anales de psicología, 2018, vol. 34, nº 1 (january), 184-196
http://dx.doi.org/10.6018/analesps.34.1.263431

ISSN print edition: 0212-9728. ISSN web edition (http://revistas.um.es/analesps): 1695-2294

Executive efficacy on Stroop type interference tasks. A validation study of a numerical and manual version (CANUM)

Francisco Gutiérrez-Martínez*, Melchor Ramos-Ortega, and J. Óscar Vila-Chaves

Universidad Nacional de Educación a Distancia (Spain).

Título: Eficacia ejecutiva en tareas de interferencia tipo Stroop. Estudio de validación de una versión numérica y manual (CANUM).
Resumen: En este trabajo presentamos CANUM, una nueva versión numérica y manual de la prueba de interferencia de Stroop. El estímulo utilizado sustituye el conflicto color-palabra de la tarea clásica por el de cantidad-número, dada la interferencia que también se genera entre el valor simbólico del número y la cantidad de veces que éste se repite. Asimismo se sustituye la respuesta vocal por una simple pulsación izquierda-derecha en el teclado del ordenador. El objetivo fue doble: primero, asegurar un índice de control ejecutivo-atencional general desvinculado del factor verbal; y, segundo, ampliar así la población en la que resulta aplicable la prueba, obviando las restricciones relativas a la competencia lectora que conlleva la tarea de Stroop clásica. Los resultados obtenidos en una muestra de escolares revelan una alta fiabilidad en términos de consistencia interna, así como una notable validez predictiva en relación con dos medidas criterio: inteligencia general y amplitud de memoria operativa. Ello avala su utilidad como instrumento de evaluación de la función ejecutivo-atencional, aplicable en un amplio rango de edad tanto con objetivos de investigación como en contextos clínicos y educativos.

Palabras clave: test de Stroop; control atencional; función ejecutiva; amplitud de memoria operativa; factor g de inteligencia.

Abstract: This paper presents CANUM, a new numerical and manual version of the Stroop interference task. The stimulus used replaces the classical color-word conflict with a quantity-number conflict, considering the interference that is also generated between the symbolic value of the number and the amount of times it is repeated. CANUM also replaces the vocal answer with a simple right-left keyboard response. The aim was twofold: firstly, to ensure a general measure of attentional control capacity not linked to the verbal factor; and secondly, to widen the population to whom the test might be applied, avoiding the restrictions on reading ability inherent to the classical Stroop task. The results obtained in a sample of school children reveal a level of high reliability in terms of internal consistency, as well as a significant predictive validity in relation to two criterial measures: general intelligence and working memory capacity. This supports its usefulness as an instrument for the assessment of executive function and controlled attention applicable across a wide age range, both for research purposes as well as for clinical and educational goals.

Key words: Stroop task; attentional control; executive function; working memory capacity; general intelligence.

Introduction

The "Stroop effect" is a well-known "color-word" interference effect in the field of cognitive psychology and across the vast field of research that has developed around the attentional capacity of individuals. Basically, the effect occurs when participants are asked to name the color in which an incongruent word is printed (e.g., "red" is written using green ink or font), ignoring the word itself (Stroop, 1935). More specifically, when the stimulus has these two-dimensions, it has repeatedly proven easier to read the words while ignoring the color in which they are written, than to name the color while ignoring the words themselves. Given this task, errors and response times (RTs) increase significantly in incongruent trials, in which color and word do not match (e.g., the word "green" written in red ink), compared to congruent trials (e.g., the word "green" written in green), or neutral trials (e.g. the word "dice" written in blue)[1].

Although the explanation of this effect is still controversial, it is generally accepted that words provoke an involuntary reading response that interferes with the requested goal of naming the color. Thus, the increases in response error and latency when facing this task might be indicative of the difficulty participants experience in resisting this interference, likewise inhibiting the preponderant reading response. Or, in other words, the phenomenon reflects a participant's capacity for controlling his or her attention when facing a conflict generated between a relatively automatic process that must be inhibited and another the participant tries to execute deliberately (MacLeod & Dunbar, 1988; Posner & Snyder, 1975).

Despite its apparent simplicity, this interference effect has proven very consistent across different task variants, engendering many studies on its nature and associated key factors (see MacLeod, 1991). At the same time, given the high attentional demands and the individual differences observed, a consensus is growing that this type of task affects fundamental aspects of cognition, at least as far as its voluntary control is concerned. Indeed, it constitutes a reference test across different fields focused on the so-called "executive functions" that are related to a person's cognitive flexibility and self-regulation (Garcia-Molina, Tirapu-Ustárroz & Roig-Rovira1, 2007; Lezak, Howieson & Loring, 2004), and whose neurological basis appears to be located in certain areas of the prefrontal cortex (Banich et al., 2000; Miyake et al., 2000). In particular, executive functioning has been linked to *intelligence* processes, in terms of the "g factor" (Friedman et al., 2006), and with the regulatory mechanisms associated with *working memory*, understood as a system of active maintenance under executive-attentional control (Baddeley, 1996; Engle, 2002). Thus, insofar as the Stroop effect is assumed to be an index of this type of control, it has also been incorporated into the investigation of these important constructs, their relationships and their common capacity to predict other measures of achievement, such as academic

*** Correspondence address [Dirección para correspondencia]:**
Francisco Gutiérrez-Martínez. Departamento de Psicología Evolutiva y de la Educación. Facultad de Psicología. U.N.E.D. C/ Juan del Rosal, 10. 28040, Madrid (Spain). E-mail: fgutierrez@psi.uned.es

[1] In the original experiment of John R. Stroop (1935), the response times in naming the color of incongruent words were compared with those words of naming the color of colored squares (experiment 2).

performance (e.g., Bull & Scerif, 2001; Imbrosciano & Berlach, 2005).

Indeed, multiple studies have shown that working memory capacity (WMC) is largely predictive of general intelligence, or Spearman's "g factor" (e.g., Ackerman, Beier, & Boyle, 2005; Colom, Abad, Rebollo, & Shih, 2005; Conway, Kane & Engle, 2003), which many have attributed to the common demand for executive-attentional control of the tasks with which these constructs are measured (e.g., Engle & Kane, 2004; Kane et al., 2007). In this same sense, there is also clear evidence of the relationship between WMC and executive-attentional control (Engle, 2002; Kane Conway, Hambrick & Engle, 2007). This relationship has been particularly evident in the significant correlations found with the Stroop task (Kane & Engle, 2003). In other words, the individuals with the highest performance on WMC tests are the ones less susceptible to interference when performing the Stroop task (Hutchison, 2011; Long & Prat, 2002; Shipstead & Broadway, 2013; Unsworth & Spillers, 2010).

However, the breadth and basis of these relationships remains a matter of debate, given that it is still unclear how the executive and memory components present in the tasks interact, or what their relative contribution to explaining the shared variance might be (Chuderski, 2014; see a review in Stelzer, Andés, Canet-Juric & Introzzi, 2016). In fact, interference and its control in the Stroop task may not only be related to the mechanism of resolving the attentional conflict, but also to contextual factors and other aspects of general processing that is required. Specifically, the relationship between working memory and executive control inherent to the Stroop task also seems related to the necessity of goal maintenance (i.e., naming the color, ignoring the word) throughout the successive trials,

while still inhibiting one's prepotent tendency of reading the word itself (Kane & Engle, 2003). Hence, this also depends on how maintaining this goal is either impeded or facilitated throughout the task. For instance, research shows that goal maintenance is more difficult if the task includes a large number of congruent trials. Namely, facing a higher ratio of congruent trials tends to suspend the executive control system and thus favour goal neglect (Hutchison, 2007, Morey et al., 2012). But also the particular sequences of congruity-incongruity may or may not contribute to this as a function of the dynamic adjustments that they impose in the activation (or dis-activation) of attentional control. For example, additional priming effects —positive or negative— may occur between successive trials (Egner, 2007; Long & Prat, 2002; Meier & Kane, 2013).

