Experience of mediated learning: An impact of Feuerstein's theory in education and psychology* (pp. 191–214). Elmsford, NY: Pergamon.

Pascual-Leone, J., & Morra, S. (1991). Horizontality of water level: A neo-Piagetian developmental review. In H. W. Reese (Ed.), *Advances in child development and behavior* (pp. 231–276). NY: Academic Press.

Piaget, J. (Ed.). (1974). *Recherches sur la contradiction. 2 Les relations entre affirmations et négations*. Paris: Presses Universitaires de France.

Piaget, J. (1980). *Les formes élémentaires de la dialectique*. Paris: Gallimard, idées.

Riggs, K. J., McTaggart, J., Simpson, A., & Freeman, R. P. J. (2006). Changes in capacity of visual working memory in 5- to 10-year-olds. *Journal of Experimental Child Psychology*, *95*, 18–26.

Siegel, L. S. (1994). Working memory and reading: A life-span perspective. *International Journal of Behavioral Development*, *17*(1), 109–124.

Towse, J. N., & Hitch, G. J. (2008). Variation in working memory due to normal development. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 109–133). New York: Oxford University Press.

Wellek, S. (2003). *Testing statistical hypotheses of equivalence*. Boca Raton, FL: Chapman and Hall/CRC.

Wertheimer, M. (1959). *Productive thinking*. New York: Harper.

Witkin, H. A. (1949). The nature and importance of individual differences in perception. *Journal of Personality*, *18*, 145–170.

Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. R., & Karp, S. A. (1962). *Psychological differentiation*. New York: John Wiley.