2-1-2022

The Emotional Word-Emotional Face Stroop task in the ABCD study: Psychometric validation and associations with measures of cognition and psychopathology

Harry R Smolker

Kai Wang

Monica Luciana

James M Bjork

Raul Gonzalez

See next page for additional authors

Authors

Harry R Smolker, Kai Wang, Monica Luciana, James M Bjork, Raul Gonzalez, Deanna M Barch, Erin C McGlade, Roselinde H Kaiser, Naomi P Friedman, John K Hewitt, and Marie T Banich

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

journal homepage: www.elsevier.com/locate/dcn

The Emotional Word-Emotional Face Stroop task in the ABCD study: Psychometric validation and associations with measures of cognition and psychopathology

Harry R. Smolker [a,*], Kai Wang [b,c], Monica Luciana [d], James M. Bjork [e], Raul Gonzalez [f], Deanna M. Barch [g], Erin C. McGlade [h], Roselinde H. Kaiser [a,i,j], Naomi P. Friedman [a,i,k], John K. Hewitt [i,k], Marie T. Banich [a,i]

[a] Institute of Cognitive Science, University of Colorado Boulder, Boulder, CO 80309, USA
[b] Key Laboratory of Brain, Cognition and Education Sciences, Ministry of Education, China
[c] School of Psychology, Center for Studies of Psychological Application, and Guangdong Key Laboratory of Mental Health and Cognitive Science, South China Normal University, Guangzhou 510631, China
[d] Department of Psychology, University of Minnesota, Minneapolis, MN 55455, USA
[e] Department of Psychiatry, Virginia Commonwealth University, Richmond, VA 23298, USA
[f] Center for Children and Families, Department of Psychology, Florida International University, Miami, FL 33199, USA
[g] Department of Psychological & Brain Sciences, Washington University, St. Louis, MO 63130, USA
[h] Department of Psychiatry, University of Utah School of Medicine, Salt Lake City, UT 84108, USA
[i] Department of Psychology & Neuroscience, University of Colorado Boulder, Boulder, CO 80309, USA
[j] Renee Crown Wellness Institute, University of Colorado Boulder, Boulder, CO 80309, USA
[k] Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO 80309, USA

ARTICLE INFO

ABSTRACT

Characterizing the interactions among attention, cognitive control, and emotion during adolescence may provide important insights into why this critical developmental period coincides with a dramatic increase in risk for psychopathology. However, it has proven challenging to develop a single neurobehavioral task that simultaneously engages and differentially measures these diverse domains. In the current study, we describe properties of performance on the Emotional Word-Emotional Face Stroop (EWEFS) task in the Adolescent Brain Cognitive Development (ABCD) Study, a task that allows researchers to concurrently measure processing speed/attentional vigilance (i.e., performance on congruent trials), inhibitory control (i.e., Stroop interference effect), and emotional information processing (i.e., difference in performance on trials with happy as compared to angry distracting faces). We first demonstrate that the task manipulations worked as designed and that Stroop performance is associated with multiple cognitive constructs derived from different measures at a prior time point. We then show that Stroop metrics tapping these three domains are preferentially associated with aspects of externalizing psychopathology and inattention. These results highlight the potential of the EWEFS task to help elucidate the longitudinal dynamics of attention, inhibitory control, and emotion across adolescent development, dynamics which may be altered by level of psychopathology.

1. Introduction

Understanding the interactions between cognition and socioemotional functioning during adolescence may unlock key insights into why adolescence coincides with a dramatic increase in the prevalence and severity of mental illness (Powers and Casey, 2015; Steinberg, 2010).

Studies evaluating neurocognitive traits associated with mental illness often report altered processing speed (e.g., Buyukdura et al., 2011), attention (e.g., Platt et al., 2017), and inhibitory control (e.g., Rock et al., 2014), as well as atypical processing of emotional information (e. g., Suslow et al., 2020). Most notably, characteristic of most all mental illnesses are deficits in behavioral inhibition (McTeague et al., 2016), as

well as gray matter abnormalities (Goodkind et al., 2015) and functional activation decrements (McTeague, et al., 2017) in the brain systems that govern behavioral inhibition. Of critical interest, therefore, is advancing understanding of associations between neurocognitive development and mental illness, which in turn requires developing tools for this research.

Researchers have employed a wide range of psychological tasks to better understand adolescent development, but such studies have often focused on a single construct of interest instead of simultaneously capturing the distinct measures of processing speed/attention, cognitive control, and emotional information processing in a single task. Introduced over 80 years ago, Stroop paradigms (Stroop, 1935) are now among the most widely studied experimental tasks to study behavioral inhibition in cognitive psychology. Theorized to tap the ability to override a prepotent response in favor of a less automatic one (i.e., inhibitory control), Stroop tasks involve the presentation of stimuli that contain two sources of information (e.g., color words and ink color) that vary in task relevance (i.e., task-irrelevant and task-relevant). On "congruent" trials, the two sources of information are redundant. On "incongruent" trials they are in conflict requiring participants to engage cognitive control to respond according to the task-relevant but generally less prepotent source of information. For example, on incongruent trials of the classic color-word Stroop task, participants view color words (e.g., the word "blue") in a different ink color (e.g., yellow ink). Responding to the ink color ("yellow") requires control to overcome the more automatic and prepotent tendency to engage in word reading ("blue") that would lead to an incorrect response. Participants are consistently slower on incongruent trials as compared to congruent trials across several variants of the task, a phenomenon referred to as the Stroop interference effect. Though previous research has generally focused on reaction time-based interference effects (MacLeod, 1991), a similar effect is frequently observed for accuracy, with participants consistently being less accurate on incongruent as compared to congruent trials (Scarpina and Tagini, 2017).

Stroop tasks have provided a multifaced window into the nature of cognitive control, including its latent structure (Miyake et al., 2000), neural substrates (Banich, 2019; Freund et al., 2020) and deviations observed as a function of psychopathology (Becker et al., 2001; Henik and Salo, 2004; Snyder et al., 2015; Williams et al., 1996). Specifically, when compared to healthy controls, patient groups show evidence of psychomotor slowing and reduced attentional vigilance as indexed by slower reaction times and more errors on congruent trials (anxiety: Becker et al., 2001; depression: Nuno et al., 2021; ADHD: Schwartz and Verhaeghen, 2008;), as well as impaired inhibitory control, as indexed by relatively increased Stroop interference effects (anxiety: Becker et al., 2001; depression: Epp et al., 2012; Nuno et al., 2021; ADHD: Lansbergen et al., 2007).

Other Stroop variants have featured task-irrelevant words that are emotional in nature, expanding the utility of the Stroop task to a probe individual differences relevant to socioemotional functioning (Epp et al., 2012). In such Stroop variants, participants are presented with words in different ink colors but instead of the words being color words (e.g.," blue"), they are either emotion words (e.g. "happy"), emotion-evoking words (e.g. "murder"), or neutral words (e.g. "doorknob"). A recent meta-analysis found that depressed patients showed larger interference effects when the task-irrelevant information is negatively valenced as opposed to positive or neutral whereas post-traumatic stress disorder is associated with impaired performance when this information is positively valenced or trauma related (Joyal et al., 2019). On the other hand, individuals with externalizing conditions such as ADHD show evidence of impairments that are not valence specific but instead more general in nature (Posner et al., 2011). However, there is inconsistency in the literature regarding the nature of disorder-specific emotional processing biases potentially due to a lack of power to detect such effects, which can be overcome with larger data sets, such as the ABCD study.