In short, the Stroop task involves a complex process of *selective attention* that involves an inhibitory function as well as sustained attention (focusing) linked to the active *maintenance* of the task goal. These broad demands explain the value of the task as an index of attentional control. But they also justify the idea that this type of measure may also reflect the domain-general executive component underlying the relationships between constructs relating to executive functions, working memory, and general intelligence, as well as the common neurological substrate in which they appear to be supported (Conway, Kane & Engle, 2003; Kane & Engle 2002).

In this respect, however, it is worth mentioning that in the classical version of the task, the verbal modality of the stimulus (S) and the vocal nature of the required response (R) may imply a specifically "verbal" bias in processing the apparent conflict. In this sense, provided the hypothesis that conflict occurs in the input phase, when S is perceived and codified in two dimensions (e.g., word and color), it has been suggested that interference may depend on the greater speed, or even the automaticity, of processing the word relative to color (SS compatibility). This would, therefore, make it necessary to control the influence of reading practice and the degree of automaticity associated with the phonological code. In fact, multiple studies in the early grades show that the level of interference is closely related to the increasing development of reading skills up until its acquisition is complete (MacLeod, 1991; Protopapas, Archonti & Skaloumbakas, 2007).

In this context, researchers have introduced task variants that use numerical stimuli instead of words, under the assumption that an understanding of numbers is acquired earlier and independently, linked to the formation of counting small quantities (Bryant, 1996; Gelman & Meck, 1983). Specifically, these variants have replaced the word-color conflict with number-magnitude conflict, taking into account the interference that also seems to be generated between its symbolic value (the value expressed by the cardinality of numbers) and the result of processing them independently along some other dimension or empirical attribute. For example, the relative size of the number itself (to indicate that a '**3**' is 'larger' than a '**5**'), or their quantity when occurring as a set (to indicate that there are 'three numbers' in '5 5 5'). In these studies (e.g., Algom, Dekel & Pansky, 1996; Wolach, McHale & Tarlea, 2004), the difficulty involved in ignoring the numerical information in the digit has been verified such that, just like in color-word interference, the tendency towards identifying the digit's symbolic numerical value might interfere with the actual task in incongruent cases, that is, in cases where there is no correspondence between the irrelevant symbolic meaning and the relevant empirical dimension.

On the other hand, another possible basis for explaining the interference on the classical Stroop task is the coincidence involved in the verbal nature of the stimulus —the word— with that of modality of the response, usually vocal. In this case, it is assumed that the interference will occur later during the output phase, when selecting a response. As the two S dimensions compete for the same vocal channel in the response, the "word" as such may impose itself against "color" only due to the mutual affinity in modality between S and R (SR compatibility). This has led to research on the Stroop effect that contrasts the classic vocal response (naming color) with that of a manual response (selecting an item or pressing a key). The results of research in this respect are not so homogeneous (see MacLeod, 1991), but, in general, the interference effect has been confirmed in the manual versions.

In practice, the advantage of these numerical variants is to isolate the Stroop effect from purely verbal factors and that of reading ability, both linked to the "word" itself as *input* and that possibly condition or limit its application. For instance, this might occur in certain populations with lower levels of education, the illiterate, or those with verbal disabilities (Sedó, 2004). But these variants also have theoretical implications. The fact that the interference effect still occurs when the modality of S or R is changed strengthens the idea that these kinds of tasks reflect a central and general capacity for executive-attentional control and that, in fact, measures devoid of the verbal factor specific to this kind of interference may be more valid.

In this respect, the present study explores a version of the task that employs numerical content and requires a manual response (CANUM) in order to strip it —at least in part— of the possible verbal bias inherent in the classical version, associated with both the presented S as well as the required R. The objective of CANUM, therefore, has been to ensure that this measure of executive-attentional control capacity is more central and non-specific than the classic Stroop involves, due possibly to its verbal nature. Additionally, at the same time we seek to extend the possibilities of the test's application by avoiding the restrictions that the classical approach presupposes, at least with regards to competence in reading.

However, in order to test the consistency of CANUM on a comparative basis in terms of convergent and construct validity, we also apply a parallel version of the classic Stroop color-word task (STROOPm), designed and implemented according to the same procedure. We describe both tests below.

CANUM: A "quantity-number" interference test with manual response

In CANUM, numerical stimuli (digits 1, 2, 3, or 4) are presented in the center of the computer screen and are arranged in a repeated way (e.g., 111) to simultaneously trigger or allow for two kinds of judgement: the digit itself as numerical symbol (number "one" in the example above), and the number of times the digit repeats itself ("three" in the example above). The participant is asked to provide a response to this latter question of magnitude, that is, the "number of times" that the base number is repeated (relevant dimension – hereinafter, Quantity or Q^a) while ignoring the numerical symbol as such (irrelevant dimension – hereafter, base Number or N^o)[2]. Now, in a way similar to the Stroop color-word test, it allows the participant to either face congruent (e.g., 22- two "twos") or incongruent (e.g., 222 - three "twos") cases and, therefore, generates similar effects of facilitation-interference. That is, we assume that when numbers and quantity match (congruent cases: $Q^a = N^o$), the task will be facilitated; whereas when they do not coincide (incongruent cases: $Q^a \neq N^o$), the act of identifying the number itself will interfere with the objective task of indicating how many times it is repeated.

This particular "numeric Stroop effect" has indeed been found in a lot of work, some already classic (e.g., Sedó, 2004; Shor, 1971; Windes, 1968). But in CANUM, in order to manually implement the response to the task (R), another digit (also ranging from one to four) is added both at the beginning and at the end of the stimulus (S), one of which corresponds to the value of the relevant dimension. In other words, the correct option of R appears on one side of the line reflecting the Q^a, and on the other side an incorrect option, which may or may not repeat the N^o. For example, in "4332", the central digits '33' make up the S, the initial '4' to the left represents an incorrect R, and the '2' to the right ends the series and corresponds to a correct R ("two threes"). Accordingly, the subject is instructed to spatially associate these options on each side of the S with two correlative keys to the left and right of the keyboard. In other words, they must respond by pressing the key that spatially corresponds to the correct response (in the example, the correct response would be to press the right key).

No visual emphasis is offered that distinguishes the two parts (central and lateral) of the stimulus configuration[3], thus making the task more complex and more demanding of selective attentional control. Note that in addition to numerical dimensions in conflict (number and quantity), this fact implies maintaining a spatial division between the part corresponding to the S (the central location) and the R options (lateral ends) provided. This is important because the inclusion of response options on either side alters —and possibly amplifies— the levels of incongruity attributable to each stimulus configuration, depending on the alternatives of R used (Q^a) and their possible correspondence with the irrelevant dimension (N^o) in the correct or incorrect option. In this sense, the previous example of "4332" actually corresponds to a neutral condition, since the lateral choices present in Q^a (4**2) do not include repetition of N^o (3), neither in the correct or incorrect sense. Yet, for example, in the case of "4222", the congruent condition (two "twos") is reinforced, since N^o (2) is repeated in the correct Q^a option (2), possibly increasing the facilitating effect. In contrast, in the case of "4442" the incongruous condition (two "fours") is reinforced, since N^o (4) is repeated in the incorrect Q^a (4) option, possibly contributing to an increase in interference.

In this regard, we have also taken into account the possible additional occurrence of other interference effects, in a way similar to that of Stroop, but directly linked to the lateral

[2] The name of the test comes from the combination of the first part of each term that designate in Spanish both dimensions, relevant and irrelevant: CANtidad vs. NUMero = CANUM.

