Psychoneuroendocrinology (2014) 50, 155—166

Available online at www.sciencedirect.com

ScienceDirect

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

Cortisol responses to a group public speaking task for adolescents: Variations by age, gender, and race

Camelia E. Hostinar[a], Mollie T. McQuillan[b], Heather J. Mirous[b], Kathryn E. Grant[c], Emma K. Adam[a,b,*]

[a] *Institute for Policy Research, Northwestern University, Evanston, IL, United States*
[b] *School of Education and Social Policy, Northwestern University, Evanston, IL, United States*
[c] *Psychology Department, DePaul University, Chicago, IL, United States*

Received 22 April 2014; received in revised form 18 August 2014; accepted 25 August 2014

KEYWORDS
Stress;
Adolescence;
Cortisol;
Reactivity;
HPA axis;
Public speaking

Abstract Laboratory social stress tests involving public speaking challenges are widely used for eliciting an acute stress response in older children, adolescents, and adults. Recently, a group protocol for a social stress test (the Trier Social Stress Test for Groups, TSST-G) was shown to be effective in adults and is dramatically less time-consuming and resource-intensive compared to the single-subject version of the task. The present study sought to test the feasibility and effectiveness of an adapted group public speaking task conducted with a racially diverse, urban sample of U.S. adolescents ($N = 191$; 52.4% female) between the ages of 11 and 18 ($M = 14.4$ years, $SD = 1.93$). Analyses revealed that this Group Public Speaking Task for Adolescents (GPST-A) provoked a significant increase in cortisol production (on average, approximately 60% above baseline) and in self-reported negative affect, while at the same time avoiding excessive stress responses that would raise ethical concerns or provoke substantial participant attrition. Approximately 63.4% of participants exhibited an increase in cortisol levels in response to the task, with 59.2% of the total sample showing a 10% or greater increase from baseline. Results also suggested that groups of five adolescents might be ideal for achieving more uniform cortisol responses across various serial positions for speech delivery. Basal cortisol levels increased with age and participants belonging to U.S. national minorities tended to have either lower basal

* Corresponding author at: Northwestern University, Annenberg Hall, Room 110, 2120 Campus Drive, Evanston, IL 60208-0001, United States. Tel.: +1 847 467 2010.
E-mail address: ek-adam@northwestern.edu (E.K. Adam).

http://dx.doi.org/10.1016/j.psyneuen.2014.08.015
0306-4530/

cortisol or diminished cortisol reactivity compared to non-Hispanic Whites. This protocol facilitates the recruitment of larger sample sizes compared to prior research and may show great utility in answering new questions about adolescent stress reactivity and development.

1. Introduction

Stress responses to social evaluation and threat increase from childhood to adolescence (Van den Bos et al., 2013). This occurs against the backdrop of a more general puberty-related up-regulation in stress and emotional reactivity (Dahl and Gunnar, 2009). In turn, the higher stress load observed among some adolescents is associated with the emergence of clinical symptomatology (Rudolph and Hammen, 1999), delinquent behaviors (Kim et al., 2003), poorer academic functioning (Flook and Fuligni, 2008), substance use (Bonilha et al., 2013) and other deleterious outcomes. Despite normative maturational changes in the activity of stress systems across adolescence, there are also widespread individual differences in how adolescents respond to the same stressors (Gunnar et al., 2009a; Stroud et al., 2009). Thus, understanding the origins and manifestations of individual differences in stress responsivity during adolescence has important ramifications for many domains of functioning.

1.1. Eliciting acute stress in the laboratory

Empirical efforts to better characterize adolescent stress reactivity often rely on laboratory paradigms to experimentally elicit a stress response from individual participants. For instance, the Trier Social Stress Test (TSST, Kirschbaum et al., 1993), the most frequently used laboratory stressor in stress research with adults worldwide, has seen several adaptations for younger participants: the Trier Social Stress Test for Children (TSST-C, Buske-Kirschbaum et al., 1997), the modified TSST (TSST-M, Yim et al., 2010), and the Leiden Public Speaking Task (Leiden-PST, Westenberg et al., 2009). These public speaking paradigms share elements of social-evaluative threat and uncontrollability, a combination that reliably activates stress physiology in adults (Dickerson and Kemeny, 2004) and adolescents (Gunnar et al., 2009b). These laboratory-based individual stress protocols require the participant to prepare and then deliver a speech, sometimes followed by a mental arithmetic task (the Leiden-PST did not include a mental arithmetic task; the TSST-C and TSST-M did).

1.2. Advantages of laboratory-based social stress paradigms for individuals

Laboratory social stress tests are standardized, easily reproducible, and widely used worldwide, allowing comparability of results to a vast body of research. They also reliably activate stress-response systems like the hypothalamic—pituitary—adrenal (HPA) axis in diverse age groups (Dickerson and Kemeny, 2004; Gunnar et al., 2009a), which can be non-invasively assessed through salivary cortisol samples. Laboratory-based social stress tasks not only elicit elevations in cortisol, they also affect other stress markers such as heart rate and self-reported stress, anxiety and negative affect (Hellhammer and Schubert, 2012; Kirschbaum et al., 1993).

These paradigms have shown their utility in answering a broad range of research questions. In children and adolescents, laboratory-based protocols inspired by the TSST have been used to link acute stress reactivity to numerous events and processes, including child maltreatment (Seltzer et al., 2013; Trickett et al., 2014), later memory recall (Rush et al., 2013), allergic asthma symptoms (Buske-Kirschbaum, 2003), rumination and depressive response styles (Stewart et al., 2013), pubertal development (Van den Bos et al., 2013), or body mass index (Francis et al., 2013). The broad range of research problems approached using TSST variants is not surprising given the well-documented associations between individual differences in stress reactivity and socio-emotional, cognitive, or physical development outcomes (Gunnar and Quevedo, 2007).

1.3. Challenges of laboratory social stress paradigms for individuals

Despite their popularity, classic individual laboratory stress protocols are very resource-intensive, requiring a considerable investment of time, staff hours, and research space. As the standard procedure is typically conducted, each subject is scheduled alone for a 1.5-h laboratory session that requires an experimenter, two judges, and two experimental rooms. At a realistic rate of four participants per week, testing a sample similar to the present study (~200 subjects) would require approximately one year, 900 staff hours, and two research spaces reserved for 300 h each. These logistical demands make it difficult to use individualized social stress protocols in studies with large sample sizes (e.g., genetic or epidemiological studies).

Because testing participants in this fashion can last many months, conducting individual social stress tests across the year can add noise due to seasonal variation in HPA activity (Rosmalen et al., 2005). HPA activity also varies between weekdays and weekends (Thorn et al., 2006), potentially prolonging testing if researchers choose to test exclusively during weekdays or weekends.

1.4. Group-based stress protocols

The TSST in a group format (TSST-G) is a recent methodological innovation in adult stress research (Von Dawans et al., 2011). In the classic single-subject task for adults, participants start by resting for 30 min, receive task instructions in a different room, prepare their speech alone for 10 min, deliver the speech for 5 min in front of the judges, then perform a 5-min mental arithmetic task. In the new group format, six participants undergo the stressor at the same time and in the same room for a total duration of 20 min. Each subject delivers a 2-min speech in front of the same two judges, followed by a serial subtraction task lasting 80 s for each subject. The order of the speeches is random and determined beforehand. The TSST-G participants similarly have 10 min to prepare their speech. They are instructed not to interact with each other during this time. During speech delivery, participants are separated by dividers and cannot see each other.

The TSST-G method is six times more efficient compared to the single-subject TSST in terms of testing time, staff hours, and research space usage. To our knowledge, published studies using the standard TSST-G protocol have thus far mostly been conducted with adults from Germany, Switzerland or Norway (Boesch et al., 2014; Buckert et al., 2014, 2012; Häusser et al., 2012; Jacobsen et al., 2014; Klaperski et al., 2014, 2013; Kumsta et al., 2013; Leder et al., 2013; Von Dawans et al., 2011, 2012). Two studies conducted in the United States have used modified versions of the task conducted with groups of 2—3 participants (Childs et al., 2006; Hackman et al., 2013). Only one of these studies has examined adolescents—i.e., a group of 55 African American adolescents aged 15—18 (Hackman et al., 2013). No prior studies have tested multiple group sizes to identify which one is optimal for a group stress protocol.