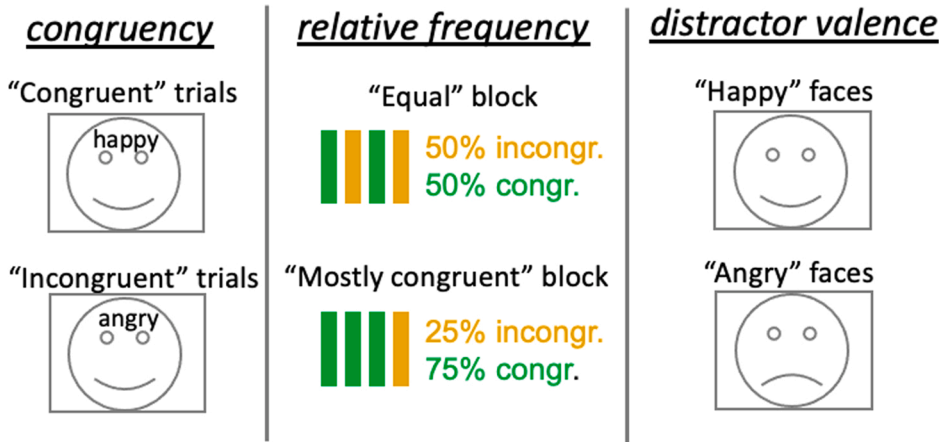While it may tap biases in emotional information processing, the color-emotional word Stroop task does not engage cognitive control to the same degree as the classic-color word Stroop as there is no actual response conflict between task-relevant and task-irrelevant dimensions. That is, in the color-emotional word Stroop task, the emotional words may be more salient than the ink color, but the meaning of the word is not response relevant as in the classic color-word Stroop. A variant of the emotional Stroop task, the Emotional Word-Emotional Face Stroop task, may better allow researchers to simultaneously measure attentional processes, inhibitory control, and emotional information processing in a way not possible in either the color-word or color-emotional word Stroop tasks alone. This task involves the presentation of an emotional word (e.g., "joy") overlaid on an image of an emotional face of either the same (i.e., congruent) or different (i.e., incongruent) emotional valence. Blending features of the classic color-word and color-emotional word Stroop tasks, the Emotional Word-Emotional Face Stroop task is well suited to simultaneously differentiate between processing speed/vigilance and inhibitory control, while also allowing for the exploration of valence-specific emotional processing biases. To meet task demands, participants must suppress the prepotent tendency to process the face (Beall and Herbert, 2008) and instead respond to the content of the less salient emotional word. Importantly, as is the case in the classic color-word Stroop, both the task-relevant and task-irrelevant stimulus dimensions contain information that is response related (namely, emotional valence), allowing for the computation of an interference effect that quantifies individual differences in inhibitory control that are distinct from general processing speed.

Due to these interpretive advantages, the Emotional Word-Emotional Face Stroop (EWEFS) task was included at the year 1 follow-up of the Adolescent Brain Cognitive Development (ABCD) longitudinal study (Bjork et al., 2017; Luciana et al., 2018; Volkow et al., 2018), which has recruited a cohort of over 11,000 sociodemographically-diverse American youth to track their neurocognitive development from ages 9–10 to 19–20, by administering annual neurocognitive and mental health phenotypic assessments, together with biennial multimodal brain scans. Another important feature of the ABCD EWEFS task is that the proportion of incongruent to congruent trials varies between its two blocks (50/50% vs. 25/75%). Prior work indicates that Stroop interference is reduced with an increasing proportion of incongruent relative to congruent trials (Logan and Zbrodoff, 1979). Subsequent research concluded that when the proportion of incongruent trials is higher, the conflict in the stimuli themselves provides an implicit reminder of the task goal ("Red word – blue ink. Oh yes, respond on the basis of ink color") (Kane and Engle, 2003). When incongruent trials are infrequent, the Stroop interference effect additionally indexes the degree to which an individual can effectively maintain task goals without such implicit reminding. As such, more cognitive control will be required when incongruent trials are relatively infrequent (25/75% block) than when the frequency of incongruent and congruent trials is equivalent (50/50% block). Fig. 1 for graphically depicts these three main task manipulations and individual differences measures of interest of the ABCD EWEFS task.

One of the major goals of the ABCD study is to address head-on the issues of rigor and reproducibility due to small sample sizes and inflated effect sizes that plague the psychological sciences (e.g., Ioannidis, 2005). Towards this aim, the ABCD dataset offers a definitive sample size for confirmation of key anticipated differences in behavior within-task, as well as between-subjects relationships with other phenotypes. As such, while much of the current report is focused on replicating common within-task manipulations that have been relatively well established, this focus provides an unprecedented lens into the true effect sizes of manipulations frequently utilized in emotional Stroop paradigms. Such information may serve as an important resource to researchers planning to utilize similar tasks to investigate issues surrounding cognition, emotion, and their interactions, while also providing an important validation of the publicly available EWEFS data.

In this paper, we 1) characterize performance on the EWEFS task, including testing for predicted effects of task manipulations; 2) evaluate

A) Task manipulations

Fig. 1. Task manipulations and measures of interest. *Panel A:* Graphical depiction of three task manipulations of interest from the ABCD Emotional Word-Emotional Face Stroop task, including "congruency", "relative frequency", and "distractor valence" manipulations. *Panel B:* Domain-specific measures of interest from the ABCD Emotional Word-Emotional Face Stroop task, including "processing speed/attention", "inhibitory control", and "emotional information processing". Congr. = congruent; incongr. = incongruent; Diffs = differences; RT = reaction time; acc. = accuracy.

the degree to which performance is associated with cognitive principal components from the baseline assessment that occurred one year prior; and 3) determine the degree to which performance on this task is associated with dimensions of psychopathology from the baseline assessment. The structure of the ABCD EWEFS task enabled the testing of several hypotheses. First, with regards to task manipulations we predicted i. a Stroop interference effect, in which participants are slower and less accurate on incongruent trials as compared to congruent trials (i.e., *congruency manipulation*), which ii. would be larger in a mostly congruent block as compared to equal frequency block (i.e., *frequency manipulation*) (Hutchison, 2011; Kane and Engle, 2003; Logan and Zbrodoff, 1979) and that iii. participants' performance will be slower and less accurate when the task irrelevant face is angry as compared to happy (i.e., *distractor valence manipulation*) in line with evidence suggesting negative emotional information may engender a deeper degree of processing that is more difficult to disengage from than positive emotional information (Norris et al., 2021). Second, with regards to the associations with other measures, we predicted that while Stroop performance will be associated to varying degrees with each of the three cognitive components at the baseline timepoint (Thompson et al., 2019), task manipulations within the EWEFS will capture additional processes related to emotion above and beyond those more general cognitive abilities. Third, with regards to psychopathology symptoms, we predict that slower processing speed (i.e., RT on congruent trials), impaired inhibitory control (i.e., larger interference effect), and alterations in emotional information processing (i.e., difference between performance on happy vs angry faces) will be associated with a range of symptom dimensions related to psychopathology (Amir et al., 2002; Epp et al., 2012; Koven et al., 2003; Richards et al., 2000).

2. Methods

2.1. Participants

Participants for all analyzes were drawn from the ABCD study® data release 3.0, including data from baseline (i.e., demographics, cognitive components, Child Behavior Checklist) and year 1 (i.e., EWEFS task). As part of the ABCD protocol, participant consent and parent assent were obtained prior to participation. Of the 11,878 participants enrolled in the ABCD study at baseline, we excluded any participants who were missing Stroop data from the pre-tabulated NDA 3.0 release (1079 participants). Additionally, participants with mean RT less than 200 ms (2 participants) or greater than 2000 ms (1 participant) for any of the Stroop conditions were excluded from analyses. This resulted in a final sample of 10,796 participants (for demographic information, see Table 1). To account for potential non-independence between participants due to familial relations (i.e., twins or sibling), all mixed effects models included family as a random effect nested within data acquisition site.

2.2. Emotional Word-Emotional Face Stroop task

The EWEFS task measures cognitive control over emotionally distracting information. On each trial, individuals categorize the emotional valence of a word as either positive or negative via a button press while ignoring a distracting face of either the same (i.e., congruent) or opposite (i.e., incongruent) valence. The task includes 96 trials, divided into two test blocks, each of which contain 48 trials. In one block there are 75% congruent trials and 25% incongruent trials (termed the "mostly congruent block"). In the other block there is an equal percentage of congruent and incongruent trials (termed the "equal block"). On each

Table 1
Sample demographics. Demographics of total sample after exclusions. Percentages may not add up to 100% because of participants declining to answer certain questions.