[3] With this label we refer, obviously, to the set of digits presented that includes the actual "stimulus" (central part) and the offered response options (lateral ends).

repetitions of base number (Nº) as R options. Thus, on the one hand, the task can be compared to the "Simon effect" (see Simon, 1990; Lu & Proctor, 1995; Hommel, 2011), given that it also poses a visuo-spatial conflict: participants must choose an alternative right-left response by pressing a corresponding key that may or may not be congruent with respect to the position that indicates the relevant dimension. In fact, whenever either of the lateral Qª options corresponds with the Nº itself, a certain visuo-spatial asymmetry is clearly generated in the configurations, capable of inducing the type of response "lateralization" found in the Simon effect. However, since the coincidence can occur both in the correct or incorrect option, its possible incidence will always converge with the corresponding Stroop effect, whether it is interfering (incongruent cases, such as "two ones": "1112" or "2111") or facilitating (congruent cases, such as "two twos": "2221" or "1222") a correct response.

On the other hand, some configurations may also give rise to a certain "flanker effect" similar to the task proposed by Eriksen (Eriksen & Eriksen, 1974; Eriksen 1997). The central digits —relevant to the computation of the Qª— are "flanked" by other digits that will be distracting if they lead to the wrong answer (i.e., that of Nº itself). In particular, we assume that the "flanker effect" will also occur in cases in which base Nº matches one of the R options provided as Qª —whether correct or incorrect—, considering them as "flanks". But in this case, this will be true only insofar as it makes discrimination more difficult between the central digits that constitute S and those lateral digits offered as R options. Thus, in some congruent cases, the facilitating effect may be counteracted by a "negative flanker effect." For example, in the case of "3222" (two "twos"), the occurrence of Nº in the correct lateral flank option '2', combined with the incorrect lateral flank option '3', causes confusion between the S-R parts on the correct flank, as it can be identified as "three twos". On the contrary, in some incongruent cases, the interference effect may in fact be intensified by the "flanker effect". For example, the case "3332" (two "threes") also combines the flanks '3' - '2'. But here the repetition of Nº in the incorrect flank option '3' generates confusion on that flank because it can lead one to identify the case as "three threes". In any of these cases, therefore, a disruptive "flanker effect" may be generated via a superficially correct but invalid description, because one of the options "flanking" as R are mistakenly included as part of the S. Thus, this "flanker effect" is less frequent, but always of an interfering nature and, for this reason, it will serve to increase the difficulty in any of the affected cases (see Appendix I).

STROOPm: A "color-word" interference test with manual response

As we have argued, to test the validity of CANUM on a comparative basis, we designed a parallel version of the classic color-word task of Stroop, to be implemented according to the same procedure: presentation of the S word (color name written in different colors), flanked by two lateral R options (color names written in white) associated in visuo-spatial correspondence to right-left keystrokes on the keyboard. Thus, given that the correct option always names the relevant dimension (the ink color of the S word), this linear configuration allows for a manual response according to the procedure followed in CANUM and under similar conditions: when a condition of congruence occurs between the relevant-irrelevant dimensions, the correct option reproduces the stimulus word (e.g., "green green red", or 'green' next to the word "green" written in green); while given an incongruent condition, it is the incorrect option that reproduces the stimulus word (e.g., "green red red", or 'red' next to the word "red" written in green).

This type of configuration may involve, as in CANUM, a strengthening of the Stroop effect (facilitator or interferer) through the possible addition of the Simon effect, due to the perceptual asymmetry produced by the co-occurrence of words and the lateralization (right-left) of the response (R), something that may give rise in cases of visuo-spatial correspondence. However, by their nature, these same configurations will not bring about cases having possible "flanker effects" in the sense observed in CANUM. Therefore, for the purposes of analyzing response latency and difficulty, we have only considered the three basic conditions: congruence (the S word occurs in one of the R options, designating the color in which it is written), incongruence (the S word occurs in one of the R options, designating a color different from the one in which it is written) and neutral (the S word occurs in none of the R options) (see Appendix II).

Approach and hypothesis

As we have seen, CANUM is a test that requires high attentional control over the interference —between stimuli and between the stimuli and responses— similar to the classic Stroop task (1935), but in a numerical and manual way. On the one hand, the numerical modality of the stimulus aims to eliminate or minimize the specific impact of the verbal factor (relevant to the S-S compatibility at the input phase). On the other hand, the new version calls for a manual response that situates the conflict in the selection of the response outside of any direct overlap between the verbal stimulus and the vocal response (which concerns S-R compatibility at the output phase). Additionally, the way in which we have implemented this response incorporates to some extent the "Simon" and "flanker" effects, such that, overall, we can consider the test as contributing to the development of a valid measure of attentional control as a reflection or manifestation of general executive-attentional competence. In order to achieve this goal, we have designed and applied both tests (CANUM and STROOPm) according to the same procedure. In addition, two working memory capacity tests and a general intelligence test (g factor) were also applied in order to assess and compare their predictive power relative to these important criteria, the relationship of which has been widely supported in the literature.

Thus, in terms of *construct* and *convergent validity* between STROOPm and CANUM, we make the following predictions:

1. Both tests will provide measures of difficulty (in terms of response accuracy and speed), according to the level of interference attributable to the different cases as a function of the congruence-incongruence conditions theoretically associated with each. In particular, and according to the approach developed in the current study, we predict:
   - A. For both tests, the best performance will occur in conditions of congruence, the worst in conditions of incongruence, and an intermediate level in neutral conditions.
   - B. More particularly, in CANUM we predict worst performance in cases with a possible "flanker" effect, whether congruent or incongruent, at least compared to their unaffected counterparts.
   - C. Overall, because of its higher demands on executive-attentional control, CANUM will be more difficult than STROOPm.

3. However, as measures of the same underlying construct, we predict a significant relationship between CANUM and STROOPm, at least in reference to the overall score provided by each.

Finally, in terms of *criterion validity:*

4. Both tests will show good predictive potential in regards to working memory capacity (WMC) and general intelligence (g factor), although we expect higher CANUM correlations than we do for STROOPm.

Method

Participants

The sample was quasi-randomly taken across primary and secondary classrooms belonging to three public schools in Conil (Cadiz, Spain), all belonging to similar socioeconomic strata. Given the goals of the research and the requirements of the various tests, we believed it appropriate to sample from pre-adolescent participants to ensure minimum competencies. Also, sampling from a wide age-range would similarly allow us to test the consistency of the proposed new instrument (CANUM) as a function of the age variable. Thus, the sample included 128 students aged between 10 and 15 years (M = 12.73, SD =2.49) after excluding those whose performance on any given test fell 2.5 standard deviations below the mean, as well as those who for any reason were unable to perform all of the tasks.

Testing Instruments

CANUM

To bring about the relevant facilitation and interference effects, CANUM includes 12 congruent configurations (3 x 4) and 12 incongruent configurations (3 x 4), all of which were generated on the basis of the following range of numbers: 1, 2, 3 and 4. In each configuration, however, 9 items are seen as reinforcing the "Simon effect" and 3 as provoking a disruptive "flanker" effect.

On the other hand, to verify and compare the expected effects according to interference conditions, CANUM included two kinds of "neutral" (N) configurations, in which the lateral response options offered are different from (do not repeat) the base number. Firstly, 12 *Na* cases were selected from the 24 generated by the range (between 1 and 4) provided (e.g., "3442"). Secondly, 12 *Nb* parallel cases were added using a base number outside (between 5 and 8) of that range (e.g., "3772").

In addition, across all conditions the cases included were duplicated to counterbalance the left-right position of the lateral response (R) options offered. The final set of 96 items was presented to participants in a pre-established sequence, while still obtained at random (see details in Appendix I).

STROOPm

In a parallel way, the STROOPm test includes 12 congruent configurations (3 x 4) and 12 incongruent configurations (3 x 4), all of which were generated on the basis of the following range of colors: red, yellow, green, and blue. Similarly, as a means of contrasting the expected effects of interference and facilitation, we included two neutral configuration types that paralleled those used in CANUM: a "semi-neutral" type, that refers to cases in which the stimulus S does not occur as a lateral response R option (e.g., "green blue red"; where the central word "blue" written in green represents S) and a "neutral" type, that refers to cases in which the stimulus S does not name a color at all (e.g., "green dinner red"; where the central word "dinner" written in green represents S).