Compared to prior group stress tasks, the present study expands the sample size, includes broader age ranges and greater ethnic and racial diversity, and also uses a public speaking script that is developmentally tailored for use with children and adolescents (introducing oneself to a hypothetical new classroom), as opposed to the standard mock job interview script used in studies with older participants from the publications mentioned above. Similar to other published protocols (e.g., Leiden-PST), we did not include a backwards mental arithmetic component to the task, in order to have a more brief protocol and also to keep the task as ecologically valid as possible.

The present study sought to test the feasibility and effectiveness of a group protocol for a public speaking task with adolescents (ages 11—18) from diverse racial backgrounds. Given that adolescence is a period of enhanced sensitivity to social evaluation by peers (Somerville, 2013), it is important to empirically evaluate the stress response to a group stress test in adolescents and ensure that the protocol is effective but does not result in excessive stress, which could raise ethical concerns and increase the risk of participant attrition. We tested the effectiveness of our group public speaking task with respect to eliciting increases in cortisol levels and subjective ratings of negative affect, as well as decreases in positive affect. We next examined whether cortisol response trajectories differed by age, gender, and race. Finally, we tested whether group size or speaking order mattered for cortisol stress reactivity.

2. Method

2.1. Participants

A diverse sample of 206 public school students from two middle schools and one high school in a large urban area in the Midwestern United States was recruited. Of these participants, 15 had to be excluded from the present analyses (*N* = 5 reported using corticosteroid-based medications, which confound cortisol assays; *N* = 3 used oral contraceptives and exhibited flat cortisol curves; *N* = 1 had abnormally high cortisol values suggestive of acute infection or other conditions; *N* = 3 were part of an atypical group of only three participants, which was too different from the conditions experienced by most participants, i.e., group sizes of 5—8 individuals; *N* = 1 left the study before the beginning of the TSST-G; *N* = 1 declined to deliver the speech; and, lastly, *N* = 1 did not provide information on race or ethnicity, which was used as a covariate in cortisol analyses).

The final sample (*N* = 191; 52.4% female) consisted of youth between the ages of 11 and 18 (*M* = 14.4 years, *SD* = 1.93). We coded race and ethnicity into a single variable according to recommendations in past research (http://www.cpc.unc.edu/projects/addhealth/data/code/race; Accessed 2/4/2014), and we created the following four categories: African American (32.5%), Hispanic (31.9%), non-Hispanic White (15.7%), and Other (19.9%; included Asian, American Indian or Alaskan Native, Native Hawaiian or Other Pacific Islander, Mixed Race, or Other). Degrees of freedom available in statistical models and small sample sizes for each of the ethnic or racial subgroups in the ''Other'' category did not allow us to examine them separately in the analyses presented below.

2.2. Procedure

The protocols used in this study were approved by the Institutional Review Boards at Northwestern University and DePaul University. All procedures were carried out with the adequate understanding and written consent of the adolescent subjects and at least one parent or legal guardian. Participants were scheduled for a full-day, weekend visit to DePaul University conducted on one of four[1] consecutive Saturdays in October of 2012. Our protocol was inspired by the original TSST-G, but we have made several modifications for use with adolescents, to reduce participant burden, to fit the constraints of the current study design, and to increase the ecological validity of the task. Compared to the original TSST-G, this protocol had the following modifications: (a) we used speech instructions designed for children or adolescents (i.e., introducing oneself to a hypothetical classroom of students, as suggested by Yim et al., 2010); (b) we only used the public speaking component and not the mental arithmetic task in order to decrease participant burden, reduce the length of the protocol, reduce the time lag between saliva samples, and increase the ecological validity of the task (i.e., make the situation more similar to what youth would encounter in daily life); (c) judges wore professional business attire instead of white lab coats to impersonate adults/officials that adolescents typically encounter in school settings; (d) participants were

[1] Pilot testing during a previous Saturday was also conducted. The 33 participants from this session were not included in these analyses due to major protocol changes that occurred between pilot testing and the rest of the study. One participant declined to participate during pilot testing.

seated instead of standing up to eliminate any impact of standing up from the heart rate record (data not included in this report); (e) we varied group size (from 5 to 8 participants instead of equal groups of 6 used by Von Dawans et al., 2011) in order to empirically test whether group size affects cortisol reactivity. However, we retained the critical aspect of the task, which is the social-evaluative threat posed by the judges and the other adolescents in the room (Dickerson and Kemeny, 2004). The Group Public Speaking Task for Adolescents (from here on referred to as GPST-A) occurred during four sessions (with start times of approximately 11:00 AM; 1:00 PM; 3:00 PM and 5:00 PM) over the course of each Saturday, with different groups of participants in each session, and it lasted approximately 1 h and 10 min per session. Participants in each of the four sessions were randomly assigned to one of two rooms resulting in 8 small groups of 5 to 8 individuals each Saturday. The present report focuses on results from the GPST-A sessions and does not include the other psychosocial or physical health assessments conducted throughout the day. In the 50 min prior to the GPST-A, measures of height, weight, blood pressure, and questionnaire reports of demographics and health were taken, along with an iPad-based measure of executive functioning.

The GPST-A session began with a baseline mood survey. Research assistants then demonstrated saliva sampling using the passive drool method. Participants were then instructed to take the first saliva sample and then either completed a form with their contact information (for a random subsample participating in a diary study) or completed a brief positive mood induction (for those not in the diary study[2]). They were introduced to the GPST-A task next. Participants were free to choose whether they wanted to stay in the room and prepare a speech they would later deliver, and all but one agreed to do so (this participant agreed to be in the room and think about the speech, but declined to deliver one). Additionally, a second participant had to leave the study early, before the beginning of the speech task.

Fig. 1 presents the room layout of the GPST-A; note that this room configuration was set up each Saturday in a regular university classroom, using collapsible room dividers to increase the portability of the protocol and applicability outside of a regular laboratory environment. Fig. 2 presents the target timeline template used each Saturday. Exact timings varied slightly from session to session and were, on average, −16.6 min, 0 min, 16.4 min, 29.7 min, 40.9 min, and 52.9 min from the beginning of the GPST-A but are shown on the graph as −15, 0, 15, 30, 40, and 50 min from GPST-A onset for simplicity; actual time of saliva collection was used in HLM models for increased accuracy. Each participant was seated between two dividers in a room, undergoing testing in groups of 5 to 8 participants (there were 10 groups of 5, 6 groups of 6, 3 groups of 7, and 12 groups of 8 participants across the four Saturdays). The dividers prevented eye contact

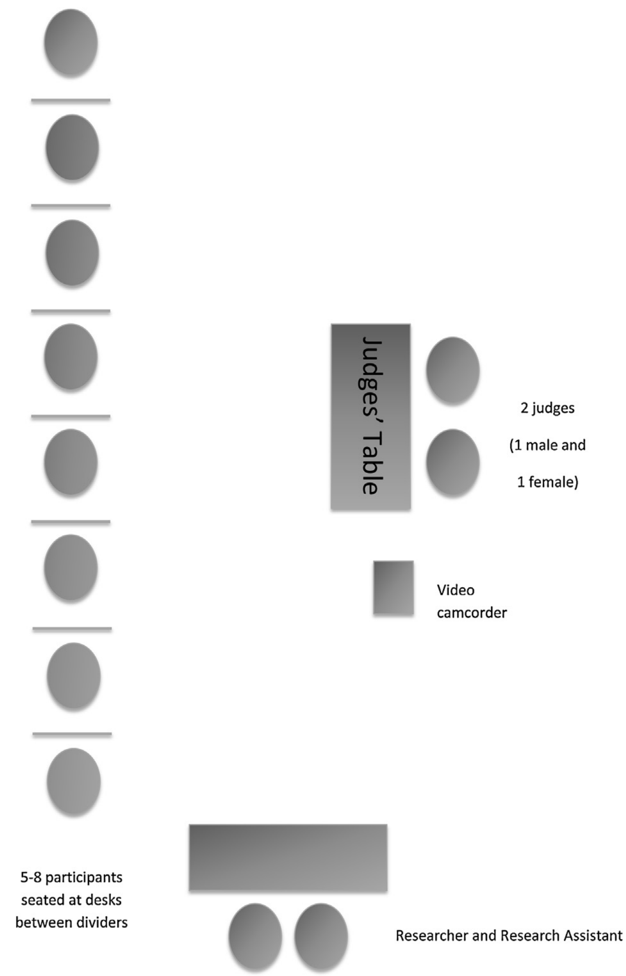

Fig. 1 Layout of the stress exposure room.

