Acta Psychologica 224 (2022) 103528

Contents lists available at ScienceDirect

Acta Psychologica

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

Beach balls: Assessing frustration tolerance in young children using a computerized task

Annabel Jiménez-Soto *, Joaquín Lorente-Loza, Juan P. Vargas, Estrella Díaz, Juan C. López

Animal Behavior and Neuroscience Lab, Department of Experimental Psychology, Universidad de Sevilla, Spain

ARTICLE INFO

Keywords:
Frustration tolerance
Childhood
Visual perception
Visual discrimination
Reaction time
Computerized task

ABSTRACT

Frustration tolerance is a skill related to emotional regulation processes and is important insofar as it affects people's social relationships and even health. Low levels of frustration tolerance in children have been associated with a greater number of externalizing symptoms such as aggression or anger. Despite its importance, there are a limited number of tasks that attempt to evaluate this construct objectively. Therefore, the aim of our work was the design of a computerized task, programmed as a videogame in order to assess frustration tolerance in children from 6 to 10 years old. The results obtained showed that the test had a good internal consistency and could be useful as an objective measure of frustration tolerance in children. In line with the literature, our data have shown no influence of gender or laterality of participants during the task and only 7% of the frustration measure could be explained by the influence of participants' age. On the other hand, the performance of the participants during the task has allowed us to classify them into six groups according to their performance, namely Low/High Frustration, Low/High Performance and Low/High Reaction Time. This test would permit to compare participants' performance with their reference group but also with their own results, facilitating the obtention of an objective assessment of frustration tolerance in young children.

1. Introduction

Missing public transportation on the way to the office, forgetting something essential or arriving late to a meeting are situations that can generate emotions related to frustration. Frustration can arise in any situation where our goals are compromised or not achieved and may significantly influence our subsequent behavior, negatively affecting our performance. There are many definitions of the phenomenon of frustration and its tolerance. The classic definition indicates that frustration is a temporary state that occurs when a response is not reinforced in presence of reinforcement expectations. Motivation or the desire to achieve a goal affects the level of frustration and this frustration, in turn, affects the behavior that follows (Amsel, 1992). In line with the classical definition, frustration is also defined as emotional and behavioral responses associated with blocked goal attainment (Leibenluft, 2011), as unexpected lack of reinforcement (Papini et al., 2019), or as goal-directed activities that are blocked (Yu et al., 2014).

Frustration, as a process dependent on emotional regulation, is related to vital aspects of people's lives such as health and social relationships (Yu et al., 2014) and can predict the outcome of future externalizing psychological symptoms (Jeronimus et al., 2017), increasing anger or aggression responses. Frustration intolerance has been strongly associated with anger and aggressive behavior in children and adolescents (Fives et al., 2011), as well as increased potential for child physical abuse (Rodríguez et al., 2015). In fact, in current models of aggression, frustration might work as a possible trigger for anger and aggression (Berkowitz, 2012).

When assessing frustration, irritability is considered an expression of frustrating lack of reward and in this regard, children with high levels of irritability commonly display increased levels of self-reported frustration (Seymour et al., 2020). Irritable children facing frustrating tasks might have difficulties using emotion-regulation strategies to shift their attention to useful stimuli (Deveney et al., 2013), being an essential competence in relation to emotional regulation processes (Tseng et al., 2017). In fact, frustration seems to diminish the attention flexibility, especially in severe irritability children, which may affect their emotional regulation processes (Deveney et al., 2013). In this regard, cognitive control is also considered crucial for emotional regulation processes (Posner & Rothbart, 2000) and its relationship with frustration has been studied for more than 40 years (Weiner & Adams, 1974).

* Corresponding author.
E-mail address: ajsoto@us.es (A. Jiménez-Soto).

https://doi.org/10.1016/j.actpsy.2022.103528
Received 2 October 2021; Received in revised form 30 January 2022; Accepted 1 February 2022
Available online 15 February 2022

Low levels of cognitive control may result in maladaptive behavioral responses, which could increase negative affect and reversely, frustration might impact negatively in cognitive control as well (Seymour et al., 2020). Data from Functional Near-Infrared Spectroscopy (fNIRS) support this idea, showing frustration could be related to both, impulse control and emotion-regulation processes, where a frustrating task displayed higher activity in areas related to cognitive appraisal, impulse control and emotion regulation processes (Ihme et al., 2018).