	Total
Number of participants	10,796
Female = yes (%)	5175 (48%)
Race/Ethnicity (%)	
Hispanic	2129 (20%)
White	5799 (54%)
Black	1499 (14%)
Asian	233 (2%)
Multiracial	1138 (11%)
Highest Parental Ed. (%)	
< HS Diploma	501 (5%)
HS Diploma/GED	939 (9%)
Some College	1312 (12%)
Associate Degree	1405 (13%)
Bachelor's Degree	2796 (26%)
Graduate Degree	3825 (35%)
Household Married (%)	7421 (69%)
Household Income (%)	
< 50 K	2777 (26%)
≥ 50 K & < 100 K	2834 (26%)
≥ 100 K	4309 (40%)
Site (%)	
site01	353 (3%)
site02	530 (5%)
site03	546 (5%)
site04	697 (6%)
site05	337 (3%)
site06	546 (5%)
site07	315 (3%)
site08	325 (3%)
site09	383 (4%)
site10	679 (6%)
site11	382 (4%)
site12	525 (5%)
site13	671 (6%)
site14	572 (5%)
site15	410 (4%)
site16	941 (9%)
site17	546 (5%)
site18	366 (3%)
site19	505 (5%)
site20	670 (6%)
site21	500 (5%)

trial, participants have 2000 ms to make a response. In the publicly available NDA data, non-responses were treated as error trials. Facial stimuli for this task were drawn from a set of adolescent emotional faces used in previous research (Guyer et al., 2008). Two images from 48 different posers were used, including one image with a happy facial expression and another with an angry expression. Each poser appeared once per block as well as once per congruency condition. All face images were of white adolescents, and the task was administered on an iPad through the Inquisit platform (www.millisecond.com) as part of the ABCD year 1 protocol. Participants responded by pressing buttons on the iPad touch screen with their left and right index fingers, with one finger indicating the emotional word is a "bad" emotion and the other indicating the word is a "good" emotion. The response mapping was counterbalanced across participants. When not initiating a response, participants were instructed to rest their fingers on boxes on the iPad screen immediately beneath the response buttons.

To test our specific hypotheses within the Stroop data, we created derived variables from the NDA 3.0 pre-tabulated data. Specifically, we utilized mean RT and accuracy data by Stroop task condition to compute derived variables. These include interference effects in the mostly congruent block (i.e., performance on incongruent trials minus performance on congruent trials divided by performance on congruent trials) and performance on happy faces in the equal block minus performance on angry faces in the equal block averaging across congruent and

incongruent trials. We focus on the interference effect from the mostly congruent block because of evidence that interference when incongruent trials are infrequent captures inherent individual cognitive control abilities without the implicit task reminders that are engendered when incongruent trials are more frequent. Specifically, the mostly congruent block likely places a greater demand on goal-maintenance, with prominent theories suggesting that goal-maintenance and inhibitory control may not be dissociable from one another (Friedman and Miyake, 2017). Furthermore, we focus our emotional information processing measures on performance in the equal block to ensure that we are averaging across an equal number of congruent and incongruent trials as the mostly congruent block only has six incongruent trials of either valence, whereas the equal block had 12 trials for all four possible congruency-valence pairings. We detected several outliers (i.e., ± 3 standard deviations) on our Stroop measures of interest (congruent RT outliers = 20, congruent accuracy outliers = 191, interference RT outliers = 57, interference accuracy outliers = 143, happy minus angry RT outliers = 68, happy minus angry accuracy outliers = 148). To determine any potential effects of outliers, we ran all analyzes both with and without these outliers. These results found no notable differences in the significance and direction of effects. As such, all reported results are from analyses that included the full sample.

2.3. Cognitive principal components at the baseline timepoint

To measure individual differences in cognitive abilities from the prior year, we utilized three distinct cognitive principal components (PCs) for each participant derived from the neurocognitive battery administered at the baseline time point. This battery consisted of tasks from the NIH Toolbox (http://www.nihtoolbox.org), including the Picture Vocabulary Task (Gershon et al., 2014), the Oral Reading Recognition Task, the Pattern Comparison Processing Speed Test (Carlozzi et al., 2015), the List Sorting Working Memory Test (Tulsky et al., 2014), the Picture Sequence Memory Test (Bauer et al., 2013), the Flanker Task (Eriksen and Eriksen, 1974), and the Dimensional Change Card Sort Task (Zelazo et al., 2013), as well as the Rey Auditory Verbal Learning Test (Daniel et al., 2014) and the Little Man Test (Acker and Acker, 1982). These PCs were calculated in the same manner described by Thompson et al. (2019) but across the entire ABCD sample. As such, while the specific loadings of tasks on each component and individual differences scores may slightly differ between the current project and those reported by Thompson et al. (2019), the overall pattern of results are similar (see Supplemental Fig. 1 for loadings of indicators on the three cognitive components). The three PCs have been suggested to represent general cognitive ability/crystallized reasoning, executive functioning/processing speed, and learning/memory (Thompson et al., 2019). We detected very few outlying factor scores values (1 for PC1, 4 for PC2, and 0 for PC3), all of which were withing + /- 3.25 SDs. Given the large sample size and relatively minor nature of these outlying values, we did not remove these participants from related analyses.

2.4. Measures of internalizing and externalizing psychopathology at the baseline timepoint

To ascertain each participant's level of symptom severity along several dimensions of psychopathology, we utilized parent report of the Child Behavior Checklist (CBCL; Achenbach and Ruffle, 2000) obtained at the baseline time point collected one year prior to the EWEFS task. Specifically, we used the CBCL syndrome subscales, which capture severity for eight distinct dimensions associated with psychopathology, including anxious/depressed, withdrawn/depressed, somatic complaints, social problems, thought problems, attention problems, rule-breaking behavior, and aggressive behavior. Additionally, we utilized three composite measures of psychopathology symptoms capturing more general symptom dimensions, including internalizing symptoms (sum of anxious/depression, withdrawn, and somatic complaints),

externalizing symptoms (sum of rule-breaking and aggressive behaviors), and total problems (sum of all syndrome subscales). To capture the full available range of psychopathology, we did not remove any participants due to outlying CBCL scores.

2.5. Analyses

We carried out our analyses through mixed effect modeling as implemented in the "lme4" package (version 1.1-26) within R-Studio (version 4.0.2). All models were run with random effects of site and family nested within site, as well as fixed effects of demographic variables, including age, sex assigned at birth, marital status of household, household income, highest household education, and race/ethnicity. Repeated measures ANOVA analyses included subject as an additional random effect. In some models we controlled for the three cognitive factor scores derived from cognitive tasks at the baseline timepoint. Mixed effects models excluded participants who were missing data for any of the dependent or independent variables in a given model through the "na.action = na.omit" option in lme4.

First, we evaluated the degree to which the task manipulations worked as hypothesized through mixed repeated measures ANOVAs testing for effects of congruency, relative frequency of incongruent trials, and valence of the distractor face, as well as their interactions, on overall RT and accuracy. Second, we investigated the relations of Stroop performance to cognitive factors derived from the year prior through both repeated measures ANOVAs and correlational analyses (i.e., Spearman's rank). Third, we used correlational analyses and mixed effects models to evaluate the degree to which psychopathology symptoms are associated with alterations in multiple aspects of Stroop performance, including performance on congruent trials, the interference effect in the mostly congruent block, and the difference in performance on happy as compared to angry faces in the equal block. To further determine the degree to which specific symptoms are associated with Stroop performance over and above more general aspects of psychopathology, we included all eight non-composite CBCL syndrome scales as predictors of our Stroop performance metrics of interest, allowing us to capture associations between Stroop performance and symptoms that are unique to that symptom as compared to others. To account for the substantial skew in accuracy on congruent trials and the accuracy interference effect, in models in which these variables were the dependent variable we modeled a Gamma distribution through a log link function. Additionally, all correlation analyses were run as Spearman correlations to account for non-normality in variables.

To aid in the interpretation of analyses, we report standardized effect size metrics. For ANOVA analyses we report the partial eta squared (η^2) (.01 = small effect; 0.06 = medium effect; 0.14 = large effect), whereas for multiple regression models we report standardized beta coefficients for continuous independent variables (i.e., cognitive PCs, CBCL scores). Standardized beta coefficients for fixed effects of interest were computed by multiplying the unstandardized coefficients for each fixed effect by the ratio of the standard deviation of the independent variables over the standard deviation of the dependent variables.