and social interaction among participants during the GPST-A. Adolescents were not allowed to use their cell phones during the procedure to prevent distractions or recordings of other participants' speeches, which would represent a major breach of confidentiality. Participants were told to prepare a 1.25-min speech to introduce themselves to a hypothetical classroom of students. Most speeches lasted 1.25 min; however, speeches were longer by up to 15 s in a few instances. Adolescents were asked to discuss both positive and negative aspects of themselves in their introduction (see Appendix 1 for complete instructions; the script used by Yim et al., 2010 for the public speaking portion of the single-subject TSST-M was employed). Participants were told that judges who were trained to evaluate their speech content and body language would enter the room and call them in a random order to begin their speech and may also return to them and ask further questions at any point. Adolescents had 3 min to prepare their speech before the judges entered the room. Participants then provided their second saliva sample at the end of their speech preparation period. The two judges, always one male and one female, entered dressed in professional business attire and sat at the front of the room with a conspicuous video camera that was set up to record the students' speeches. Judges called on students in a random order by their divider number to begin

[2] Each Saturday half the participants were randomly assigned to undergo a brief positive mood induction procedure before the GPST-A, however the manipulation had no significant effects on self-reported affect or cortisol levels ($p > 0.50$), thus this factor was excluded to maximize degrees of freedom for analyses. Results did not change when controlling for this parameter.

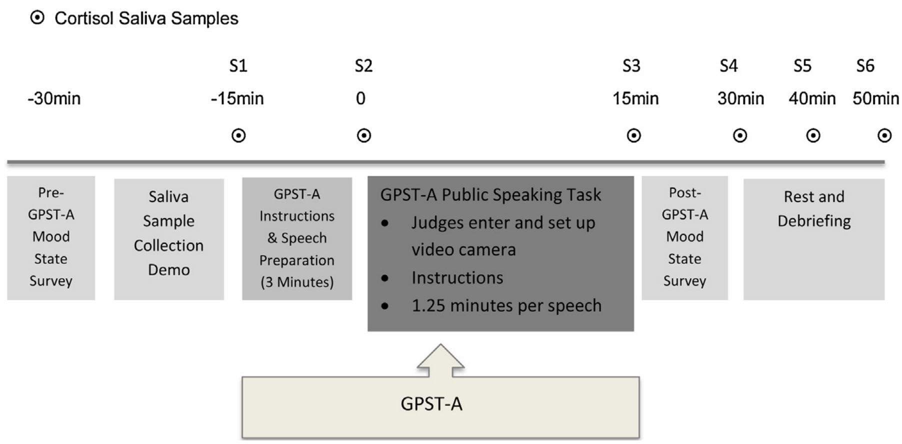

Fig. 2 Session timeline and target salivary cortisol collection times (samples 1—6: S1 through S6).

their speech and retained a neutral expression throughout each speech. Judges also prompted students to continue the task if they ended early or were quiet for at least 20 s (see Appendix 1 for a list of possible prompts). Immediately following the group speech task, the judges left the room and participants provided the third saliva sample. The participants then completed a post-task mood survey, provided the fourth saliva sample, and moved to a different room for debriefing and recovery. During the debriefing and recovery session, all participants were informed about the goals of the study and were reassured about their performance, including being told that their speeches were not actually evaluated for content. Any questions participants had about the procedure were answered, and participants then signed forms acknowledging that they were debriefed. A fifth saliva sample was collected after completion of this debriefing. A randomly selected subset of participants was then provided directions about a home diary study. Lastly, adolescents provided their sixth and final saliva sample before moving to another room for the next set of assessments.

2.3. Measures

2.3.1. Salivary cortisol

Participants expelled saliva through a straw into pre-labeled vials. The samples were stored in a laboratory freezer at −20 °C until being shipped to the University of Trier, Germany for being assayed using a time-resolved fluorescence immunoassay (dissociation-enhanced lanthanide fluorescent immunoassay [DELFIA]; intra-assay CV < 7%, inter-assay CV < 9%). Samples were assayed in duplicate and averaged.

2.3.2. Demographics and health behaviors

In the 50 min prior to the Trier, participants completed a paper-and-pencil questionnaire regarding their demographic information and health behaviors that might influence cortisol levels, including medications taken, caffeine use, number of cigarettes smoked, and typical number of hours of sleep per night. Beyond oral contraceptive use, which was used as a basis for exclusion, none of these health behaviors were significantly associated with either baseline cortisol or the cortisol increase from sample 2 to sample 4 in response to the task, thus their role was not considered any further.

2.3.3. Brief mood survey

Adolescents rated their mood approximately 30 min before and again immediately after the GPST-A. They used 5-point Likert-type items (from 0 = ''Not at all'' to 4 = ''Very much'' to rate how they currently felt on each of 10 affect items: negative (sad, anxious, embarrassed, upset) or positive (happy, self-confident, calm, interested, alert, energetic). Composite measures of Negative Affect and Positive Affect were created by averaging the negative and positive items for the pre- or post-GPST-A ratings. The scales constructed in this fashion had high reliability for both Negative Affect (Cronbach's alphas of 0.71 for pre-GPST-A and 0.68 for post-GPST-A) and Positive Affect (alphas of 0.73 for pre-GPST-A and 0.78 for post-GPST-A).

2.4. Data analysis plan

2.4.1. Statistical analyses

A three-level multilevel growth-curve analysis (Raudenbush and Bryk, 2002; Singer and Willet, 2003) was implemented using the HLM 6.08 software. Use of a multilevel model was necessary given that cortisol samples were nested within persons, who were nested within GPST-A groups, in order to adjust for the non-independence of observations associated with nesting. This also allowed us to model cortisol reactivity using exact time of each sample while also examining person-level and group-level factors predicting differences in the stress response growth curves. Variables that changed with each cortisol sample, such as time of sample collection, were Level 1 or moment-level variables. Three time parameters were included at this level to model the functional shape of the cortisol growth curve identified based on visual inspection and the session timeline. The intercepts were set to represent sample 2, which we are considering our baseline sample, by coding it as time zero, whereas the time of all other samples was coded as either minutes before sample 2 (time term for anticipatory responses captured by a positive slope from sample 1 to sample 2) or as minutes after sample 2 for samples assessing reactivity to the task (samples 3—6). Both linear (time-after-sample-2) and quadratic (time-after-sample-2-squared) time terms were entered to account for the curvilinearity observed in the second (Trier reactivity) portion of the cortisol trajectory. Level 2 included person-level variables (e.g., demographics, exact time of day for each person's speech, and speech order). The only characteristic of GPST-A groups included at Level 3 of the model was group size. All continuous variables were grand-mean centered, with the exception of time-since-sample-2 variables given the coding described above. Race was introduced as three dummy codes which represented African Americans, Hispanics, or Other as ''1'', respectively, whereas non-Hispanic Whites were the reference group and were coded as ''0'' across all three variables. Gender was also dummy-coded, with female as the reference.