The relationship between cognitive development, gender and frustration tolerance is controversial. A few years ago, Oliva et al. (2011) adapted one of the most used questionnaires for emotional skills assessment, the Emotional Quotient Inventory (EQ-i, YV), originally developed by Bar-On and Parker (BarOn & Parker, 2001). In the scales reported for students aged 12–17, there were no significant differences due to age in frustration tolerance. Similarly, Deater-Deckard, Petrill, & Thompson (2007) found no significant differences in frustration tolerance among preschoolers aged four to eight years, either regarding age or gender. In addition, a general neuropsychological assessment made by Sebastian (2013) with a battery of neuropsychological and behavioral tests in 90 Psychology students did not report any significant gender differences for frustration tolerance levels either. Nevertheless, Seymour et al. (2020) did find differences between boys and girls using a go/no go frustration tolerance task. In this task boys showed higher levels of frustration than girls F (1,103) = 4.09; p = .046; d = 0.44, although they found no differences according to the age of the participants.

Low frustration tolerance is commonly present in the depressed population, causing them difficulties in social environments and increasing the risk of suicide in these patients (Korzenev et al., 2012). Similarly, people with high levels of anxiety demonstrated higher frustration scores when measured by questionnaires (Netter et al., 2014). Low frustration tolerance is a common behavioral characteristic among people with brain damage (Johnstone & Stonnington, 2009). It is present in patients with borderline personality disorder (Reich & Zanarini, 2001) and patients with substance use disorders (Ramirez-Castillo et al., 2019). In children with pediatric bipolar disorder, irritability and frustration are considered one of the most impairing symptoms of the condition (Carlson et al., 2003). In addition, ADHD is one of the disorders where tolerance to frustration is usually compromised. Anastopoulos et al. (2011) found that only 15% of participants with typical development showed high levels of emotional dysregulation, while up to 50% of people with ADHD showed this characteristic, concluding that it could be a key part of the disorder. In fact, the emotional regulation deficits of individuals with ADHD usually cause a low ability to tolerate frustration (Brown, 2013; Maedgen & Carlson, 2000). A study of 8- to 17-year-old children diagnosed with ADHD (Seymour et al., 2019) observed that when faced with a frustrating task, ADHD participants show less frustration tolerance, tending to drop out of the task earlier than typically developing participants. Looking for differences between subtypes, Flores (2009) analyzed a sample of participants with ADHD diagnosed, reporting that up to 35% of hyperactive-impulsive subtype and the 17% of inattentive subtype showed lower levels of frustration tolerance than control participants.

Nowadays, evaluation of frustration tolerance is usually carried out through standardized psychological tests, behavior scales and self-report questionnaires. Although there are good standardized measures for assessing this construct, such as the Frustration Discomfort Scale (Harrington, 2005), the Basic Psychological Needs Satisfaction and Frustration Scale BPNSFS (Chen et al., 2015), or the EQ-i, YV mentioned above (BarOn & Parker, 2001), few of these measures are suitable for youth (Trip & Bora, 2012) and less even for young children. On the other hand, although some computerized tests have been constructed to assess frustration with good results in adults (McElroy & Rodriguez, 2008) or young adults (Moreno et al., 2000), the presence of computerized neuropsychological tests to evaluate frustration tolerance in healthy children are scarce. For example, Tseng et al. (2017) tested the reliability and validity of a task to measure frustration in healthy children, but in older children (from 9 to 14 years old). Therefore, the main objective of the present study was to create a measurement tool of frustration tolerance to be used in school-age children. In this regard, we developed a computerized task that tries to induce frustration in the participants and after that, we analyzed its psychometric properties and suitability for the studied construct. For task design, we assumed frustration as the subjective emotion of dislike associated with high levels of effort and low levels of success in a task (Perlman et al., 2014). Many frustration tasks try to measure tolerance to frustration in humans by giving false feedback or making the task unsolvable. In our case, we decided to do it by restricting the possibilities of reinforcement to the minimum possible in order to minimize the possibilities of a learned helplessness effect.