Likewise, in order to counterbalance the left-right positions of the lateral response options, the total set of configurations were doubled. As in CANUM, the resulting set includes 120 items, which were presented to participants in a pre-established, yet random, order (see details in Appendix II).

Working memory tests: RST and RxST

Two measures adapted for children were employed as working memory capacity tests, both of which follow the double-task structure of the classic reading span test (RST) of Daneman and Carpenter (1980), based on reading unconnected sentences. In particular, we applied an adaption of this same test, carried out by Carriedo and Rucián (2009), as well as a parallel test of "reasoning span" (RxST), using the resolution of simple analogies instead of basic reading. Gutiérrez-Martínez y Ramos (2014) adapted the latter test for the use in children, and it is that version which is used in the current study.

The structure of the adapted RST and RxST used here is equivalent to that of the original test (for detailed description see Elosúa, Gutiérrez-Martínez, García-Madruga, Luque & Gárate, 1996; and Gutiérrez-Martínez, García-Madruga, Carriedo, Vila & Luzón, 2005). For this reason, its application is similar. The items (sentences or analogies) are presented in a successive number of increasing series (from two to five), and include three essays at each level, which in total thus make up four blocks or levels of increasing difficulty.

Test of general intelligence

On the other hand, Raven's Progressive Matrices (RPM) test (see Raven, Court & Raven, 1996) was used as a measure of general intelligence. It is recognised as providing a scale for estimating "g factor" or fluid intelligence. The task requires participants to reason about the relationships that make up an incomplete set of abstract forms (in a 3 x 3 matrix) in order to select from amongst several options the shape that correctly complete the set. The test includes 60 matrices.

Procedure

The participants performed all of the tests in counterbalanced order. The tests were presented using computers and applied individually, except in the case of RAVEN, the application of which was collective with experimenter support.

The two WMC tests were administered on a computer using the *E-prime* software (Schneider, Eschman & Zuccolotto, 2002), following a repetitive sequential process: in each test the participant had to complete a series of items of increasing difficulty, corresponding to the level tested (2, 3, 4 and 5). In RST, processing consisted of reading sentences and in RxST of reading and solving analogies. At the end of each series a question mark (i.e., "?") appears on the screen and then the participant tries to remember, in the order of appearance, the key words: in RST, the last words of each sentence; and in RxST, the words chosen to complete the analogies. The "integrated criterion" developed by Elosúa et al. (1996) was used to evaluate a participant's performance. This criterion assigns an integer score that corresponds to the level achieved (between 2 and 5), plus a decimal score (between 0.1 and 0.9) that qualifies the actual performance within that level by considering the three trials.

In a similar way, both STROOPm and CANUM were administered via the open-source software known as PEBL (http://pebl.sourceforge.net/), which allows recording both reaction times as well as error hit rates (Mueller & Piper, 2014). As previously described, the stimuli (words or numbers) were presented in the center of the screen, and the response options on either lateral side. In this way, the participant can choose the correct option by pressing the corresponding key: the "P" key as the right side option, and the "Q" key as the left side option. Pressing the space bar enabled a participant to move onto the next item. Through this procedure the participant was able to proceed at his or her own speed, and in so doing the response times (RT) of each test could be recorded. For this reason, the participant was provided instruction through an initial set of practice items (24 random cases selected from "neutral" conditions), while being requested to "try to go as quickly as possible without making mistakes". Feedback was provided with the word "incorrect" appearing on the screen after each failed attempt, until refreshing it by pressing the space bar to begin the next test item.

Measures

Given its nature, the results obtained from tests like that of Stroop have usually been operationalized as measures of time rather than measures of accuracy. In particular, "interference" or "facilitation" scores, calculated as the difference between the reaction or response time (RT) used in incongruent or congruent cases —respectively— with the time required by the neutral cases. This make sense given what is requested of the participant (speed, but without mistakes). Thus, one expects the two parameters —speed and accuracy— to trade-off, insofar as an inverse relationship holds for each test trial: the higher the speed of executing the item, the lower probability of being accurate on it.

Nevertheless, this is not necessarily the case for the whole test. This is true especially when taking into account that, in addition to the general stimulus condition presented in each trial (incongruent, congruent, or neutral), other variables —such as the proportion of congruent-incongruent items in the list, or the particular sequence of presentation— may also affect the level of difficulty from one trial to the next, and/or result in inconsistencies between the measures themselves (speed or accuracy). In fact, according to the double mechanism of attention and memory postulated by Kane and Engle (2003), the difficulty in resolving attentional conflicts is what is reflected in increasing RT measures (via response latency), while forgetting the goal of the task (i.e., goal neglect) is that which results in errors (via response inaccuracy). Thus, fluctuations or local variations in these measures are to be expected. In other words, according to the particular sequence of congruity-incongruity, either delays in processing speed or an increase in error rates will tend to occur, and therefore this speed-accuracy tradeoff may be variable throughout the course of successive trials of the task.

In the present study we did not directly control the kinds of factors alluded to above (ratio of congruity-incongruity per list, and presentation sequence). We simply applied a pre-established sequence of the different conditions and cases, albeit obtained initially at random. Consequently, it makes sense to take both sides of the tradeoff (speed and accuracy) into account when evaluating global individual differences, that is, those related to performance as a whole on the test. And hence, in the current study, besides speed and accuracy, we consider separately a combined measure of efficacy (E) (dividing response accuracy by response speed, within each condition and in the total; i.e., E = accuracy/speed) as yet another operational measure of executive-attentional difficulty of the task, and thus, of the overall efficacy in executive control shown by the participant.

Indeed, we expect this measure of efficacy to be more consistent with that of our hypothesis. Therefore, and in order to compare them, we will record the results of reliability and validity obtained across all three measures (speed, accuracy, and efficacy), both in STROOPm and in CANUM, and in reference to their different conditions.

Data analysis

The reliability of the tests has been estimated in terms of "internal consistency" by means of Cronbach's alpha coefficient. For this, the various conditions of the tasks (congruent, incongruent and neutral) were taken as underlying factors contributing to the same construct.

The validity of the construct is assessed, firstly, by analysing the expected differences in difficulty between the various conditions within each test and between the two tests, in both cases in reference to the total scores obtained. To do so, given the heterogeneity of the sample in terms of school level and age, we first evaluate the normality of distributions by means of the Kolmogorov-Smirnov test. In this sense, we verify that in many of the variables we cannot assume normality. For this reason, in order to test for significance in performance differences, we applied the non-parametric Wilcoxon test for related samples.

For the same reason, the expected linear relationship between the two tests was estimated by calculating the Spearman correlation coefficient between the total scores. This correlation analysis was also performed with respect to the criterion measures (WMC and general intelligence) in order to verify the predictive capacity of the two tests studied.

Results and discussion

Reliability and condition difficulty

Table 1 shows the mean percentage of hits and the mean response times (RT) per item, recorded across each test and condition[4], as well as the combined measure for efficacy. In addition, the reliability coefficient (Cronbach's alpha) associated with each of these measures is presented. In this respect, as can be seen, the RT measure and the efficacy measure are shown to be reliable, with excellent coefficients (> .90) for both STROOPm and CANUM. However, in STROOPm only the efficacy measure provides evidence consistent with expectations about various conditions relative to its presumed nature across different conditions (Hypothesis 1a): the congruent condition shows the most effective performance (*Q* = 83.6) by facilitating performance. After that, the neutral (*N* = 70.7) and semi-neutral (*S* = 67.2) conditions show no effects in either direction and so, as expected, can be taken as baseline references. And, finally, the incongruent condition (*I* = 62.5) has the lowest score, presumably due to the inherent interference effect.

Table 1. Means (*SD*) in measures of difficulty according to interference condition in the STROOPm and *CANUM* tests (*N* = 128).