3. Results

3.1. Effects of task conditions on performance

To test for our predicted effects of congruency, distractor valence and the interaction between congruency and frequency on performance, we conducted repeated measures ANOVAs with RT and accuracy as dependent variables, and congruency, frequency, distractor valence, their interactions, and demographic measures as independent variables (for demographics see Table 1; for means by condition, see Table 2 and Fig. 2; for descriptive statistics of variables of interest, see Supplemental Table 1). We focus our results on our a priori effects of interest, including main effects of congruency and distractor valence, as well as the

Table 2

Means by task manipulation conditions in the Emotional Word-Emotional Face Stroop task. Con = congruent; Inc = incongruent; Eq = equal block; MC = mostly congruent block; Con & Inc = averaging across congruent and incongruent trials; Eq & MC = averaging across trials in the equal and mostly congruent block; Angry & Happy = averaging across trials with angry and happy valenced faces. SD = standard deviation; ms = milliseconds.

Congruency	Relative frequency	Distractor valence	Mean accuracy (SD) [%]	Mean RT (SD) [ms]
Con & Inc	Eq & MC	Angry & Happy	94.0(9.9)	1139(168)
Con	Eq & MC	Angry & Happy	96.1(6.9)	1088(124)
Inc	Eq & MC	Angry & Happy	92.4(11.8)	1165(135)
Con & Inc	Eq	Angry & Happy	94.1(9.3)	1130(161)
Con & Inc	MC	Angry & Happy	94.2(10.5)	1148(175)
Con & Inc	Eq & MC	Angry	94.2(9.9)	1144(167)
Con & Inc	Eq & MC	Happy	94.2(9.8)	1133(170)
Con	Eq	Angry	95.3(8.1)	1113(160)
Inc	Eq	Angry	93.3(9.9)	1149(164)
Con	Eq	Happy	96.0(7.4)	1082(156)
Inc	Eq	Happy	92.0(10.8)	1173(158)
Con	MC	Angry	96.2(0.6.3)	1112(148)
Inc	MC	Angry	91.8(13.1)	1201(184)
Con	MC	Happy	97.0(5.6)	1077(142)
Inc	MC	Happy	91.8(13.1)	1200(188)

interaction between congruency and frequency. We refer the readers to Tables 3–6 for ANOVA tables of full models.

In models predicting RT, there was a moderately sized and highly significant main effect of congruency in which participants were slower on incongruent as compared to congruent trials ($\eta^2 = .177$, p < .00001) and a small but significant main effect of valence in which participants were slightly slower when angry faces were the task-irrelevant distractor as compared to happy faces, but the effect size was miniscule ($\eta^2 = .003$, p < .00001). Furthermore, as predicted, the interaction between congruency and frequency was highly significant though small, with slower performance on incongruent trials within the mostly congruent as compared to the equal blocks ($\eta^2 = .010$, p < .00001) and no difference between performance for congruent trials across the two block types (see Table 2 and Supplementary Fig. 1 for condition means and Table 3 for full ANOVA results). To determine the degree to which task manipulations affected RT over and above individual differences in baseline measure of cognition, we carried out additional ANOVAs that include the three cognitive PCs as independent variables. When doing so, all task effects on RT remained (Table 4).

For accuracy, there was a medium sized main effect of condition in which participants were more accurate on congruent as compared to incongruent trials ($\eta^2 = .060$, p < .00001). Main effects of relative frequency ($\eta^2 = .000$, p = .099) and valence ($\eta^2 = .000$, p = .502) on accuracy were negligible. All interaction effects were small but highly significant, including the congruency by relative frequency interaction, our primary interaction of interest. As predicted, this interaction effect revealed a small but significant effect in which the difference between accuracy on congruent and incongruent trials was larger in the mostly congruent as compared to equal blocks ($\eta^2 = .004$, p < .00001), with greater accuracy on congruent and lower accuracy on incongruent trials in the mostly congruent as compared to the equal block (see Table 2 and Supplementary Fig. 1 for condition means and Table 5 for ANOVA results). To determine the degree to which task manipulations affected accuracy over and above individual differences in baseline measure of cognition, we carried out additional ANOVAs that included the three cognitive PCs as independent variables. When doing so, all task effects on accuracy remained (Table 6).
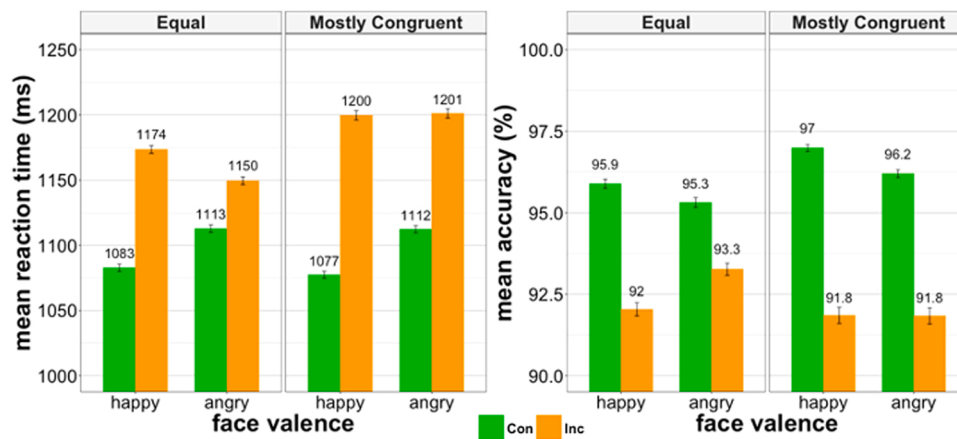
Stroop Performance by Task Manipulations

Fig. 2. Bar plot of condition means for reaction time and accuracy. Green bars = congruent trials, orange bars = incongruent trials. Error bars show 95% confidence interval. Values show condition means prior to controlling for covariates. Con = congruent; Inc = incongruent; ms = milliseconds.

Table 3

ANOVA table for reaction time. Sum Sq. = total sum of squares; Mean Sq = mean squares; NumDF = numerator degrees of freedom; DenDF = denominator degrees of freedom; η^2 = partial eta squared; congr. = congruency manipulation; freq. = frequency; r. freq. = relative frequency manipulation; val. = distractor valence manipulation.

Domain	Variable	Sum Sq	Mean Sq	NumDF	DenDF	F-value	η^2	p-value
task	*congruency (congr.)*	*154,640,861*	*154,640,861*	*1*	*75,537*	*16,191.3*	*.177*	*< 0.00001*
task	*relative freq. (r. freq.)*	*6,979,469*	*6,979,469*	*1*	*75,537*	*730.8*	*.010*	*< 0.00001*
task	*distractor valence (val.)*	*2,423,763*	*2,423,763*	*1*	*75,537*	*253.8*	*.003*	*< 0.00001*
task	*congr. by r. freq.*	*9,456,413*	*9,456,413*	*1*	*75,537*	*990.1*	*.013*	*< 0.00001*
task	*congr. by val.*	*10,341,708*	*10,341,708*	*1*	*75,537*	*1082.8*	*.014*	*< 0.00001*
task	*freq. by val.*	*1,243,372*	*1,243,372*	*1*	*75,537*	*130.2*	*.002*	*< 0.00001*
task	*congr. by r. freq. by val.*	*567,247*	*567,247*	*1*	*75,537*	*59.4*	*.001*	*< 0.00001*
demo	*age*	*4,988,411*	*4,988,411*	*1*	*10,394*	*522.3*	*.048*	*< 0.00001*
demo	*sex*	*132,164*	*132,164*	*1*	*10,571*	*13.8*	*.001*	*.0002*
demo	*race/ethnicity*	*733,878*	*183,469*	*4*	*6407*	*19.2*	*.012*	*< 0.00001*
demo	*max. parental education*	*10,239*	*10,239*	*1*	*8941*	*1.1*	*.000*	*0.301*
demo	*total parental income*	*7486*	*7486*	*1*	*9064*	*0.8*	*.000*	*0.376*
demo	*parent marital status*	*357,359*	*59,560*	*6*	*9138*	*6.2*	*.004*	*< 0.00001*

Table 4

ANOVA table for reaction time controlling for cognitive factor scores. Sum Sq. = total sum of squares; Mean Sq = mean squares; NumDF = numerator degrees of freedom; DenDF = denominator degrees of freedom; η^2 = partial eta squared; congr. = congruency manipulation; freq. = frequency; r. freq. = relative frequency manipulation; val. = distractor valence manipulation.