Our test is designed for use in children aged 6–10 years because they are at a crucial time in their psychological development and it could provide an objective measure of a child's frustration tolerance levels. If this is true, the task should be useful to produce frustration in the participants, who would reduce their overall performance in Set 3, despite being identical to set 1. Additionally, this reduction should be largely explained by performance in Set 2 and beyond a fatigue effect during the task. Furthermore, we do not expect differences in frustration levels influenced by participants' age, nor influences of gender or manual preference. Finally, we expect participants with higher frustration scores will show higher reaction times and lower general performance. The paucity of objective measures, especially for children, may render our task potentially useful for the study of phenotypes and endophenotypes of those disorders in which tolerance to frustration is impaired (Jiménez-Soto et al., 2020; Vargas et al., 2016).

2. Materials and methods

2.1. Participants

A total of 160 children were evaluated. From the original sample of 180 participants, we eliminated 3 participants who did not complete the task properly, 8 participants for being diagnosed or in the evaluation process for any psychological condition, and 9 participants for being identified as outliers according to the Mahalanobis distance analysis. From the final sample (n = 160), there were 81 males and 79 females, 14 left-handed and 164 right-handed. Table 1 shows the demographic characteristics of the sample.

The participants were recruited from a collaborating Primary School at Seville in Spain. The groups were composed of 23 participants from 1st Grade (6 years old mean), 14 from 2nd Grade (7 years old mean), 41 from 3rd Grade (8 years old mean), 46 from 4th Grade (9 years old mean) and 36 from 5th Grade (10 years old mean). All participants had to provide informed consent signed by their parents to participate in the study. To carry out this study we have the approval of the Biomedical Research Ethics Coordinating Committee de Andalucía, Junta de Andalucía (Spain) with the code (1221-N-17).

2.2. Experimental task

The task used to evaluate frustration tolerance in this study was named the *Beach Balls Task*. Our task is based on the work carried out by

Table 1
Sample demographic characteristics.

Grade (mean age)	1st (6 years old)		2nd (years old)		3rd (8 years old)		4th (9 years old)		5th (10 years old)	
Gender	F[a]	M	F	M	F	M	F	M	F	M
Right-handed	9	12	9	5	21	16	19	22	19	14
Left-handed	0	2	0	0	0	4	1	4	1	2

[a] Abbreviations: F: female. M: male.

Moreno et al. (2000) to evaluate frustration using a computerized task and has been developed using JAVA JDK to ease its portability between operating systems.

The test was designed to be as simple as possible so children would be able to understand it easily. Sets of four beach balls of different sizes are successively shown in the central part of the screen. The goal of the task was to click with the mouse on the smallest ball in each group within the five-second time interval each trial lasts. The appearance of the task is shown in Fig. 1.

The time required in each trial was shown at the top of the screen, where a descending bar getting shorter indicated the remaining time to respond. The corresponding feedback was given after each trial in a counter at the bottom of the screen, where the hits (in green) and errors/omissions (in red) were also displayed.

All the instructions were shown on the screen at the beginning of the task to avoid possible influences by the administrator. The instructions displayed on the screen were: “Select with the mouse the smallest ball in each group. You can see your hits and errors on the screen. Try to be quick because you only have five seconds to respond”. Underneath the instructions, the app showed a button with the inscription “Understood” and when it was clicked a cross on the central part of the screen appeared. At the bottom, a button with the message “Start” allowed the task begins when it was clicked.

The task “Beach Balls” has a total of 80 trials of five second duration divided into three sets. Sets 1 and 3 are identical, having the same number of trials in which the balls appear, same order and position in each trial. Set 1 and Set 3 included 20 trials with a difference of 5% between ball sizes. In contrast, Set 2 had 40 trials and was the one trying to produce frustration in participants. To achieve it, we made the differences between ball sizes smaller, maintaining the five-second limitation to the response. In this set, the beach balls had only a difference of 2% in their sizes, which makes the discrimination between these groups more complicated. Due to the limited response time and the minimal difference between the ball sizes in this set, we tried to induce a state of frustration in the participants in order to affect their performance during the task.

When a participant completed the task, the screen turned to black and appeared a button with the inscription “End” to finish and exit. The program recorded the percentage size of each ball, which one was the smaller, the ball chosen, the response latency and if the response was correct, incorrect or an omission for every trial. For each session, we had information about the total number of hits, errors and omissions of the participant during the task as well as the total time required to complete it.