		Hits (%)		**RT per item** (ms)		**Efficacy** (Hits/RT)	
		STROOPm	*CANUM*	STROOPm	*CANUM*	STROOPm	*CANUM*
Neutral	**(N[1]-*Nb*[2])**	96.91	*93.00*	1480	*1780*	70.74	*58.89*
		(4.44)	*(7.6)*	(443)	*(652)*	(19.28)	*(20.3)*
Semi-neutral	**(S[1]-*Na*[2])**	96.24	*94.27*	1564	*1902*	67.16	*55.95*
		(3.24)	*(5.9)*	(527)	*(714)*	(18.47)	*(18.8)*
Congruent	**(C[1]-*Cs*[2])**	97.53	*92.01*	1259	*2202*	83.64	*49.39*
		(3.56)	*(8.8)*	(379)	*(940)*	(22.07)	*(20.3)*
	(- *Cf*[2])		*79.69*		*2099*		*42.67*
			(19.4)		*(785)*		*(17.7)*
Incongruent	**(I[1] - *Is*[2])**	83.53	*90.58*	1448	*2206*	62.51	*47.10*
		(12.17)	*(8.7)*	(450)	*(893)*	(19.42)	*(17.0)*
	(- *If*[2])		*79.56*		*2165*		*42.54*
			(18.6)		*(930)*		*(17.5)*
Total		*94.09*	*91.01*	*1463*	*2014*	*69.65*	*51.11*
		(3.86)	*(6.37)*	*(446)*	*(743)*	*(18.70)*	*(17.6)*
Cronbach's Alpha		0.62	*0.78*	0.96	*0.96*	0.96	*0.95*

([1]) relative to STROOPm – N: Neutral; S: Semi-neutral; C: Congruent; I: Incongruent

([2]) relative to *CANUM – Nb: Neutral b; Na: Neutral a; Cs: Congruent+Simon; Cf:Congruent+flankers; Is: Incongruent+Simon; If: Incongruent+flankers;*

In fact, it is the measure of efficacy (E) that reveals highly significant differences (p < .001) across the various conditions in all cases and in the expected direction, without any mismatches (Table 2). This does not occur with the accuracy measure, which fails to differentiate between neutral and congruent conditions; nor with the RT measure, which yields clearly inconsistent data: RT measures turn out being greater in the neutral conditions (N and Q) than in the incongruent (I) condition, and hence the differences in these cases are unexpectedly positive. Taking into account, moreover, that these differences concern precisely the traditional indices of "interference" and "facilitation", it is evident this lone temporal measure is weak as an operative index of the executive-attentional capabilities that supposedly underlie the performance on the test.

[4] In the tests not all the analyzed conditions contain the same number of items, thus in order to make them comparable, instead of taking direct measurements of hits and response latencies, percentage accuracy and average RTs per item have been used.

Table 2. Mean differences in the STROOPm and *CANUM* conditions (row *minus* column) across the various measures of difficulty (*N* = 128).

	Hits (%)					**RT per item** (ms)					**Efficacy** (Hits/RT)				
STROOPm	**-N**	**-S**	**-I**			**-N**(2)	**-S**(2)	**-I** (1)			**-N**	**-S**	**-I**		
C-	0.62	1.28**	14.00***			_**-221**_***	_**-304**_***	-188***			12.90***	16.47***	21.13***		
N-		0.67*	13.38***				-83***	_**32**_				3.58***	8.23***		
S-			12.71***					_**116**_***					4.65***		
CANUM	***-Na***	***-Cs***	***-Is***	***-Cf***	***-If***	***-Na***	***-Cs***(2)	***-Is***(1)	***-Cf***	***-If***	***-Na***	***-Cs***	***-Is***	***-Cf***	***-If***
Nb-	-1.27*	0.99	2.42**	13.31***	13.44***	-122***	_**-422**_***	_**-426**_***	-319***	-385***	2.94**	9.50***	11.79***	16.22***	16.35***
Na-	-	2.26**	3.69***	14.58***	14.71***	-	_**-300**_***	_**-304**_***	-196***	-263***	-	6.56***	8.85***	13.28***	13.41***
Cs-		-	1.43	12.33***	12.46***		-	_**-3**_	103*	36		-	2.28*	6.72***	6.85***
Is-			-	10.89***	11.02***			-	107*	40			-	4.43***	4.56***
Cf-				-	0.13				-	-66				-	0.13
If-					-					-					-

* *p* < .05; ** *p* <.01; *** *p* <.001.

(1): In bold and italic, the differences in the *RT* of the *incongruent conditions* (I), which can be taken as direct indices of the associated interference, especially the difference in the neutral condition (N-I).

(2): In bold and italic, the differences in the *RT* of the *congruent conditions* (C), which can be taken as direct indices of the associated facilitation, especially the difference in the neutral condition (C-N).

With regards to the new CANUM test, the difficulty measures in Table 1 for each general condition have been broken down according to expectations for each added effect: Simon only (s) or also with flankers (f). In fact, the overall results in conditions of congruence (C) and incongruity (I) are misleading, given that the performance in flanker conditions (Cf and If) was different from the others (Cs and Is) across all measures. This is due to the fact that both conditions (Cf and If) in these cases tend to be equal, producing a similar reduction in RTs relative to their counterparts (Cs and Is), but likewise impairing response accuracy. Taken together, this is thus reflected as a significant drop in the combined efficacy measure. In short, the cases Cf and If are resolved faster but with much less success (i.e., fewer hits). However, these flanker cases do not diminish the internal consistency of the test. Rather they appear to behave together as a new condition (f) generating more interference than incongruent cases exhibiting only Simon effects (Is), a result that lines up with Hypothesis 1b.

On the other hand, we can observe that the differences between the various measures in CANUM are not as marked as in STROOPm, and all reflect the same general pattern: performance in the congruent condition (Cs) is somewhat better compared to the incongruent (Is). But, contrary to expectations, it is not really facilitative compared to the neutral conditions (Na and Nb), where the highest scores are reached. In other words, in CANUM —unlike that of STROOPm— the congruent condition with only the Simon effect (Cs) appears to have functioned more like another level of interference: less effectively than the proper interference condition (Is), but clearly more so than either neutral condition (Na and Nb). Leaving aside the flanker cases, this implies the following sequence of difficulty, from least to most difficult: neutral, congruent, and incongruent. This does not exactly correspond to our general hypothesis (Hypothesis 1a).

However, and in reference to this pattern, the combined measure of efficacy is again the one that appears to reveal it in a more consistent way. Thus, as seen in Table 2, this measure reflects significant differences in the expected direction, except for in the conditions with "flanker" effects (in which no difference occurs between Cf and If). But this is theoretically acceptable as a more precise delimitation of the conditions in terms of levels of interference and difficulty. That is, 'f' would be the condition of greatest interference, including in an undifferentiated way the two types of cases, Cf and If. In fact, such indifferentiation may also explain why the ordinary congruent condition (Cs) was not shown to be "facilitative" compared to the neutral conditions. To the extent that the "flanker" effect distorts Stroop conditions of congruity-incongruity, the attentional control required may imply greater "alertness" or "vigilance," even in ordinary congruent cases (Cs). This would translate into an increase in RTs (indeed, it is similar to that which occurs in incongruent "Is" cases — see Table 1), thus nullifying some of its facilitating effect, and thus making them more difficult than the neutral cases. Ultimately, even cases without any direct "flanker" confusion (e.g., in case "2224", with no corresponding R flank) also do not conform to the congruent verbal description "two twos" since this configuration perceptively appears as "three twos", and thereby also calls for an additional discrimination incompatible with the facilitation effect. In sum, the condition Cs implies some level of interference, and therefore contributes to increasing the difficulty of the CANUM test as a whole, compared to that of STROOPm.