The analysis proceeded as follows. First, a basic model with the three time predictors entered at Level 1 and errors nested according to the three levels described above was fit to assess whether there was a significant curvilinear cortisol rise from baseline to peak in response to the task. Paired-samples *t*-tests were then used to investigate changes in negative or positive affect from pre- to post-GPST-A. In the second HLM, time of day for each person's speech start time and demographics (age, gender, and race codes) were entered simultaneously at Level 2 to assess variations in cortisol intercepts, anticipatory responses, or reactivity to the task across these person-level characteristics. Lastly, two separate models were conducted to test the effects of speech order (entered at Level 2) or group size (entered at Level 3) on cortisol reactivity to the task captured by the linear and quadratic time-since-sample-2 terms. Given that group size and speech order were significantly correlated ($r(189) = 0.30$, $p < 0.001$), they were tested separately to avoid multicollinearity. All models included random effects for each term at Level 2, and random intercepts for Level 3; other Level 3 random slopes were dropped due to having near-zero variance. Maximum likelihood estimation with robust standard errors was used in all multilevel models.

3. Results

3.1. Did the protocol elicit stress responses?

We first examined whether the protocol elicited a cortisol increase. There was a significant curvilinear rise in cortisol from baseline to peak for the trajectory after sample 2 (linear term: $\beta = 0.02$, $SE = 0.003$, $p < 0.001$; quadratic term: $\beta = -0.003$, $SE = 0.00004$, $p < 0.001$). As Fig. 3 indicates, maximal values occurred approximately 30 min after the beginning of the GPST-A, which is consistent with the typical time course of peak HPA activation reported in a previous meta-analysis (i.e., 21—30 min from stressor task onset, Dickerson and Kemeny, 2004). The peak cortisol value (i.e., sample 4) was, on average, a 60.3% increase from the baseline captured by sample 2. A substantial proportion of participants (63.4%, $N = 121$) displayed a cortisol increase greater than zero, whereas 36.6% ($N = 70$) had either stable (i.e., no change in) cortisol levels ($N = 4$) or showed a decrease across the session ($N = 66$). Among those with any elevation, cortisol increases ranged from 1.6% to 1430.8% (Med = 76.7% increase, $M = 148.1\%$, SD = 216.1%) from baseline, corresponding to a range of 0.001 to 0.55 μg/dl increases in cortisol concentrations. There are no clear guidelines for determining a definite cortisol response, but some have suggested a 10% increase as a meaningful cut-off (Gordis et al., 2006) and 59.2% ($N = 113$) of our sample exceeded this threshold.

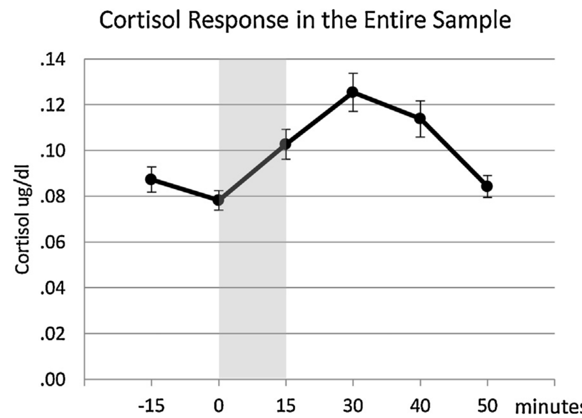

Fig. 3 Observed mean salivary cortisol levels across the session. Error bars are standard errors of the mean (SEMs). The shaded region marks the public speaking portion of the task.

With respect to anticipatory responses, the first (pre-Trier instruction) sample was 11.5% higher than the second one (difference was not significant, $\beta = -0.002$, $SE = 0.003$, $p = 0.36$), suggesting that, on average, there was not an anticipatory stress response (i.e., an increase in cortisol in response to the instructions before the GPST-A began).

When examining changes in self-reported affect, participants also reported perceiving significant increases in negative affect, $t(182) = 5.54$, $p < 0.001$, and decreases in positive affect, $t(182) = -7.01$, $p < 0.001$, from baseline measures (30 min pre-GPST-A) to the period immediately after the GPST-A (see Fig. 4 for means and standard errors). Perceived negative affect increased by an average of 59.7%, but remained lower on average than a rating of 1 (''A little''). Observed scores ranged from 0 to 3 on both the pre- and post-GPST-A negative affect measure, which had a maximum possible value of 4. Positive Affect decreased approximately 16.7%, but post-GPST-A mean levels remained in the same range as pre-GPST-A, between a rating of 2 (''Moderate'') and 3 (''A good amount''). Observed values ranged from 0 to 4 on the pre-GPST-A positive affect measure, and from 0 to 3.67 on post-GPST-A levels.

3.2. Variations by age, gender, and race

We examined the role of demographics in predicting cortisol trajectories next. Age, gender, and race were simultaneously entered at Level 2 of the model, along with time of day (see Table 1 for results). As expected, adolescents had increasing cortisol intercepts — i.e., baseline or sample 2 levels — with increasing age ($\beta = 0.05$, $SE = 0.03$, $p = 0.046$, see Fig. 5 for estimated curves). Older adolescents also had greater anticipatory responses compared to their younger counterparts ($\beta = 0.004$, $SE = 0.001$, $p < 0.001$), as evidenced by the fact that they began having a rise in cortisol from sample 1 to sample 2, i.e. during instructions and speech preparation but before the public speaking

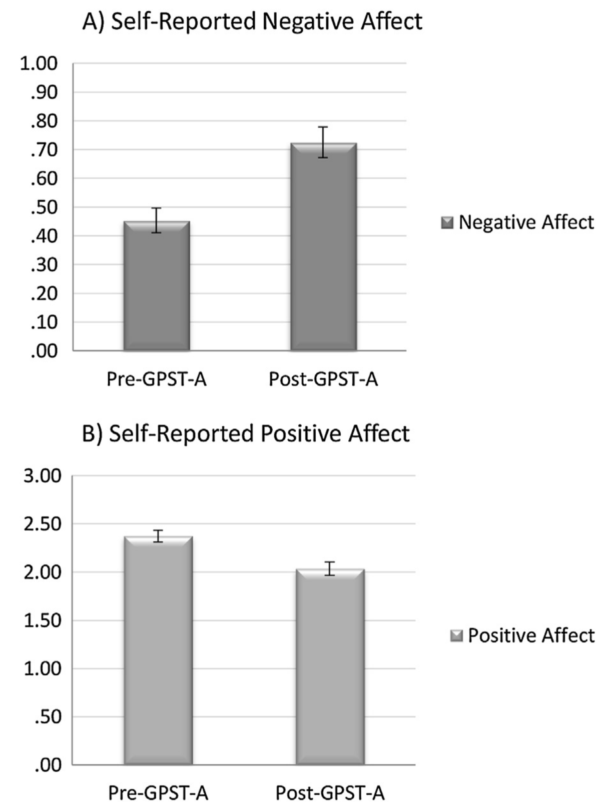

Fig. 4 Mean self-reported mood states before and after the TSST in the entire sample. Mood states were rated on a scale from 0 = ''Not at all'' to 4 = ''Very much'' and averaged to yield summary measures of negative or positive affect. Error bars represent SEMs.

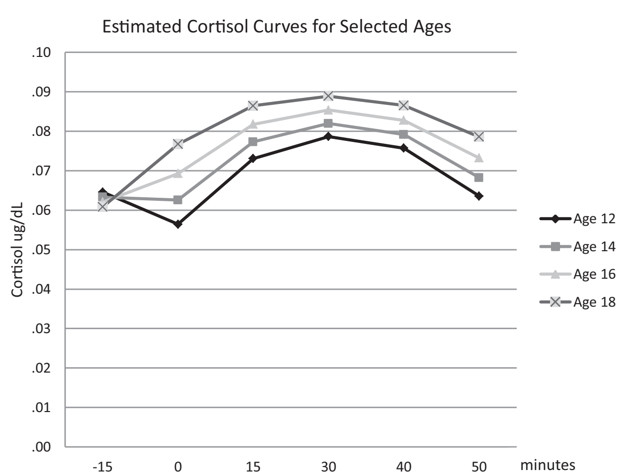

Fig. 5 Estimated cortisol curves based on HLM results. Meaningful ages were selected for plotting instead of ±1 SD from the mean age. The GPST-A began at the 0-min mark. Models revealed significantly increasing intercepts with age and significant age differences in anticipatory slopes (i.e., linear slopes from sample 1 to sample 2), but no differences in reactivity to the task (i.e., curvilinear trajectories from samples 2 through 6) after accounting for intercept differences.

task began. In contrast, the youngest participants showed slight declines from sample 1 to sample 2 (Fig. 5). Males tended to have higher intercepts than females, but this difference was not statistically significant ($\beta = 0.23$, $SE = 0.12$, $p = 0.056$), with no other gender differences in anticipatory responses or task reactivity. African Americans and ''Other'' minorities tended to have lower intercepts compared to non-Hispanic Whites ($\beta = -0.27$, $SE = 0.14$, $p = 0.055$, and $\beta = -0.50$, $SE = 0.17$, $p = 0.004$, respectively). Minorities in the ''Other'' group also had a significantly higher anticipatory response ($\beta = 0.01$, $SE = 0.005$, $p = 0.003$) compared to non-Hispanic Whites. There was a non-significant trend for Hispanics to have lower task reactivity compared to non-Hispanic Whites (linear term: $\beta = -0.01$, $SE = 0.007$, $p = 0.11$; quadratic term: $\beta = 0.0002$, $SE = 0.0001$, $p = 0.079$).