By analyzing the detailed performance of the participants between sets one and three we were able to check if frustration has occurred during the task. Since Set 1 and Set 3 were exactly the same, being the trials and items identical in both sets if a participant performs better in Set 1, we understood that was because of the influence of Set 2, or frustrating set, which has had an impact on participants subsequent behavior. If participants performed better in Set 3, we concluded there was a good level of frustration tolerance, which makes Set 2 unable to affect negatively to the task performance.

In order to assess frustration and the psychometrical properties of the task, we have calculated a set of new variables from the original ones. On the one hand, these variables are the sum of hits, errors, omissions and latencies of each set and the total sum of hits, errors, omissions and latencies for the complete task. We called them summation scores. On the other hand, these variables permitted us to calculate the indexes that comprised the overall performance of a participant in each set.

To calculate the performance index of each set, we subtracted the number of errors and omissions from the number of hits during the set, thus we have an index for each set of trials and a General Performance Index for the complete task. These indexes allowed us to explore the performance of participants globally during the different sets and to facilitate their subsequent understanding. Comparing the indexes of Set 1 (Index 1) and Set 3 (Index 3) we were able to check if frustration has occurred. It served as a Frustration Index. The final Frustration Index was the result of subtracting Index 1 from Index 3. Negative values in this variable showed a decrease in task efficiency after Set 2, so frustration might be assumed.

2.3. Procedure

This research has followed the parameters of the CODE OF ETHICS OF THE WORLD MEDICAL ASSOCIATION. DECLARATION OF HELSINKI and has approval of the ethics committee of the Consejería de Igualdad, Salud y Asuntos Sociales from Junta de Andalucía, Spain. The

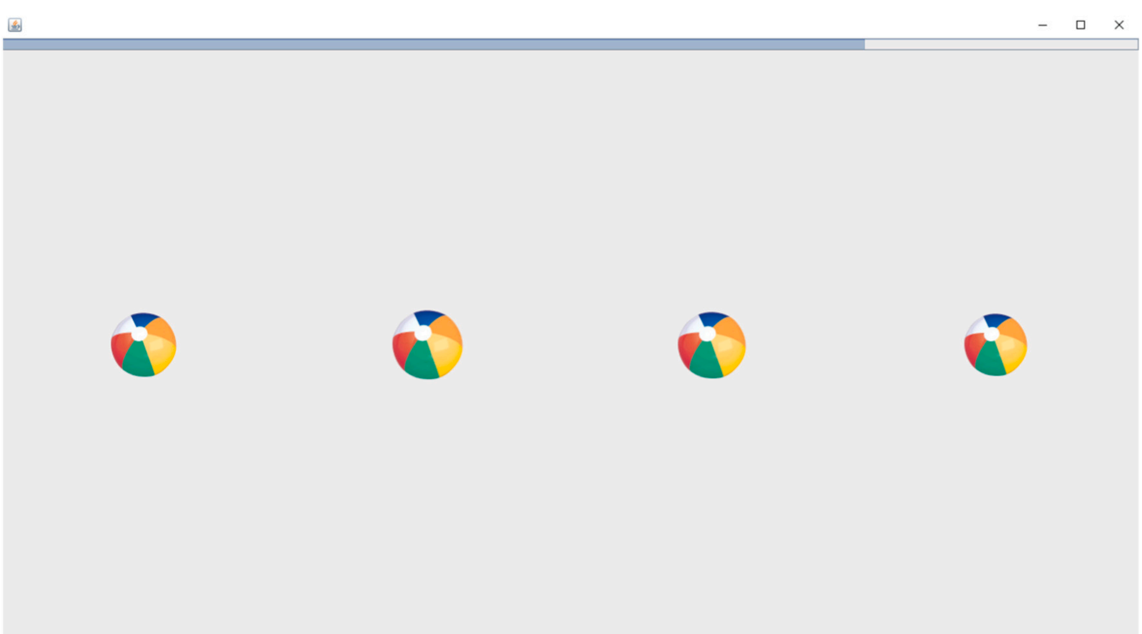

Fig. 1. Appearance of the task on screen. The upper bar shows response time and the lower bar shows hits (green) and errors (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

data collection was carried out in the IT room of the collaborating School. The room was equipped with five computers of identical properties and participants were evaluated in groups of five. Tests were done from 09.00 am to 13.00 am. and we used the software IBM SPSS Statistics 26 for data analysis. All participants provided informed consent, signed by their fathers, mothers or legal guardians to participate in the research. At the end of the task, participants were informed of the task objective and the measure we were trying to obtain trying to induce frustration through Set 2.