Construct and criterion validity

Indeed, as predicted (Hypothesis 1c), CANUM was more difficult than STROOPm across all measures according to global scores (Table 1)[5]. Thus, the percentage of hits is significantly lower in CANUM (-3.08, Wilcoxon, z = -4.98, *n* = 128; *p* <.001), and mean RTs significantly higher (+551 ms,

[5] This greater difficulty was also experienced across the various conditions (congruent, incongruent, and neutral). But because in theory we cannot assume a complete parallelism between these —especially in cases with possible "flanker" effects— here we choose not to look further into this contrast between particular conditions.

Wilcoxon, z = -9.40, *n* = 128; *p* < .001). Consequently, the combined measure of efficacy is also significantly lower in CANUM (-19.14, Wilcoxon, z = -9.56, *n* = 128; *p* < .001). All this is in line with our prediction that CANUM places a greater set of demands on interference control, both in processing the input stimulus (numerical Stroop effect) and in response management (additional "Simon" and “flanker" effects).

However, with reference to these same overall scores, the expected relationship between STROOPm and CANUM as measures of the same construct is also confirmed (Hypothesis 2). As can be seen in Table 3, the correlations between both tests are equally high and significant in the RTs and efficacy measure (*r* = .79 in both; *p* < .01), representing a shared variance of 62 %. Also, although to a lesser degree, the correlation in accuracy, or hits (*r* = .54; r^2 = .29; *p* < .01) is also significant. All of this therefore serves to confirm its convergent validity in the expected direction. That is, the two constructs, CANUM and STROOPm, seem to reflect the same kind of general competence, presumably linked to its broad and common executive-attentional demands related to managing the interference.

Table 3. Spearman correlations between the various measures of difficulty in the STROOPm and CANUM tests (*N* = 128).

		STROOPm			*CANUM*		
		% Hits	RT	E	*% Hits*	*RT*	*E*
STROOPm	% Hits						
	RT	.00					
	Efficacy (E)	.14	-.98**				
CANUM	% Hits	.54**	.12	-.04			
	RT	.05	.79**	-.77**	-.09		
	Efficacy (E)	.06	-.79**	.79**	.09	-.97**	

** *p* < .01.

On the other hand, the correlation matrix (Table 3) indicates that for both tests, the RTs contribute more and hits less, to the combined measure of efficacy. Thus, within tests the correlation between RT and efficacy is quite high (*r* = .98 in STROOPm y *r* = .97 in CANUM; *p* < .01), explaining practically 95 % of the common variance. And between tests it remains high (*r* > .75; *p* < .01), which cannot be said for the percentage of hits. This makes sense when considering that, proportionally, few errors are made, even in the conditions with most interference. In addition, and in line with Kane and Engle’s (2003) approach, this result also suggests that the difficulty in both tests is due more to the attentional control requirements than to the failure to maintain the goal of the task. In this sense, therefore, both tests appear to offer a sufficiently conflicting context (in terms of the proportion of incongruent items), so that the difficulty mainly manifests itself in processing speed. Hence, the degree of efficacy is more related to response latencies (RT) than it is with the hit-rate, which tends to remain quite high.

In any case, as can be also seen in Table 3, we confirm that for the test as a whole, the latency and accuracy of the responses, as distinct manifestations of processing difficulty, do not maintain a simple inverse relationship. Correlations between the percentage of hits and RTs are non-existent for both tests, nor occur in any of the conditions. This is consistent with the idea that difficulties in handling interference can be reflected in a changeable way throughout the trials, with miss-rates or delays in RT being predominant depending on each case. This result is endorsed by research that emphasizes the incidence of contextual factors in the successive processing of items. But, also, compared with isolated measures of RT and response accuracy, it supports the value of the combined measure of efficacy as a very valid index of executive-attentional control, at least as a measure of overall performance.

On the other hand, the results also clearly reflect the expected relationship between CANUM and STROOPm with the criteria of general intelligence and working memory (Hypothesis 3), which assumes that interference management plays a key role in executive capacity underlying the relationship between these constructs (Table 4). In this respect we first highlight the high correlation between the criterion themselves (i.e., WMC and g factor). This serves to support their consistency as such and corresponds to that found in previous literature regarding their nature and their relationship. The predictive capacity of the two tests also corroborates the validity of the measure of efficacy, especially against the simple measure of accuracy. As can be seen, while all correlations are significant, the highest correspond to measures of efficacy (E) in both tests. Additionally, in line with Hypothesis 3, the correlations tend to be even higher in the CANUM test. However, given the equivalence in observed significance levels (*p* < .01), the apparent advantage of CANUM does not necessarily imply greater validity in reference to the underlying construct.

Table 4. Means (*SD*) in the criterion and Spearman correlations across various measures of difficulty in the STROOPm and CANUM tests (*N* = 128)

				STROOPm			*CANUM*		
	M (SD)	RAVEN	RST	%Hits	RT	E	*%Hits*	*RT*	*E*
RAVEN	*43.2 (7.6)*			.34**	-.46**	.52**	*.25***	*-.57***	*.64***
RST	*2.5 (0.4)*	.62**		.15	-.62**	.65**	*.05*	*-.59***	*.62***
RxST	*2.8 (0.5)*	.59**	.61**	.22*	-.45**	.48**	*.13*	*-.52***	*.56***

* *p* < .05; ** *p* < .01.

Although less obvious, the trend is equally noticeable with respect to the efficacy measure obtained in the particular interference conditions of each test. But perhaps what is striking here is that the magnitude of the correlations shows no clear correspondence with respect to the level of interference supposedly associated with each condition, not even in the measure of efficacy. On the contrary, it seems that in both tests the neutral conditions tend to produce the highest correlations, and the incongruent and the flankers the lowest. That is, it seems that managing more demanding interference levels is not more predictive of either criterion, g-factor or WMC, as expected. In other words, interference management itself does not appear to provide any special predictive status when compared to performance under neutral conditions.

Table 5. Spearman correlations between the criterion measures and the of efficacy measure (E) in the STROOPm and CANUM conditions (*N* =128-124)

	E STROOPm				*E CANUM*				
	C	N	S	I	*Nb*	*Na*	*Cs*	*Is*	*f*
RAVEN	.50**	.54**	.50**	.50**	*.61***	*59***	*.58***	*.61***	*.59***
RST	.65**	.68**	.62**	.60**	*.61***	*62***	*.57***	*.57***	*.53***
RxST	.47**	.47**	.47**	.46**	*.55***	*57***	*.52***	*.50***	*.52***

** *p* <. 01

This inconsistency, however, may be rather artificial, when considering that these are global measures of interference control. That is, they reflect the cumulative effect of the set of items or trials in each condition, regardless of local variations from trial to trial. This obviously conceals which components of performance are most relevant to the observed relationships between the criteria (WMC and g factor) and each condition of congruence. As we have already noted, both in latencies (dependent on the resolution of attentional conflict) and in errors (presumably more linked to goal neglect), there may be local variations in the demands —related to the actual sequencing of congruent-incongruence trials— conditioning the degree of relatedness between the constructs, as well as in the measures (speed and accuracy) in which that relationship is predominantly manifested (Egner, 2007; Kane & Engle, 2003). General measures, as aggregates or averages of local performances, obscure or hide these changeable influences trial-to-trial. In fact, it may be that the relationship between WMC and attentional control is not general, but instead rather selective and sensitive to change (Meier & Kane, 2013). Specifically, when the set of items contains a sufficient amount of conflict, the differences between high and low WMC participants in response accuracy tends to be considerably reduced. This, in our case, would explain the discriminatory and predictive weakness of this measure, both in STROOPm and in CANUM. In a similar vein, the contribution of control and memory aspects in predicting general fluid intelligence is also not very stable across several studies (Chuderski et. al., 2012; Chudersky, 2014).