3.3. Variations by GPST-A parameters

We then examined the role of group size and speech order in predicting cortisol reactivity, testing them in separate models given their significant correlation. Group size was entered at Level 3 of the previously-tested model and had small but statistically significant effects on cortisol reactivity to the task (linear term: $\beta = 0.004$, $SE = 0.002$, $p = 0.022$; quadratic term: $\beta = -0.00005$, $SE = 0.00003$, $p = 0.10$). Visual inspection suggested that GPST-A groups of 7 or 8 tended to produce more cortisol than groups of 5 or 6 adolescents. Follow-up HLM analyses comparing groups of 7 or 8 versus 5 or 6 using a dummy code entered at Level 3 instead of group size revealed a trend-level effect for groups of 7 or 8 to have higher reactivity (linear: $\beta = 0.006$, $SE = 0.004$, $p = 0.087$; quadratic: $\beta = -0.00002$, $SE = 0.0003$, $p = 0.39$).

In the next model, speech order was entered at Level 2 since it was an individual-level variable. Speech order had significant effects on cortisol responses to the task (linear term: $\beta = 0.003$, $SE = 0.001$, $p = 0.023$; quadratic term: $\beta = -0.00003$, $SE = 0.00002$, $p = 0.064$). Visual inspection suggested that the 6th, 7th, and 8th serial positions produced slightly more cortisol than the first five positions. Indeed, follow-up analyses using a dummy code instead of the speech order variable showed that the group of participants going 6th, 7th or 8th had greater reactivity compared to everyone else (linear term: $\beta = 0.02$, $SE = 0.006$, $p = 0.009$; quadratic term: $\beta = -0.0002$, $SE = 0.00009$, $p = 0.041$).

4. Discussion

The present study sought to test the feasibility and effectiveness of a group social stress task (public speaking) designed to elicit mild to moderate cortisol reactivity, with the aim of developing a task that was ecologically valid, could be set up outside a typical laboratory setting, and be suitable for use with a diverse sample of U.S. adolescent females and males. Recent reports employing group social stress paradigms have shown their effectiveness in adult samples (Von Dawans et al., 2011, 2012). Given the potential for group testing to increase social-evaluative threat during adolescence due to enhanced sensitivity to social evaluation by peers (Somerville, 2013), we aimed to examine whether this group laboratory stressor would be both effective and mild enough to remain ethical and feasible with adolescents

Table 1 HLM estimates for the first model. Level 1 included three time terms to model variation across samples (slope from sample 1 to sample 2, time-after-sample-2 and time-after-sample-2-squared). The Level 2 model included person-level variables, whereas Level 3 modeled nesting of individuals within GPST-A groups and only included a random intercept.

Fixed Effect	Estimate	SE	*t*	*df*	*p*-Value
Estimates for baseline (S2) cortisol level					
Intercept	−2.626	0.083	−31.495	30	<0.001*
Time of Day	−0.030	0.019	−1.57	184	0.118
Age	0.050	0.025	2.003	184	0.046*
Male	0.226	0.117	1.922	184	0.056△
Hispanic	−0.063	0.119	−0.527	184	0.599
Black	−0.266	0.138	−1.931	184	0.055△
Other	−0.496	0.165	−3.000	184	0.004*
Estimates for slope from S1 to S2					
Intercept	−0.006	0.002	−2.816	184	0.006*
Time of Day	0.001	0.001	0.921	184	0.358
Age	0.004	0.001	3.746	184	<0.001*
Male	−0.002	0.004	−0.514	184	0.608
Hispanic	0.007	0.007	1.098	184	0.274
Black	0.008	0.005	1.652	184	0.1
Other	0.014	0.005	3.105	184	0.003*
Estimates for time-after-S2					
Intercept	0.021	0.006	3.309	184	0.001*
Time of Day	0.0002	0.001	0.159	184	0.874
Age	−0.002	0.001	−1.508	184	0.133
Male	0.000	0.006	0.024	184	0.981
Hispanic	−0.012	0.007	−1.605	184	0.11
Black	−0.009	0.006	−1.385	184	0.168
Other	0.010	0.009	1.129	184	0.261
Estimates for time-after-S2-squared					
Intercept	−0.0003	0.0001	−3.453	184	0.001*
Time of Day	0.00001	0.00002	0.366	184	0.714
Age	0.00003	0.00002	1.507	184	0.133
Male	−0.00003	0.0001	−0.432	184	0.666
Hispanic	0.0002	0.0001	1.762	184	0.079△
Black	0.0001	0.0001	1.098	184	0.274
Other	−0.0001	0.0001	−1.06	184	0.291

* $p < .05$;
△ $p < .10$.

(i.e., not cause excessive distress or lead to massive sample attrition due to refusals; only two participants declined to participate in the GPST-A during either pilot testing or the rest of the study, and one chose to leave the study early for other reasons).

The protocol was successful in significantly increasing cortisol production and eliciting significant changes in negative and positive affect in the expected direction. The average for the peak cortisol sample was approximately 60% over baseline levels and 63.4% of participants exhibited a nonzero increase in cortisol, with 59.2% of the total sample having at least a 10% increase in concentrations from baseline (comparable to the rate of response in studies using single-subject TSST paradigms with adolescents, e.g. Gordis et al., 2006). The size of the average increase in cortisol is also within the range noted in single-subject TSST studies conducted with adolescents (reviewed in Gunnar et al., 2009a). The latter review suggested that studies with 7—12-year-olds sometimes failed to produce a positive HPA response, whereas the vast majority of TSST studies with adolescents aged 13 and older successfully elicited a 9% to 100% average increase in cortisol (Gunnar et al., 2009a). The current protocol was stronger than the Leiden-PST (18% increase for responses to the task, Westenberg et al., 2009), but milder than the TSST-C (90% increase, Buske-Kirschbaum, 2003) and the TSST-M (250% increase, Yim et al., 2010). Responses were also milder than those in the adult TSST-G (Von Dawans et al., 2011), but this is to be expected given typical increases in reactivity with age and previous observations that adults show stronger reactivity to TSST-like tasks compared to children/adolescents, especially in males (Yim et al., 2010). Mild to moderate stressors are often ideal compared to more potent ones because they allow individual differences in stress reactivity to be expressed, while also avoiding ceiling effects in measurement. There are no clear cut-offs for what constitutes a meaningful increase in cortisol in response to a stressor. What we know is that cortisol declines slowly across the day, thus stable or decreasing levels are fairly clear indices of failures to mount an HPA response. Even though the percentage of participants showing a cortisol decrease is comparable to that in other single-subject TSST studies with adolescents, it is unclear what the normative response rate should be in different samples, given the multiplicity of factors that might lead to either an appraisal of the situation as non-stressful or to a physiologically blunted pattern of reactivity (e.g., personality characteristics, experiential factors such as chronic stress, habituation, Kudielka et al., 2009). Future studies should explore predictors of responding to these social stress tests with a cortisol increase to clarify this issue.