Domain	Variable	Sum Sq	Mean Sq	NumDF	DenDF	F-value	η^2	p-value
cognitive	*PC1-general cog. ability*	*1,083,441*	*1,083,441*	*1*	*8976*	*114.0*	*.013*	*< 0.00001*
cognitive	*PC2- exec. func./speed*	*13,121,875*	*13,121,875*	*1*	*9364*	*1380.2*	*.128*	*< 0.00001*
cognitive	*PC3- learning/memory*	*1,665,194*	*1,665,194*	*1*	*9193*	*175.2*	*.019*	*< 0.00001*
task	*congruency (congr.)*	*143,994,268*	*143,994,268*	*1*	*70,413*	*15,145.8*	*.177*	*< 0.00001*
task	*relative freq. (r. freq.)*	*6,192,861*	*6,192,861*	*1*	*70,413*	*651.4*	*.009*	*< 0.00001*
task	*distractor valence (val.)*	*2,211,859*	*2,211,859*	*1*	*70,413*	*232.7*	*.003*	*< 0.00001*
task	*congr. by r. freq.*	*8,818,243*	*8,818,243*	*1*	*70,413*	*927.5*	*.013*	*< 0.00001*
task	*congr. by val.*	*9,769,951*	*9,769,951*	*1*	*70,413*	*1027.6*	*.014*	*< 0.00001*
task	*r. freq. by val.*	*1,131,025*	*1,131,025*	*1*	*70,413*	*119.0*	*.002*	*< 0.00001*
task	*congr. by r. freq. by val.*	*508,895*	*508,895*	*1*	*70,413*	*53.5*	*.001*	*< 0.00001*
demo	*age*	*982,959*	*982,959*	*1*	*9899*	*103.4*	*.010*	*< 0.00001*
demo	*sex*	*373,822*	*373,822*	*1*	*9781*	*39.3*	*.004*	*< 0.00001*
demo	*race/ethnicity*	*1,049,341*	*262,335*	*4*	*6333*	*27.6*	*.017*	*< 0.00001*
demo	max. parental education	2538	2538	1	7826	0.3	.000	.605
demo	total parental income	3802	3802	1	8430	0.4	.000	.527
demo	parent marital status	47,855	7976	6	8547	0.8	.001	.540

Table 5
ANOVA table for accuracy. Sum Sq. = total sum of squares; Mean Sq = mean squares; NumDF = numerator degrees of freedom; DenDF = denominator degrees of freedom; η^2 = partial eta squared; congr. = congruency manipulation; freq. = frequency; r. freq. = relative frequency manipulation; val. = distractor valence manipulation.

Domain	Variable	Sum Sq	Mean Sq	NumDF	DenDF	F-value	η^2	p-value
task	*congruency (congr.)*	*31.976*	*31.976*	*1*	*75,537*	*4878.8*	*.061*	*< 0.00001*
task	*relative freq. (r. freq.)*	0.018	0.018	1	75,537	2.7	.000	0.099
task	*distractor valence (val.)*	0.003	0.003	1	75,537	0.5	.000	0.502
task	*congr. by r. freq.*	*1.749*	*1.749*	*1*	*75,537*	*266.9*	*.004*	*< 0.00001*
task	*congr. by val.*	0.9	0.9	1	75,537	137.3	.002	< 0.00001
task	*r. freq. by val.*	*0.287*	*0.287*	*1*	*75,537*	*43.8*	*.001*	*< 0.00001*
task	*congr. by r. freq. by val.*	*0.145*	*0.145*	*1*	*75,537*	*22.1*	*.000*	*< 0.00001*
demo	*age*	*1.362*	*1.362*	*1*	*10,502*	*207.8*	*.020*	*< 0.00001*
demo	*sex*	*0.597*	*0.597*	*1*	*10,526*	*91.1*	*.009*	*< 0.00001*
demo	race/ethnicity	0.386	0.096	4	7055	14.7	.008	< 0.00001
demo	max. parental education	0.003	0.003	1	8925	0.5	.000	0.474
demo	*total parental income*	*0.041*	*0.041*	*1*	*9017*	*6.2*	*.001*	*0.01254*
demo	*parent marital status*	*0.478*	*0.08*	*6*	*9097*	*12.2*	*.008*	*< 0.00001*

Table 6
ANOVA table for accuracy controlling for cognitive factor scores. Sum Sq. = total sum of squares; Mean Sq = mean squares; NumDF = numerator degrees of freedom; DenDF = denominator degrees of freedom; η^2 = partial eta squared; congr. = congruency manipulation; freq. = frequency; r. freq. = relative frequency manipulation; val. = distractor valence manipulation.

Domain	Variable	Sum Sq	Mean Sq	NumDF	DenDF	F-value	η^2	p-value
cognitive	*PC1-general cog. ability*	*2.0167*	*2.0167*	*1*	*8532*	*309.8*	*.035*	*< 0.00001*
cognitive	*PC2- exec. func./speed*	*4.8881*	*4.8881*	*1*	*8852*	*750.8*	*.078*	*< 0.00001*
cognitive	*PC3- learning/memory*	*1.8467*	*1.8467*	*1*	*9131*	*283.6*	*.030*	*< 0.00001*
task	*congruency (congr.)*	*29.4157*	*29.4157*	*1*	*70,413*	*4518.1*	*.060*	*< 0.00001*
task	*relative freq. (r. freq.)*	0.0191	0.0191	1	70,413	2.9	.000	0.087
task	*distractor valence (val.)*	0.0075	0.0075	1	70,413	1.1	.000	0.284
task	*congr. by r. freq.*	*1.6151*	*1.6151*	*1*	*70,413*	*248.1*	*.004*	*< 0.00001*
task	*congr. by val.*	*0.7608*	*0.7608*	*1*	*70,413*	*116.9*	*.002*	*< 0.00001*
task	*r. freq. by val.*	*0.2446*	*0.2446*	*1*	*70,413*	*37.6*	*.001*	*< 0.00001*
task	*congr. by r. freq. by val.*	*0.1217*	*0.1217*	*1*	*70,413*	*18.7*	*.000*	*.00002*
demo	age	0.0195	0.0195	1	9501	3.0	.000	0.083
demo	*sex*	*0.3972*	*0.3972*	*1*	*9800*	*61.0*	*.000*	*< 0.00001*
demo	race/ethnicity	0.1675	0.0419	4	3868	6.4	.007	.00004
demo	max. parental education	0.002	0.002	1	7909	0.3	.000	0.577
demo	total parental income	0.0017	0.0017	1	8496	0.3	.000	0.613
demo	parent marital status	0.0904	0.0151	6	8604	2.3	.002	0.031

3.2. Associations of EWEFS performance with cognitive principal components from one year prior

Repeated measures ANOVAs demonstrated associations between overall RT and overall accuracy at year 1 with the PCs derived from performance at baseline: General cognitive ability (RT: η^2 = .013, p < .00001; Accuracy: η^2 = .035, p < .00001), EF/processing speed (RT: η^2 = .128, p < .00001; Accuracy: η^2 = .078, p < .00001), and Learning/Memory (RT: η^2 = .019, p < .00001; Accuracy: η^2 = .030, p < .00001) (Tables 5 and 6). For all these effects, higher cognitive PCs were associated with faster RT and higher accuracy overall. To test for associations between these cognitive factor scores and Stroop measures of interest, we carried out correlational analyses which revealed small to medium sized zero-order correlations between the three cognitive factor scores and performance on congruent trials (Spearman's rho = 0.193–0.393), negligible to small correlations with the interference effect (Spearman's rho = 0.000–0.09), and negligible relations with the difference in performance for happy as compared to angry faces (Spearman's rho = 0.000–0.040) (for full correlation table, see Supplemental Fig. 2). In light of these findings, we carried out an additional *post hoc* mixed effects model in which the EF/processing speed PC was predicted by the interference effect and congruent performance simultaneously to determine the degree to which our Stroop measures of interest may be differentially associated with this PCs. These analyzes revealed that the distinct Stroop measures were independently significantly associated with EF/processing speed from the year prior, though the effect size for the interference effect was quite small (congruent RT:

β(SE) = $-$0.27(0.01), p < .00001; congruent accuracy: β(SE) = 0.12 (0.01), p < .00001 accuracy interference: β(SE) = 0.04(0.01), p < .00001).