In short, the management of each trial may be influenced not only by the level of interference associated with its congruence condition, but also by the particular sequences in which it is integrated. These sequences act as contexts that may alter or modulate the difficulty encountered. Therefore, the mean scores of accurate hits or response times, and thus the combined measure of efficacy (E) they generate, do not necessarily reflect in a "pure" way the associated interference condition. Thus the most demanding do not necessarily prove more predictive of our chosen criterion measures, general intelligence and working memory capacity (WMC). On the contrary, precisely because neutral trials are themselves characterised by not involving any level of interference, they may possibly be less susceptible to the influences of the local sequences in which they occur. That is, the speed-accuracy trade-off will be more stable or coherent with these kinds of items, and perhaps for this reason the efficacy measure across neutral conditions tends to be somewhat more predictive of cognitive competence measures than do the others.

Conclusions

Overall, therefore, the results obtained support the consistency of the two tests studied, STROOPm and CANUM, as reliable and valid measures of the executive-attentional capacities that supposedly underlie Stroop tasks, at least in terms of "efficacy" and in reference to global measures. But even with respect to particular conditions, the efficacy measure has been relatively robust, consistently reflecting the difficulty attributable to the levels of interference supposedly associated. In the case of CANUM, in addition, performance may be considered a relatively independent measure of any verbal factors, since it is based on a numerical stimulus and a manual response. In this sense, as we have seen, it most likely involves other interference effects ("Simon" and "flanker" types) that add to the basic Stroop effect, increasing its difficulty. This is in fact what has occurred, reflected in higher response times and lower overall efficacy in performance.

In predictive terms, however, the greater demands CANUM sets on managing interference do not seem to imply higher levels of predictive capacity for WMC and general intelligence. Although somewhat smaller, correlations with STROOPm were equally significant. In this respect, as we have seen, what stands out is the predictive superiority of the combined measure of efficacy, compared to the simpler measures of accuracy and response latency. This is because it is the measure that most consistently reflects the validity of both tests, both in terms of construct validity as well as criterion validity. In short, the efficacy of performance in both CANUM and STROOPm is here revealed as a valid measure of some central cognitive capacity that seems to underlie both tasks. And this central capacity, given the theoretical assumptions on which it is based, probably appertains to executive-attentional control required in general by the innovative or unfamiliar tasks and, in particular, those that are more demanding from a regulatory point-of-view under conditions of interference.

In relation to our objectives, therefore, we have verified the validity of CANUM, which makes it possible to value its independent nature regarding verbal and reading competencies. That is, unlike the Stroop task, CANUM is applicable regardless of the language and reading capability of individuals. Thus, its potential across populations on which it can be applied as a diagnostic and evaluation tool is clearly greater. This we understand as having a distinct advantage for its use in various applied contexts as well as research into the limitations of executive-attentional functions and the cognitive constructs these relate to.

Acknowledge.- We would like to thank the teachers and students of the "La Atalaya" secondary school Institute and the "Tomás Iglesias" and "Fernández Pozar" primary schools, from Conil (in Cádiz, Spain) for their collaboration and participation in this research.

References

Ackerman, P., Beier, M.E. & Boyle, M.O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131*(1), 30-60.

Algom, D., Dekel, A., & Pansky, A. (1996). The perception of number from the separability of the stimulus: The Stroop effect revisited. *Memory & Cognition, 24*, 557-572.

Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology: Human Experimental. Psychology, 49*(1), 5-28.

Banich, M.T., Milham, M.P., Atchley, R., Cohen, N.J., Webb, A., Wszalek, T., …Magin, R. (2000). FMRI studies of Stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. *Journal of Cognitive Neuroscience, 12*, 988-1000.

Bryant, P. (1996). Mathematical Understanding in the Nursery School Years. In *Learning and Teaching Mathematics. An Internacional perspective* (pp. 53-67), Psychology Press ltd, Publishers, UK.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19*, 273-293. doi: 10.1207/S15326942DN1903_3

Carriedo, N. & Rucian, M. (2009). Adaptación para niños de la prueba de amplitud lectora de Daneman & Carpenter (PALn). *Infancia y Aprendizaje, 32*(3), 449-485.

Chuderski, A. (2014). Which Working Memory Components Predict Fluid Intelligence: The Roles of Attention Control and Active Buffer Capacity. *Psychology, 5,* 328-339. http://dx.doi.org/10.4236/psych.2014.55043.

Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage Capacity Explains Fluid Intelligence while Executive Control Does Not. *Intelligence, 40*, 278-295. http://dx.doi.org/10.1016/j.intell.2012.02.010.

Colom, R., Abad, F.J., Rebollo, I. & Shih, P.C. (2005). Memory span and general intelligence: A latent-variable approach. *Intelligence, 33*(6), 623-642.

Conway, A.R.A., Kane, M.J., & Engle, R.W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7*(12), 547-552.

Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*(4), 450-466.

Elosúa, M.R., Gutiérrez-Martínez, F., García-Madruga, J.A., Luque, J.L., & Gárate, M. (1996). Adaptación española del Reading Span Test de Daneman y Carpenter. *Psicothema, 8,* 383-395.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19- 23.

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* (Vol. 44, pp. 145-199). NY: Elsevier.

Egner, T. (2007). Congruency sequence effects and cognitive control. *Cognitive, Affective, & Behavioral Neuroscience, 7*(4), 380-390.

Eriksen, B.A., & Eriksen, C.W. (1974). Effects of noise letters upon identification of a target letter in a non- search task. *Perception and Psychophysics 16,* 143–149. doi:10.3758/bf03203267

Eriksen, C.W. (1997). La tarea de los flancos y la competición de respuestas: Un instrumento útil para investigar una variedad de problemas cognitivos. *Estudios de Psicología, 57*, 93-108.

Friedman, N.P., Miyake, A., Corley, R.P., Young, S.E., DeFries, J.C., & Hewitt, J.K., (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*(2), 172-179.

García-Molina, A., Tirapu-Ustárroz, J., & Roig-Rovira, T. (2007). Validez ecológica en la exploración de las funciones ejecutivas. *Anales de Psicología, 23*-2, 289-299.

Gelman, R., & Meck, E. (1983). Preschooler's counting: principles before skill, *Cognition, 13*, 343-360.

Gutiérrez-Martínez, F. & Ramos, M. (2014). La memoria operativa como capacidad predictora del rendimiento escolar. Estudio de adaptación de una medida de memoria operativa para niños y adolescentes. *Psicología Educativa, 20*(1), 1-10. http://dx.doi.org/10.1016/j.pse.2014.05.001.

Gutiérrez-Martínez, F., García-Madruga, J. A., Carriedo, N., Vila, J. O., & Luzón, J. M. (2005). Dos pruebas de amplitud de memoria operativa para el razonamiento. *Cognitiva, 17*(2), 183-207.

Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica 136*, 189-202.

Hutchison, K.A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 645-662. http://dx.doi.org/10.1037/0278-7393.33.4.645.

Hutchison, K. A. (2011). The interactive effects of listwide control, item-based control, and working memory capacity on Stroop performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 851-860. http://dx.doi.org/10.1037/a0023437.

Imbrosciano, A., & Berlach, R.G. (2005). The Stroop test and its relationship to academic performance and general behavior of young students. *Teacher Development, 9*(1), 131-144. http://doi.org/10.1080/13664530500200234.

Kane, M.J., & Engle, R.W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychonomic Bulletin & Review, 9*, 637-671.

Kane, M.J., & Engle, R.W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132*, 47-70. doi:10.1037/0096-3445.132.1.47.

Kane, M.J, Conway, R. A., Hambrick, D. Z. & Engle, R.W. (2007). Variation in working memory capacity as variation in executive attention and control. In A.R.A. Conway, C. Jarrold, M.J. Kane, A. Miyake, & J.N. Towse (Eds.), *Variation in Working Memory* (pp. 21 - 48). NY: Oxford University Press.

Lezak, M.D., Howieson, D.B., & Loring, D.W. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.

Long, D.L., & Prat, C.S. (2002). Working memory and Stroop interference: An individual differences investigation. *Memory & Cognition, 30*, 294-301. doi:10.3758/BF03195290

Lu, C.H. & Proctor, R.W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review, 2*, 174-207. http://dx.doi.org/10.3758/BF03210959

MacLeod, C.M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163-203.