Subjective ratings of positive and negative affect also changed significantly in the expected direction, with effects being fairly mild (approximately a 60% increase in negative affect, resulting in a final rating that was still below ''A little'', and a 17% decrease in positive affect, corresponding to ''A good amount'' of positive affect post-GPST-A). The range of scores on positive and negative affect was similar between pre-test and post-test, with negative affect scores not exceeding 3 on a self-report scale that ranges from 0 to 4. Together, these results alleviated concerns that group testing in the same room with similarly-aged participants might have rendered this task too upsetting or stressful due to adolescents' enhanced sensitivity to social evaluation by peers (Somerville, 2013). Two participants refused the task (one during pilot testing and one during the study) and another adolescent asked to leave the study early, but all were free to opt out, suggesting that this protocol does not cause massive attrition. In addition to its feasibility and effectiveness, the task has high ecological validity and is developmentally-appropriate, given that introducing oneself and speaking publicly in front of one or two adults and several similarly-aged peers are typical scenarios in school contexts.

With respect to variations by age, gender, and race, the present study primarily revealed effects on intercepts (i.e., baseline levels) and anticipatory responses, with no statistically significant differences in reactivity to the task itself. These findings suggest that the GPST-A may be equally well-suited for use across ages 11—18, with both genders, and diverse racial groups. Nevertheless, future studies should replicate these results with larger sample sizes to increase the power to detect what may be smaller effects than could be observed here. In the same vein, future studies should tease apart differences between ethnic or racial groups subsumed under the construct of ''Other'', which was not permitted by degrees of freedom in the current analysis. Variations in cortisol levels that were observed are consistent with prior literature. For instance, there was a significant increase in cortisol intercepts and in anticipatory responses with age, both of which are consistent with prior reports on the effects of puberty on HPA activity (Adam, 2006; Van den Bos et al., 2013). Future studies should examine the effects of pubertal maturation stage and timing of puberty, to understand the biological and psychological origins of these age-related changes in patterns of HPA activity.

Prior literature concerning gender differences in adolescent cortisol reactivity is mixed, ranging from studies showing no differences (Gunnar et al., 2009c; Stroud et al., 2009; Sumter et al., 2010) to some finding increased reactivity to performance stressors in adolescent males (Bouma et al., 2009; Zijlmans et al., 2013). The present study found marginally higher intercepts in males and no gender differences in reactivity. However, it must be noted that three female participants who reported using oral contraceptives exhibited flat cortisol slopes, consistent with prior work (Bouma et al., 2009), leading us to exclude them before conducting any analyses. We recommend that future studies should collect this information from female participants.

There was also a general pattern of dampened cortisol levels in participants belonging to U.S. minorities compared to non-Hispanic Whites. There was a trend for African Americans to have lower intercepts and ''Other'' racial minorities had significantly lower intercepts, whereas Hispanic youth had a trend-level reduction in cortisol reactivity compared to non-Hispanic Whites. Previous studies have noted that African American and Hispanic youth have flatter diurnal cortisol rhythms compared to Caucasians, with lower levels than would be expected early in the day and higher bedtime levels than would be normal (DeSantis et al., 2007). This flattened diurnal profile has been associated with chronic stress in meta-analytic results (Miller et al., 2007) and is thought to be, at least in part, due to a down-regulation of the HPA axis after repeated acute stressors (Gunnar and Vazquez, 2001). Future studies should examine experiential and bio-behavioral predictors that may explain the racial differences observed here.

Lastly, the present study tested the effects of GPST-A group size and speech order, revealing that both seemed to have some small effects on cortisol reactivity. The pattern of results suggested that participants who were 6th, 7th or 8th in the public speaking sequence tended to have greater reactivity than the rest of the participants, which also explained the small but significant effect of group size on reactivity, since larger groups would include more participants in those serial positions. Given that judges invited participants to begin their speech in a random order that was not known by participants, it seems that adolescents who had to wait until the end to deliver their speeches tended to have a slightly greater cortisol buildup. These results recommend that researchers should either statistically control for the effect of speech order in order to parse out its contribution, or use TSST-G groups of 5 adolescents if the goal is to achieve a more uniform response. The adult version of the TSST-G had reported no significant effects of serial position on cortisol reactivity in groups of 6 (Von Dawans et al., 2011), supporting the recommendation that smaller groups of 5 or 6 and using the same group size across all participants would be ideal in future studies.

In this sample, we did not find any evidence of main effects of time of day on reactivity, though initial cortisol levels were lower later in the day, as expected. Given that meta-analyses reviewing single-subject social stress tests have found greater effect sizes for cortisol reactivity in afternoon studies compared to morning ones (due to the fact that basal afternoon levels are lower and more stable than morning levels, making it easier to detect increases, Dickerson and Kemeny, 2004), it would be important to replicate our findings before concluding that there is no effect of time of day on reactivity when the group procedure is used. Certainly, conducting sessions in more narrow time bands than was done here would reduce any possible variability due to time of day. Although including sessions at varying times of day could be considered a limitation of the current study, were additional research to confirm that time of day has little to no impact on level of reactivity when the group protocol is utilized, this would further increase the time savings associated with the GPST-A.

In addition to variability due to group size and testing across the day, another limitation of the present study was that the perceived stress ratings were only measured 30 min pre-GPST-A and immediately after it, thus we were unable to examine anticipatory responses or patterns of recovery using these self-report measures.

4.1. Recommendations for future research

Despite significantly eliciting cortisol reactivity and mild increases in negative affect, the group protocol used here was a milder stressor than some of the single-subject procedures designed for use with children and adolescents (e.g., the TSST-M; Yim et al., 2010) and the adult version of the TSST-G (Von Dawans et al., 2011). We highlighted the advantages of a milder stressor above (e.g., avoiding teenage attrition, allowing individual differences in reactivity to manifest). It is likely that the modifications we made to the standardized TSST-G might have led to the lower reactivity observed in the current study. For instance, we only used the public speaking portion of the task and not the mental arithmetic. Future research wishing to elicit greater reactivity could use both challenge tasks. Employing a smaller number of participants in each group (five are recommended) would help to allow time for both tasks and thus potentially increase reactivity by prolonging the stressor. Conducting all sessions in the afternoon may also heighten reactivity, as previously discussed, even though we did not find evidence of this in our study. Instructing participants to stand up for their speech, as is standard for TSST procedures, may be another change to the current protocol that could amplify reactivity. Keeping both group size and stressor duration fixed for each participant would reduce variability in the data and improve power to detect signal from noise.

We chose to conduct our study in a classroom setting on a University campus (resembling school contexts); we consider this an advantage of the current procedure, as it demonstrates that public speaking tasks can be portable—-that is, they can be successfully conducted (albeit with slightly lower reactivity) outside of typical laboratory settings. However, investigators wishing to maximize reactivity may wish to explore other settings for this group-based protocol (e.g., a neutral, sparsely-furnished laboratory room). Importantly, regardless of context, we recommend that the best use of this methodology with this age group should prohibit cell phone usage (as was done here) during all the procedures to prevent distractions and the potential recordings of other participants' speeches, as this would infringe on their privacy.

In sum, the present study revealed that the GPST-A employed here was successful in eliciting significant cortisol reactivity, increasing self-reported negative affect and mildly decreasing positive affect amongst U.S. adolescents, while at the same time avoiding excessive stress responses that would raise ethical concerns or provoke substantial participant attrition. The group protocol described here has good ecological validity, is more portable and dramatically less time-consuming and less resource-intensive than single-subject social stress paradigms, facilitating the recruitment of much larger sample sizes and applicability to a broader range of populations and settings than have been employed in prior research. This methodology adds a new option to the strong existing set of social stress paradigms designed to elicit cortisol reactivity. Due to its efficiency and its applicability to a broader range of research settings and across diverse samples of adolescents of various ages, the GPST-A has the potential for answering new questions about adolescent stress reactivity and development.

Conflict of interest statement

Authors have no financial or other conflicts of interest related to the submitted manuscript.

Contributors

Emma K Adam and Kathryn E Grant designed the study and recruited the sample. Mollie T McQuillan, Heather J Mirous, and Emma K Adam conducted the experiments and collected the data. Camelia E Hostinar and Emma K Adam analyzed the data and interpreted the results. Camelia E Hostinar wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

Role of the funding source

This work was funded through the National Institutes of Health (NIH Grant 5R21AA21073-212 to Emma K Adam and Kathryn E Grant). The funding agency had no role in the collection, analysis, and interpretation of data, in the writing of the report, or in the decision to submit the paper for publication.