3.3. Associations of EWEFS performance with CBCL symptom dimensions from one year prior

Correlational analyses demonstrated negligible to small correlations of Stroop performance on congruent trials with the total problem (Spearman's rho: congruent accuracy = $-$0.093; congruent RT = 0.055) and externalizing (Spearman's rho: congruent accuracy = $-$0.092; congruent RT = 0.041) subscales, whereas associations with internalizing were negligible (Spearman's rho: congruent accuracy = $-$0.024; congruent RT = 0.009). Furthermore, the non-composite CBCL syndrome scores at baseline also showed associations with performance on congruent trials (Spearman's rho = 0.001–0.135) and the interference effects (Spearman's rho = 0.000–0.067), as well as negligible correlations with the difference in performance for happy as compared to angry faces (Spearman's rho = 0.000–0.031) at year 1 (for full correlation matrix, see Supplemental Fig. 2). Models in which we predicted the Stroop measures of interest by all eight non-composite syndrome scales revealed small Bonferroni-corrected associations of performance on congruent trials with attention problems (accuracy: β(SE) = $-$0.148(0.015); RT: (SE) = 0.117(0.013)), social problems (accuracy: β(SE) = $-$0.070(0.016); RT: β(SE) = 0.073(0.015)), aggressive behaviors (accuracy: β(SE) = $-$0.060(0.017)), and thought problems (RT: β(SE) = $-$0.047(0.014)), as well as associations between the

interference effect and social problems (accuracy: β(SE) = $-$ 0.057 (0.015)) (for full results see Table 7). Results from multiple regression analyses largely aligned with correlational analyses.

4. Discussion

We investigated properties of performance on the EWEFS task from year 1 of the ABCD study, including testing the efficacy of task manipulations and evaluating relations of performance to cognitive PCs and psychopathology dimensions from the baseline timepoint administered one year prior. In the following sections, we first discuss the observed pattern of task manipulation effects. We then discuss analyzes evaluating the degree to which Stroop performance is associated yet dissociable from cognitive principal components at the baseline timepoint one year prior, and conclude with a consideration of observed associations between Stroop performance and measures of psychopathology at baseline.

4.1. Task manipulations of congruency and frequency worked as designed

Our results suggest that this task does indeed evoke the Stroop interference effect and this effect is most pronounced when incongruent trials are relatively infrequent. We also observe that RTs are slightly slower for angry as compared to happy task-irrelevant faces, but this difference was very small. The fact that performance on incongruent but not congruent trials was affected by the relative frequency manipulation suggest two important points. First, the longer RTs and increased errors on incongruent trials in the mostly congruent block compared to the equal block indicates that the high frequency of incongruent trials in the equal block did in fact instantiate a more proactive control state. Second, because performance on congruent trials was unchanged by the relative frequency manipulation, we can rule out the possibility that the enlarged interference effect in the mostly congruent block was due to a facilitation effect on congruent trials and not interference on incongruent trials. While these results were largely replications of preexisting findings, they provide a degree of clarity as to some inconsistencies in the literature while also providing an unprecedented lens into the true effect size of popular emotional Stroop manipulations in this age range. Importantly, we found accuracy to be quite high across the task potentially approaching a ceiling effect, suggesting that future analyses may be best served to focus on RT measures over accuracy.

While we did in fact observe a predicted slowing on angry as compared to happy faces, the effect size was small. As such, caution is warranted in interpreting this effect, and the EWEFS may have limited utility for investigating valence-specific emotional information processing biases among children in this age range. However, given that socioemotional functioning can dynamically change across adolescent

development (Del Piero et al., 2016), we note that effects of valence in the EWEFS may change as the study progresses and over the course of pubertal development. Here, participants were between 10 and 11 years old at the time of the EWEFS task, and thus likely not yet reaching or just beginning puberty. Because changes in socioemotional functioning during adolescence are thought to be catalyzed by puberty (e.g., Goddings et al., 2012), it may be that the ages investigated in the current study are prior to when emotional information processing biases are most pronounced. Relatedly, in the wave of the ABCD study that was the focus of this analysis, prevalence and severity of psychopathology was limited (see Supplemental Table 1 for descriptive statistics). As children undergo adolescent development, increases in psychopathology are likely and may in turn be associated with larger-effect-size associations with EWEFS performance. We plan to investigate these possibilities possibility through future ABCD study data releases.

4.2. Stroop performance is associated with measures of cognition and psychopathology from the year prior

Of particular interest was determining the degree to which Stroop performance is both associated with and dissociable from cognitive measures determined one year prior. Suggesting an overlap between neurocognitive measures and performance on the EWEFS task, scores on all three cognitive PCs were associated with overall performance (i.e., mean RT and accuracy) regardless of condition. These associations were strongest for the PC capturing executive function and processing speed, two constructs the EWEFS task can distinguish through the interference effect and performance on congruent trials, respectively. Indeed, post hoc analyses revealed that RT and accuracy on congruent trials, as well as the interference effect on accuracy, were all independently associated with this PC, however the effect size for accuracy interference was quite small. Thus, we view the EWEFS task as complementary to other cognitive tasks within the ABCD study in trying to disentangle distinct constructs relevant to cognition, more specifically processing speed and inhibitory control. These associations between Stroop performance at year 1 and cognitive ability at baseline may also suggest some degree of stability in individual differences in cognition over child development. This may be particularly true for EF/processing speed, which accounted for roughly 9% and 15% of variance in congruent accuracy and RT, respectively. Indeed, previous research suggests that EFs are stable over time, at least from adolescence to adulthood and almost all this stability is attributable to genetic factors (Friedman et al., 2016). Our results extend this work by suggesting that the stability in individual differences in cognitive ability can be detected across tasks during late childhood, if only across one year. It is worth noting, however, that performance on congruent trials may not actually be indexing individual differences in cognition per se, but could instead be capturing other constructs, such as

Table 7
Results from mixed effects models predicting EWEFS performance by CBCL syndrome scales. RT = reaction time; Prob. = problems; Com. = complaints; Anx./ Dep. = anxious/depressed; SE = standard error; * indicate if effect was significant at Bonferroni-corrected alpha of .001. All betas are standardized.

CBCL subscale	Congruent trials		Interference effect		Happy – angry faces	
	Accuracy β (SE)[Bonf. sig.]	RT β (SE)[Bonf. sig.]	Accuracy β (SE)[Bonf. sig.]	RT β (SE)[Bonf. sig.]	Accuracy β (SE)[Bonf. sig.]	RT β (SE)[Bonf. sig.]
Internalizing/externalizing models						
Internalizing	-0.005(0.013)	.036(0.012)	.007(0.012)	.005(0.012)	.020(0.012)	-0.026(0.012)
Externalizing	-0.107(0.013)*	-0.012(0.012)	-0.067(0.012)*	.013(0.012)	-0.015(0.012)	.005(0.012)
Non-composite syndrome scales models						
Aggressive	-0.060(0.017)*	-0.025(0.015)	-0.011(0.016)	-0.006(0.016)	.028(0.016)	.005(0.016)
Rule Breaking	.046(0.016)	-0.022(0.014)	-0.025(0.014)	.023(0.015)	-0.026(0.015)	.025(0.015)
Attention Prob.	-0.148(0.015)*	.117(0.013)*	-0.043(0.014)	-0.003(0.014)	-0.027(0.014)	-0.042(0.014)
Thought Prob.	.040(0.015)	-0.047(0.014)*	.037(0.014)	-0.010(0.014)	.020(0.014)	.011(0.014)
Social Prob.	-0.070(0.016)*	.073(0.015)*	-0.057(0.015)*	.002(0.015)	-0.016(0.015)	.009(0.015)
Somatic Com.	.011(0.013)	-0.003(0.010)	-0.019(0.011)	.020(0.011)	-0.000(0.011)	-0.009(0.011)
Withdrawn	.043(0.013)	-0.020(0.012)	.023(0.013)	-0.005(0.013)	.027(0.013)	-0.030(0.013)
Anx./Dep.	-.001(0.015)	-0.020(0.013)	.022(0.014)	.000(0.014)	-0.004(0.014)	.006(0.014)

motivation and/or motoric function. Further research is needed to disentangle these possibilities.