MacLeod, C.M. & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14*, 126-135.

Meier, M.E. & Kane, M.J. (2013). Working Memory Capacity and Stroop Interference. Global vs. Local Indices of Executive Control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39(3)*, 748-759.

Morey C.C., Elliott, E.M., Wiggers, J., Eaves, S.D., Shelton, J.T., Mall, J.T. (2012). Goal-neglect links Stroop interference with working memory capacity. *Acta Psychologica, 141*, 250-260.

Mueller, S.T. & Piper, B.J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods, 30*(222), 250-259. doi:10.1016/j.jneumeth.2013.10.024.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A. & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100.

Posner, M.I. & Snyder, C.R.R. (1975). Attention and cognitive control. In R.L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (55-85). Hillsdale, NJ: Erlbaum.

Protopapas, A., Archonti, A. & Skaloumbakas, C. (2007). Reading ability is negatively related to Stroop interference. *Cognitive Psychology*, *54*, 251-282. doi:10.1016/j.cogpsych.2006.07.003

Raven, J.C., Court, J.H. & Raven J. (1996). *Matrices progresivas.* Publicaciones de psicología aplicada. Madrid: TEA.

Schneider, W., Eschmann, A. & Zuccolotto, A. (2002). *E-prime User's Guide.* Pittsburgh, Psychology Software Tools Inc.

Sedó, M.A. (2004). Test de las cinco cifras: una alternativa multilingüe y no lectora al test de Stroop. *Revista de Neurología, 38*(9), 824-828.

Shipstead, Z. & Broadway, J.M. (2013). Individual differences in working memory capacity and the Stroop effect: Do high spans block the words? *Learning and Individual Differences, 26*, 191-195.

Shor, R.E. (197 I). Symbol processing speed differences and symbol interference effects in a variety of concept domains. *Journal of General Psychology, 85*, 187-205.

Simon, J.R. (1990). The effects of an irrelevant directional cue on human information processing. In R.W. Proctor & T.G. Reeve (Eds.), *Stimulus-response compatibility: An integrated perspective* (pp. 31-86). Amsterdam: North-Holland.

Stelzer, F., Andés, M.L., Canet-Juric, L. & Introzzi, I. (2016). Memoria de Trabajo e Inteligencia Fluida. Una Revisión de sus Relaciones. *Acta de Investigación Psicológica, 6*(1), 2302-2316.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 28*, 643-662.

Unsworth, N. & Spillers, G. J. (2010). Working memory capacity: Attention, memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, 62, 392-406. http://dx.doi.org/10.1016/j.jml.2010.02.001.

Windes, L.D. (1968). Reaction time for numerical coding and naming of numerals. *Journal of Experimental Psychology, 78*, 318-322.

Wolach, A.H., McHale, M.A. & Tarlea, A. (2004). Numerical stroop effect. *Perceptual and Motor Skills, 98*(1), 67-77.

(Article received: 13-07-2016; revised: 26-12-2016; accepted: 09-02-2017)

Appendix I. Conditions, cases and items in the CANUM test.

Figure 1a. Types of CANUM conditions in reference to the dimensions in conflict, Number (Nº) and Quantity (Qª)

Type of condition		included items	Examples	Description relative to the dimensions Nº-Qª of the stimulus
Congruent (Nº=Qª)	Congruent + Simon Congruent + flankers	*Cs* 9x2=18 *Cf* 3x2=6	113 – "one one" 112 – "one one" / "two ones" (error due to flankers)	base Nº occurs in the correct R option in Qª
Incongruent (Nº ≠ Qª)	Incongruent + Simon Incongruent + flankers	*Is* 9x2=18 *If* 3x2=6	133 – "one three" 122 – "one two" / "two twos" (error due to flankers)	base Nº occurs in the incorrect R option in Qª
Semi-neutral (Nº ≠ Qª)		*Na* 12x2=24	132 – "one three"	base Nº does not occur in the R options, but is within the range of Qª (1-4)
Neutral (Nº ≠ Qª)		*Nb* 12x2=24	152 – "one five"	base Nº does not occur in the R options, nor is within the range of Qª (5-8)
		Total=96		

Figure 1b. Matrix of cases in CANUM with the items selected in each condition

Nº Cª	1		2		3		4		5		6		7		8	
1	112	*Cf*	122	*If*	132	*Na*	142		152	*Nb*	162		172		182	
	113	*Cs*	123		133	*Is*	143	*Na*	153		163	*Nb*	173		183	
	114	*Cs*	124	*Na*	134		144	*Is*	154		164		174	*Nb*	184	
2	2111	*Is*	2221	*Cs*	2331		2441	*Na*	2551		2661		2771		2881	*Nb*
	2113	*Na*	2223	*Cf*	2333	*If*	2443		2553	*Nb*	2663		2773		2883	
	2114		2224	*Cs*	2334	*Na*	2444	*Is*	2554		2664	*Nb*	2774		2884	
3	31111	*Is*	32221	*Na*	33331	*Cs*	34441		35551		36661		37771	*Nb*	38881	
	31112		32222	*Is*	33332	*Cs*	34442	*Na*	35552		36662		37772		38882	*Nb*
	31114	*Na*	32224		33334	*Cf*	34444	*If*	35554	*Nb*	36664		37774		38884	
4	411111	*Is*	422221		433331	*Na*	444441	*Cs*	455551		466661	*Nb*	477771		488881	
	411112	*Na*	422222	*Is*	433332		444442	*Cs*	455552		466662		477772	*Nb*	488882	
	411113		422223	*Na*	433333	*Is*	444443	*Cs*	455553		466663		477773		488883	*Nb*

Appendix II. Conditions, cases and items in the STROOPm test.

Figure 2a. Types of STROOPm conditions in reference to the dimensions in conflict, word and color

Type of condition		items included		Description relative to the dimensions of stimulus (S), *color-word*
Congruent (color=word)	C	12x2=24	red red green (in red)	The S word occurs in the correct color R option
Incongruent (color ≠ word)	I	12x2=24	red red green (in green)	The S word occurs in the incorrect color R option
Semi-neutral (color ≠ word)	S	24x2=48	red blue green (in red)	The S word does not occur in the R options, even though it remains being a name of a color within the range provided.
Neutral (color ≠ word)	N	12x2=24	red polo green (in red)	The S word does not occur in the R options, nor is it a name of a color.
		Total=120		

Figure 2b. Matrix of STROOPm cases with the items used across each condition

Word Color	red		green		yellow		blue		(neutral words)	
red	red red green	C	red green yellow	S	red yellow green	S	red blue green	S	red lbook yellow	N
	red red yellow	C	red green green	I	red yellow yellow	I	red blue red	I	red sleeve blue	N
	red red blue	C	red green blue	S	red yellow blue	S	red blue yellow	S	red crib green	N
green	green red yellow	S	green green yellow	C	green yellow red	S	green blue red	S	green well blue	N
	green red red	I	green green red	C	green yellow yellow	I	green blue blue	I	green polo yellow	N
	green red blue	S	green green blue	C	green yellow blue	S	green blue yellow	S	green mole red	N
yellow	yellow red green	S	yellow green red	S	yellow yellow red	C	yellow blue red	S	yellow cow blue	N
	yellow red red	I	yellow green green	I	yellow yellow green	C	yellow blue blue	I	yellow dinner red	N
	yellow red blue	S	yellow green blue	S	yellow yellow blue	C	yellow blue green	S	yellow cute green	N
blue	blue red green	S	blue green red	S	blue yellow red	S	blue blue yellow	C	blue bank green	N
	blue red red	I	blue green green	I	blue yellow yellow	I	blue blue green	C	blue hair yellow	N
	blue red yellow	S	blue green yellow	S	blue yellow green	S	blue blue red	C	blue finger red	N