Acknowledgements

We would like to thank the participants for their time and effort. This project was funded by NIH grant 5R21AA21073-212 to Drs. Emma K. Adam and Kathryn E. Grant. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix A. Instructions for participants and judges

Speech script from the single-subject TSST-M by Yim et al. (2010) was used.

Instructions provided to participants by the experimenter: Now, we are going to start the performance portion of the study. This section will be video recorded, but you will have the choice afterward of either allowing us or not allowing us to keep the recording.

You will be asked to deliver a speech. Imagine that you are in a new class with about 20 other students and the teacher requested that you make a presentation introducing yourself. You need to talk about your personality and why you would be liked by other students in the class. You must mention at least one good thing and one bad thing about yourself. Please refer to yourself by the number we will provide you with, rather than your name. You will deliver your speech in front of two judges. The judges have special training in observing behavior. During your speech, they will listen, take notes, and analyze your behavior and responses.

Instructions for Judges: Your job as a judge is to present a serious and stoic, but not hostile, demeanor towards the participant. This means that you should **not** smile, nod your head, or provide affirmative verbal responses, such as ''mm-hmm.'' You should keep frequent eye contact with the participant and pretend to be evaluating their speech.

After entering the room where the participants will be, be seated and provide the following instructions to them.

Judge 1: You should speak for the entire time period, until you are told to stop. You will be called in random order to give your speech using the numbers on your divider. When we call your number, please stand up and begin your speech. Even after you are finished with your speech, you may be called on at any time to elaborate further or answer questions. When you are not speaking, please sit quietly. Only stand when it is your turn to speak.

Judge 2: Call the randomly selected first number card and begin timer.

If participant stops talking:

The first time the participant stops talking, wait 10 seconds and then say:

''You should talk for the whole time. Please continue.''

After that, continue to prompt them after each 20 seconds of silence by saying things like:

''You still have time remaining. Please continue.''

''Remember to talk about at least one good thing and one bad thing about yourself.''

''You're supposed to talk for the entire time.''

''You need to keep talking.''

References

Adam, E.K., 2006. Transactions among adolescent trait and state emotion and diurnal and momentary cortisol activity in naturalistic settings. Psychoneuroendocrinology 31, 664—679, http://dx.doi.org/10.1016/j.psyneuen.2006.01.010.

Boesch, M., Sefidan, S., Ehlert, U., Annen, H., Wyss, T., Steptoe, A., La Marca, R., 2014. Mood and autonomic responses to repeated exposure to the trier social stress test for groups (TSST-G). Psychoneuroendocrinology 43, 41—51, http://dx.doi.org/10.1016/j.psyneuen.2014.02.003.

Bonilha, A.G., de Souza, E.S.T., Sicchieri, M.P., Achcar, J.A., Crippa, J.A.S., Baddini-Martinez, J., 2013. A motivational profile for smoking among adolescents. J. Addict. Med. 7, 439—446, http://dx.doi.org/10.1097/01.ADM.0000434987.76599.c0.

Bouma, E.M.C., Riese, H., Ormel, J., Verhulst, F.C., Oldehinkel, A.J., 2009. Adolescents' cortisol responses to awakening and social stress; effects of gender, menstrual phase and oral contraceptives. The TRAILS study. Psychoneuroendocrinology 34, 884—893, http://dx.doi.org/10.1016/j.psyneuen.2009.01.003.

Buckert, M., Kudielka, B.M., Reuter, M., Fiebach, C.J., 2012. The COMT Val158Met polymorphism modulates working memory performance under acute stress. Psychoneuroendocrinology 37, 1810—1821, http://dx.doi.org/10.1016/j.psyneuen.2012.03.014.

Buckert, M., Schwieren, C., Kudielka, B.M., Fiebach, C.J., 2014. Acute stress affects risk taking but not ambiguity aversion. Front. Neurosci. 8, 82, http://dx.doi.org/10.3389/fnins.2014.00082.

Buske-Kirschbaum, A., Jobst, S., Wustmans, A., Kirschbaum, C., Rauh, W., Hellhammer, D., 1997. Attenuated free cortisol response to psychosocial stress in children with atopic dermatitis. Psychosom. Med. 59, 419—426.

Buske-Kirschbaum, A., 2003. Blunted cortisol responses to psychosocial stress in asthmatic children: a general feature of atopic disease? Psychosom. Med. 65, 806—810, http://dx.doi.org/10.1097/01.PSY.0000095916.25975.4F.

Childs, E., Vicini, L.M., De Wit, H., 2006. Responses to the trier social stress test (TSST) in single versus grouped participants. Psychophysiology 43, 366—371, http://dx.doi.org/10.1111/j.1469-8986.2006.00414.x.

Dahl, R.E., Gunnar, M.R., 2009. Heightened stress responsiveness and emotional reactivity during pubertal maturation: implications for psychopathology. Dev. Psychopathol. 21, 1—6, http://dx.doi.org/10.1017/S0954579409000017.

DeSantis, A.S., Adam, E.K., Doane, L.D., Mineka, S., Zinbarg, R.E., Craske, M.G., 2007. Racial/ethnic differences in cortisol diurnal rhythms in a community sample of adolescents. J. Adolesc. Health 41, 3—13, http://dx.doi.org/10.1016/j.jadohealth.2007.03.006.

Dickerson, S.S., Kemeny, M.E., 2004. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol. Bull. 130, 355—391, http://dx.doi.org/10.1037/0033-2909.130.3.355.

Flook, L., Fuligni, A.J., 2008. Family and school spillover in adolescents' daily lives. Child Dev. 79, 776—787, http://dx.doi.org/10.1111/j.1467-8624.2008.01157.x.

Francis, L.A., Granger, D.A., Susman, E.J., 2013. Adrenocortical regulation, eating in the absence of hunger and BMI in young children. Appetite 64, 32—38, http://dx.doi.org/10.1016/j.appet.2012.11.008.

Gordis, E.B., Granger, D.A., Susman, E.J., Trickett, P.K., 2006. Asymmetry between salivary cortisol and alpha-amylase reactivity to stress: relation to aggressive behavior in adolescents. Psychoneuroendocrinology 31, 976—987, http://dx.doi.org/10.1016/j.psyneuen.2006.05.010.

Gunnar, M., Quevedo, K., 2007. The neurobiology of stress and development. Annu. Rev. Psychol. 58, 145—173, http://dx.doi.org/10.1146/annurev.psych.58.110405.085605.

Gunnar, M., Vazquez, D.M., 2001. Low cortisol and a flattening of expected daytime rhythm: potential indices of risk in human development. Dev. Psychopathol. 13, 515—538.

Gunnar, M.R., Frenn, K., Wewerka, S.S., Van Ryzin, M.J., 2009a. Moderate versus severe early life stress: associations with stress reactivity and regulation in 10—12-year-old children. Psychoneuroendocrinology 34, 62—75, http://dx.doi.org/10.1016/j.psyneuen.2008.08.013.

Gunnar, M.R., Talge, N.M., Herrera, A., 2009b. Stressor paradigms in developmental studies: what does and does not work to produce mean increases in salivary cortisol. Psychoneuroendocrinology 34, 953—967, http://dx.doi.org/10.1016/j.psyneuen.2009.02.010.

Gunnar, M.R., Wewerka, S., Frenn, K., Long, J.D., Griggs, C., 2009c. Developmental changes in hypothalamus—pituitary—adrenal activity over the transition to adolescence: normative changes and associations with puberty. Dev. Psychopathol. 21, 69—85, http://dx.doi.org/10.1017/S0954579409000054.

Hackman, D.A., Betancourt, L.M., Brodsky, N.L., Kobrin, L., Hurt, H., Farah, M.J., 2013. Selective impact of early parental responsivity on adolescent stress reactivity. PLoS One 8, e58250, http://dx.doi.org/10.1371/journal.pone.0058250.