Results of the present study suggest that aspects of Stroop performance, including measure of general processing speed/attention and inhibitory control, are preferentially associated with externalizing symptoms, as well as attentional problems from the year prior. Specifically, in multiple regression models including externalizing and internalizing as simultaneous predictors of performance, higher levels of externalizing symptoms were associated with lower accuracy on congruent trials and, to a lesser but notable degree, a larger interference effect, suggesting externalizing-specific impairments in processing speed/attention and inhibitory control. A similar pattern of results has been observed previously in the same age group, albeit on different tasks and using a case-control framework (Brunnekreef, et al., 2007). Taken together, the current study and prior research suggest that externalizing psychopathology during pre-adolescence may be preferentially associated with attentional and difficulties self-regulating/inhibiting behaviors as opposed to symptoms specific to internalizing psychopathology. However, other possibilities exist. For instance, because the current sample is in childhood when the incidence and severity of internalizing psychopathology is relatively low, it may that internalizing-specific impairments in processing speed/attention and inhibitory control are less pronounced at this developmental stage or, as noted in the Section 4.3, parents may be better able to report on their child's externalizing than internalizing behaviors.

Additional analyses using more specific syndrome subscales revealed that Stroop performance was not only associated with externalizing symptomology, but was also associated with attention and, to a lesser but notable degree, social problems. Importantly, these associations were largely unchanged even after controlling for all other syndrome subscales, suggesting that associations between attention and social problems with performance were distinct from the more general associations with externalizing symptomology. It thus appears that Stroop performance may be influenced both by difficulties in self-regulating aggressive behavior and rule following (i.e., externalizing-specific subscales), but also issues with regulating attention and difficulties in processing socioemotional information, information which is central to the EWEFS task due to the use of emotional faces as distractors. Importantly, the current results suggest the associations with social problems may not by valence specific (i.e., stronger for happy or angry faces) but may instead be associated with performance regardless of the valance of the distracting face.

4.3. Limitations and future directions

This study is not without limitations. Because we wanted to utilize the publicly available pre-tabulated data, our analyses focused exclusively on summary values provided in the NDA. However, methodological decisions may influence these summary values, including the decision to treat errors of omission (i.e., no response during trial interval) and errors of commission (i.e., wrong response) the same. While there is no gold standard regarding how to treat these two forms of errors in Stroop paradigms, it is possible that they may arise from distinct mechanisms (Rezaei, 2019) and should thus be treated separately. In the context of the pre-tabulated data, we are not able to investigate this possibility.

Additionally, the CBCL may not be the ideal instrument to characterize psychopathology status and lacks important aspects of a proper clinical interview. Furthermore, the utilization of parent reports on the CBCL may also be an important limitation, with some evidence suggesting that while there is some convergence between parent and youth report, there is considerable disagreement, particularly for internalizing psychopathology (Huang, 2017). However, not only are parent reports frequently used for assessing psychopathology in the current age group, we note that parent reports are more in line with youth self-report measures than other informants (Huang, 2017) and explain an

appreciable amount of variance in youth self-report, particularly in the United States (Rescorla et al., 2013). Further research is needed to validate the parent report CBCL data in the ABCD study, including explorations of its relation to analogous self-report measures.

As the ABCD study progresses and participants move through adolescence, we plan to continue to track performance on the EWEFS task and its relations with other measures of cognition and psychopathology. We are particularly interested in tracking how performance metrics capturing processing speed/attention, inhibitory control, and emotional information processing may change and interact across adolescence and in conjunction with psychopathology, but also how these shifts in performance may coincide with developmental changes in the brain.

CRediT authorship contribution statement

Declaration of Competing Interest

Acknowledgements

Data Statement

All data used in this project were drawn from the publicly available data sets obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). The authors attempt to describe all methods in full detail, including describing the specific variables employed, the rationale behind all exclusions, as well as detailing the specific analyses that were run.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.dcn.2021.101054.

References

Achenbach, T.M., Ruffle, T.M., 2000. The child behavior checklist and related forms for assessing behavioral/emotional problems and competencies. Pediatr. Rev. 21 (8), 265–271.

Acker, W., Acker, C., 1982. Bexley Maudsley Automated Processing Screening and Bexley Maudsley Category Sorting Test Manual. NFER-Nelson, Windsor, England.

Amir, N., Freshman, M., Foa, E., 2002. Enhanced Stroop interference for threat in social phobia. J. Anxiety Disord. 16 (1), 1–9.

Banich, M.T., 2019. The Stroop effect occurs at multiple points along a cascade of control: evidence from cognitive neuroscience approaches. Front. Psychol. 10, 2164.

Bauer, P.J., Dikmen, S.S., Heaton, R.K., Mungas, D., Slotkin, J., Beaumont, J.L., 2013. III. NIH toolbox cognition battery (CB): measuring episodic memory. Monogr. Soc. Res. Child Dev. 78 (4), 34–48.

Beall, P.M., Herbert, A.M., 2008. The face wins: stronger automatic processing of affect in facial expressions than words in a modified Stroop task. Cogn. Emot. 22 (8), 1613–1642.

Becker, E.S., Rinck, M., Margraf, J., Roth, W.T., 2001. The emotional Stroop effect in anxiety disorders: general emotionality or disorder specificity? J. Anxiety Disord. 15 (3), 147–159.

Bjork, J.M., Straub, L.K., Provost, R.G., Neale, M.C., 2017. The ABCD study of neurodevelopment: Identifying neurocircuit targets for prevention and treatment of adolescent substance abuse. Curr. Treat. Options Psychiatry 4 (2), 196–209. https://doi.org/10.1007/s40501-017-0108-y (Epub 2017 Apr 20. PMID: 29038777; PMCID: PMC5639722).

Brunnekreef, A.J., De Sonneville, L.M., Althaus, M., Minderaa, R.B., Oldehinkel, A.J., Verhulst, F.C., Ormel, J., 2007. Information processing profiles of internalizing and externalizing behavior problems: evidence from a population-based sample of preadolescents. J. Child Psychol. Psychiatry 48 (2), 185–193.

Buyukdura, J.S., McClintock, S.M., Croarkin, P.E., 2011. Psychomotor retardation in depression: biological underpinnings, measurement, and treatment. Prog. Neuro-Psychopharmacol. Biol. Psychiatry 35 (2), 395–409.

Carlozzi, N.E., Beaumont, J.L., Tulsky, D.S., Gershon, R.C., 2015. The NIH toolbox pattern comparison processing speed test: normative data. Arch. Clin. Neuropsychol. 30 (5), 359–368.

Daniel, M.H., Wahlstrom, D., Zhang, O., 2014. Equivalence of Q-interactive and paper administrations of cognitive tasks: WISC-V. Q-Interact. Tech. Rep. 8.

Del Piero, L.B., Saxbe, D.E., Margolin, G., 2016. Basic emotion processing and the adolescent brain: task demands, analytic approaches, and trajectories of changes. Dev. Cogn. Neurosci. 19, 174–189.

Epp, A.M., Dobson, K.S., Dozois, D.J., Frewen, P.A., 2012. A systematic meta-analysis of the Stroop task in depression. Clin. Psychol. Rev. 32 (4), 316–328.

Eriksen, B.A., Eriksen, C.W., 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept. Psychophys. 16 (1), 143–149.

Freund, M.C., Bugg, J.M., Braver, T.S., 2020. A representational similarity analysis of cognitive control during color-word Stroop. bioRxiv.

Friedman, N.P., Miyake, A., 2017. Unity and diversity of executive functions: individual differences as a window on cognitive structure. Cortex 86, 186–204.

Friedman, N.P., Miyake, A., Altamirano, L.J., Corley, R.P., Young, S.E., Rhea, S.A., Hewitt, J.K., 2016. Stability and change in executive function abilities from late adolescence to early adulthood: a longitudinal twin study. Dev. Psychol. 52 (2), 326–340.