Häusser, J.A., Kattenstroth, M., van Dick, R., Mojzisch, A., 2012. We are not stressed: social identity in groups buffers neuroendocrine stress reactions. J. Exp. Soc. Psychol. 48, 973—977, http://dx.doi.org/10.1016/j.jesp.2012.02.020.

Hellhammer, J., Schubert, M., 2012. The physiological response to trier social stress test relates to subjective measures of stress during but not before or after the test. Psychoneuroendocrinology 37, 119—124, http://dx.doi.org/10.1016/j.psyneuen.2011.05.012.

Jacobsen, H.B., Bjørngaard, J.H., Hara, K.W., Borchgrevink, P.C., Woodhouse, A., Landrø, N.I., Harris, A., Stiles, T.C., 2014. The role of stress in absenteeism: cortisol responsiveness among patients on long-term sick leave. PLoS One 9, e96048, http://dx.doi.org/10.1371/journal.pone.0096048.

Kim, K.J., Conger, R.D., Elder, G.H., Lorenz, F.O., 2003. Reciprocal influences between stressful life events and adolescent internalizing and externalizing problems. Child Dev. 74, 127—143.

Kirschbaum, C., Pirke, K.-M., Hellhammer, D.H., 1993. The ''Trier Social Stress Test''—a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 28 (1—2), 76—81.

Klaperski, S., von Dawans, B., Heinrichs, M., Fuchs, R., 2013. Does the level of physical exercise affect physiological and psychological responses to psychosocial stress in women? Psychol. Sport Exercise 14, 266—274, http://dx.doi.org/10.1016/j.psychsport.2012.11.003.

Klaperski, S., von Dawans, B., Heinrichs, M., Fuchs, R., 2014. Effects of a 12-week endurance training program on the physiological response to psychosocial stress in men: a randomized controlled trial. J. Behav. Med., http://dx.doi.org/10.1007/s10865-014-9562-9.

Kudielka, B.M., Hellhammer, D.H., Wüst, S., 2009. Why do we respond so differently? Reviewing determinants of human salivary cortisol responses to challenge. Psychoneuroendocrinology 34, 2—18, http://dx.doi.org/10.1016/j.psyneuen.2008.10.004.

Kumsta, R., Chen, F.S., Pape, H.-C., Heinrichs, M., 2013. Neuropeptide S receptor gene is associated with cortisol responses to social stress in humans. Biol. Psychol. 93, 304—307, http://dx.doi.org/10.1016/j.biopsycho.2013.02.018.

Leder, J., Häusser, J.A., Mojzisch, A., 2013. Stress and strategic decision-making in the beauty contest game. Psychoneuroendocrinology 38, 1503—1511, http://dx.doi.org/10.1016/j.psyneuen.2012.12.016.

Miller, G.E., Chen, E., Zhou, E.S., 2007. If it goes up, must it come down? Chronic stress and the hypothalamic—pituitary—adrenocortical axis in humans. Psychol. Bull. 133, 25—45, http://dx.doi.org/10.1037/0033-2909.133.1.25.

Raudenbush, S.W., Bryk, A.S., 2002. Hierarchical Linear Models: Applications and Data Analysis Methods, second ed. Sage, Thousand Oaks, CA.

Rosmalen, J.G.M., Oldehinkel, a.J., Ormel, J., de Winter, A.F., Buitelaar, J.K., Verhulst, F.C., 2005. Determinants of salivary cortisol levels in 10—12 year old children; a population-based study of individual differences. Psychoneuroendocrinology 30, 483—495, http://dx.doi.org/10.1016/j.psyneuen.2004.12.007.

Rudolph, K.D., Hammen, C., 1999. Age and gender as determinants of stress exposure, generation, and reactions in youngsters: a transactional perspective. Child Dev. 70, 660—677.

Rush, E.B., Quas, J.A., Yim, I.S., Nikolayev, M., Clark, S.E., Larson, R.P., 2013. Stress, interviewer support, and children's eyewitness identification accuracy. Child Dev. 00, 1—14, http://dx.doi.org/10.1111/cdev.12177.

Seltzer, L.J., Ziegler, T., Connolly, M.J., Prososki, A.R., Pollak, S.D., 2013. Stress-induced elevation of oxytocin in maltreated children: evolution, neurodevelopment, and social behavior. Child Dev. 85, 501—512, http://dx.doi.org/10.1111/cdev.12136.

Singer, J.D., Willet, J.B., 2003. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press, Oxford.

Somerville, L.H., 2013. The teenage brain: sensitivity to social evaluation. Curr. Dir. Psychol. Sci. 22, 121—127, http://dx.doi.org/10.1177/0963721413476512.

Stewart, J.G., Mazurka, R., Bond, L., Wynne-Edwards, K.E., Harkness, K.L., 2013. Rumination and impaired cortisol recovery following a social stressor in adolescent depression. J. Abnorm. Child Psychol. 41, 1015—1026, http://dx.doi.org/10.1007/s10802-013-9740-1.

Stroud, L.R., Foster, E., Papandonatos, G.D., Handwerger, K., Granger, D.a, Kivlighan, K.T., Niaura, R., 2009. Stress response and the adolescent transition: performance versus peer rejection stressors. Dev. Psychopathol. 21, 47—68, http://dx.doi.org/10.1017/S0954579409000042.

Sumter, S.R., Bokhorst, C.L., Miers, A.C., Van Pelt, J., Westenberg, P.M., 2010. Age and puberty differences in stress responses during a public speaking task: do adolescents grow more sensitive to social evaluation? Psychoneuroendocrinology 35, 1510—1516, http://dx.doi.org/10.1016/j.psyneuen.2010.05.004.

Thorn, L., Hucklebridge, F., Evans, P., Clow, A., 2006. Suspected non-adherence and weekend versus week day differences in the awakening cortisol response. Psychoneuroendocrinology 31, 1009—1018, http://dx.doi.org/10.1016/j.psyneuen.2006.05.012.

Trickett, P.K., Gordis, E., Peckins, M.K., Susman, E.J., 2014. Stress reactivity in maltreated and comparison male and female young adolescents. Child Maltreat. 19, 27—37, http://dx.doi.org/10.1177/1077559513520466.

Van den Bos, E., de Rooij, M., Miers, A.C., Bokhorst, C.L., Westenberg, P.M., 2013. Adolescents' increasing stress response to social evaluation: pubertal effects on cortisol and alpha-amylase during public speaking. Child Dev. 00, 1—17, http://dx.doi.org/10.1111/cdev.12118.

Von Dawans, B., Fischbacher, U., Kirschbaum, C., Fehr, E., Heinrichs, M., 2012. The social dimension of stress reactivity: acute stress increases prosocial behavior in humans. Psychol. Sci. 23, 651—660, http://dx.doi.org/10.1177/0956797611431576.

Von Dawans, B., Kirschbaum, C., Heinrichs, M., 2011. The trier social stress test for groups (TSST-G): a new research tool for controlled simultaneous social stress exposure in a group format. Psychoneuroendocrinology 36, 514—522, http://dx.doi.org/10.1016/j.psyneuen.2010.08.004.

Westenberg, P.M., Bokhorst, C.L., Miers, A.C., Sumter, S.R., Kallen, V.L., van Pelt, J., Blöte, A.W., 2009. A prepared speech in front of a pre-recorded audience: subjective, physiological, and neuroendocrine responses to the Leiden public speaking task. Biol. Psychol. 82, 116—124, http://dx.doi.org/10.1016/j.biopsycho.2009.06.005.

Yim, I.S., Quas, J.A., Cahill, L., Hayakawa, C.M., 2010. Children's and adults' salivary cortisol responses to an identical psychosocial laboratory stressor. Psychoneuroendocrinology 35, 241—248, http://dx.doi.org/10.1016/j.psyneuen.2009.06.014.

Zijlmans, M.A.C., Beijers, R., Mack, S., Pruessner, J.C., de Weerth, C., 2013. Cortisol responses to social evaluation in 10- to 15-year-old boys and girls. Stress 1—10, http://dx.doi.org/10.3109/10253890.2013.764494.