Gershon, R.C., Cook, K.F., Mungas, D., Manly, J.J., Slotkin, J., Beaumont, J.L., Weintraub, S., 2014. Language measures of the NIH toolbox cognition battery. J. Int. Neuropsychol. Soc. 20 (6), 642–651.

Goddings, A.-L., Burnett Heyes, S., Bird, G., Viner, R.M., Blakemore, S.J., 2012. The relationship between puberty and social emotion processing. Dev. Sci. 15 (6), 801–811.

Goodkind, M., Eickhoff, S.B., Oathes, D.J., Jiang, Y., Chang, A., Jones-Hagata, L.B., Ortega, B.N., Zaiko, Y.V., Roach, E.L., Korgaonkar, M.S., Grieve, S.M., Galatzer-Levy, I., Fox, P.T., Etkin, A., 2015. Identification of a common neurobiological substrate for mental illness. JAMA Psychiatry 72 (4), 305–315. https://doi.org/10.1001/jamapsychiatry.2014.2206 (PMID: 25651064; PMCID: PMC4791058).

Guyer, A.E., Monk, C.S., McClure-Tone, E.B., Nelson, E.E., Roberson-Nay, R., Adler, A.D., Ernst, M., 2008. A developmental examination of amygdala response to facial expressions. J. Cogn. Neurosci. 20 (9), 1565–1582.

Henik, A., Salo, R., 2004. Schizophrenia and the stroop effect. Behav. Cogn. Neurosci. Rev. 3 (1), 42–59.

Huang, C., 2017. Cross-informant agreement on the child behavior checklist for youths: a meta-analysis. Psychol. Rep. 120 (6), 1096–1116.

Hutchison, K.A., 2011. The interactive effects of listwide control, item-based control, and working memory capacity on Stroop performance. J. Exp. Psychol.: Learn. Memory Cogn. 37 (4), 851–860.

Ioannidis, J.P., 2005. Why most published research findings are false. PLoS Med. 2 (8), e124.

Joyal, M., Wensing, T., Levasseur-Moreau, J., Leblond, J., T. Sack, A., Fecteau, S., 2019. Characterizing emotional Stroop interference in posttraumatic stress disorder, major depression and anxiety disorders: a systematic review and meta-analysis. PLoS One 14 (4), e0214998.

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. J. Exp. Psychol.: General 132 (1), 47–70.

Koven, N.S., Heller, W., Banich, M.T., Miller, G.A., 2003. Relationships of distinct affective dimensions to performance on an emotional Stroop task. Cogn. Ther. Res. 27 (6), 671–680.

Lansbergen, M.M., Kenemans, J.L., Van Engeland, H., 2007. Stroop interference and attention-deficit/hyperactivity disorder: a review and meta-analysis. Neuropsychology 21 (2), 251–262.

Logan, G.D., Zbrodoff, N.J., 1979. When it helps to be misled: facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. Memory Cogn. 7 (3), 166–174.

Luciana, M., Bjork, J.M., Nagel, B., Barch, D.M., Gonzalez, R., Nixon, S.J., Banich, M.T., 2018. Adolescent neurocognitive development and impacts of substance use: overview of the Adolescent Brain and Cognitive Development (ABCD) baseline neurocognition battery. Dev. Cogn. Neurosci. 32, 67–79.

MacLeod, C.M., 1991. Half a century of research on the Stroop effect: an integrative review. Psychol. Bull. 109 (2), 163–203.

McTeague, L.M., Goodkind, M.S., Etkin, A., 2016. Transdiagnostic impairment of cognitive control in mental illness. J. Psychiatr. Res. 83, 37–46. https://doi.org/10.1016/j.jpsychires.2016.08.001. Epub 2016 Aug 5. PMID: 27552532; PMCID: PMC5107153.

McTeague, L.M., Huemer, J., Carreon, D.M., Jiang, Y., Eickhoff, S.B., Etkin, A., 2017. Identification of common neural circuit disruptions in cognitive control across psychiatric disorders. Am. J. Psychiatry 174 (7), 676–685. https://doi.org/10.1176/appi.ajp.2017.16040400. Epub 2017 Mar 21. PMID: 28320224; PMCID: PMC5543416.

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., Wager, T.D., 2000. The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. Cogn. Psychol. 41 (1), 49–100.

Platt, B., Waters, A.M., Schulte-Koerne, G., Engelmann, L., Salemink, E., 2017. A review of cognitive biases in youth depression: attention, interpretation and memory. Cogn. Emot. 31 (3), 462–483.

Posner, J., Maia, T.V., Fair, D., Peterson, B.S., Sonuga-Barke, E.J., Nagel, B.J., 2011. The attenuation of dysfunctional emotional processing with stimulant medication: an fMRI study of adolescents with ADHD. Psychiatry Res.: Neuroimag. 193 (3), 151–160.

Powers, A., Casey, B.J., 2015. The adolescent brain and the emergence and peak of psychopathology. J. Infant Child Adolesc. Psychother. 14 (1), 3–15.

Rescorla, L.A., Ginzburg, S., Achenbach, T.M., Ivanova, M.Y., Almqvist, F., Begovac, I., Verhulst, F.C., et al., 2013. Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. J. Clin. Child Adolesc. Psychol. 42 (2), 262–273.

Rezaei, M., 2019. Neuropsychological decomposing Stroop interference into different cognitive monitoring: an exploratory factor analysis. Basic Clin. Neurosci. 10 (5), 475–483.

Richards, A., Richards, L.C., McGeeney, A., 2000. Anxiety-related Stroop interference in adolescents. J. General Psychol. 127 (3), 327–333.

Rock, P.L., Roiser, J.P., Riedel, W.J., Blackwell, A.D., 2014. Cognitive impairment in depression: a systematic review and meta-analysis. Psychol. Med. 44 (10), 2029–2040.

Scarpina, F., Tagini, S., 2017. The stroop color and word test. Front. Psychol. 8, 557.

Schwartz, K., Verhaeghen, P., 2008. ADHD and Stroop interference from age 9 to age 41 years: a meta-analysis of developmental effects. Psychol. Med. 38 (11), 1607–1616.

Snyder, H.R., Miyake, A., Hankin, B.L., 2015. Advancing understanding of executive function impairments and psychopathology: bridging the gap between clinical and cognitive approaches. Front. Psychol. 6, 328.

Steinberg, L., 2010. A dual systems model of adolescent risk-taking. Dev. Psychobiol.: J. Int. Soc. Dev. Psychobiol. 52 (3), 216–224.

Stroop, J.R., 1935. Studies of interference in serial verbal reactions. J. Exp. Psychol. 18 (6), 643–662.

Suslow, T., Hußlack, A., Kersting, A., Bodenschatz, C.M., 2020. Attentional biases to emotional information in clinical depression: a systematic and meta-analytic review of eye tracking findings. J. Affect. Disord. 274, 632–642.

Thompson, W.K., Barch, D.M., Bjork, J.M., Gonzalez, R., Nagel, B.J., Nixon, S.J., Luciana, M., 2019. The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: findings from the ABCD study's baseline neurocognitive battery. Dev. Cogn. Neurosci. 36, 100606.

Tulsky, D.S., Carlozzi, N., Chiaravalloti, N.D., Beaumont, J.L., Kisala, P.A., Mungas, D., Gershon, R., et al., 2014. NIH Toolbox Cognition Battery (NIHTB-CB): list sorting test to measure working memory. J. Int. Neuropsychol. Soc. 20 (6), 599–610.

Volkow, N.D., Koob, G.F., Croyle, R.T., Bianchi, D.W., Gordon, J.A., Koroshetz, W.J., Weiss, S.R., et al., 2018. The conception of the ABCD study: from substance use to a broad NIH collaboration. Dev. Cogn. Neurosci. 32, 4–7.

Williams, J.M.G., Mathews, A., MacLeod, C., 1996. The emotional Stroop task and psychopathology. Psychol. Bull. 120 (1), 3–24.

Zelazo, P.D., Anderson, J.E., Richler, J., Wallner-Allen, K., Beaumont, J.L., Weintraub, S., 2013. II. NIH Toolbox Cognition Battery (CB): measuring executive function and attention. Monogr. Soc. Res. Child Dev 78 (4), 16–33.