3. Results

For the analysis of the psychometric properties of the task, and given that the test trial scores were dichotomous (correct/incorrect), we calculated reliability using the Kuder-Richardson test (KR20) which yielded a reliability index of 0.870, a score considered to be a good indicator of the internal consistency of the test. To evaluate the content validity of the instrument, six collaborating judges participated. They independently scored the characteristics of the test regarding the construct we were trying to measure and its suitability to the application context. The group of judges was composed of three research Professors of the Degree and Postgraduate course in Psychology, a psychologist specialized in ADHD, a Developmental Psychologist and a Child Neuropsychologist. The judges agreed to give the maximum adjustment score on this test for its final version concerning: the aim of the task, trials length, task length, distribution of sets and the frustration measure.

In addition to the validity evaluation conducted by the team of judges, we wanted to test the effectiveness of the task to produce frustration in the participants. With this aim, we analyzed the means of the performance rates through the *t* statistic, which showed significant differences between the scores of Set 1 and Set 3 (t (159) = 44,161; $p <$ .000) despite being identical. The Index 1 (M = 16.64; SD = 4.766), Index 2 (M = 8.31; SD = 11.23) and Index 3 (M = 14.79; SD = 6.08) scores indicated that participants in general reduced their effectiveness during the task, obtaining worse scores during Set 3, after having concluded the frustration set.

To see if reduced effectiveness found in Set 3 could be due to frustration generated by the task, we also analyzed the linear relationship between the performance of the participants in Set 2 and the performance of Set 3 through Pearson's correlation analysis (r = 0.625; $p <$ .000). The effect size was large (r^2 = 0.391; $p <$ .000), indicating that the influence of Set 2 on Set 3 was high and could explain up to 39% of the variability found in the data. In Fig. 2, can be observed linear relationship between these variables, where it can be seen that the better performance in the set of frustration, the better results during the final set of the task.

In order to discard the possible influence of fatigue during the task, that is, if the performance decrease depends on the tiredness, we divided Set 3 into two halves subsets of ten trials each. Then, we compared the mean performance within these subsets verifying there were significant differences between performance subsets (t (159) = 73,155; p = .000), with an advantage for the second half (M = 8.99; SD = 1.554) in comparison to the first (M = 8.41; SD = 1.861). These results suggest that fatigue had no significant influence during the development of the test since the overall performance was significantly better in the last subset of the task.

To analyze these results in more depth, we divided participants into a Frustrated group (those with a negative Frustration Index score) and a Non-Frustrated group (those with a Frustration Index score equal to or greater than 0). The distribution of participants into the Frustrated group and the Non-frustrated group was similar. From the total of 160 subjects, 81 obtained negative scores on the Frustration Index and 79 obtained scores equal to or greater than 0.

Using this distribution, we observe that results are similar to the overall sample, with both groups scoring better in the second half of Set 3. Frustrated group (M = 7.60; SD = 2.010) for the first subset and (M = 8.51; SD = 1.762) for the second subset and Non-Frustrated group (M = 9.23; SD = 1.250) for the first subset and (M = 9.48; SD = 1.119) for the second set. These results might confirm that fatigue during the task was not a decisive factor in the measure of frustration.

Through the analysis of variances, we first studied the possible relationships between task performance and the school year of belonging through *Grade* variable. Almost all the relationships studied were sensitive to the influence of this variable. Tables 2–4 show the complete analysis.

The effect sizes of these relationships were important in some cases. The variable Grade could explain up to 26% of the variability found in the total hits (r^2 = 0.264; p = .000), up to 17% of errors/time spent (r^2 = 0.173; p = .000) and up to 11% of omissions (r^2 = 0.110; p = .001). Concerning to performance indexes, Grade could explain up to 11% of the variability found in the scores of Index 1 (r^2 = 0.110; p = .001), besides 22% of the differences of Index 2 (r^2 = 0.229; p = .000) and Index 3 (r^2 = 0.223; p = .000). In fact, Grade could explain up to 26% of the General Performance Index (r^2 = 0.264; p = .000). These results confirm that Grade variable, associated with the mean age of

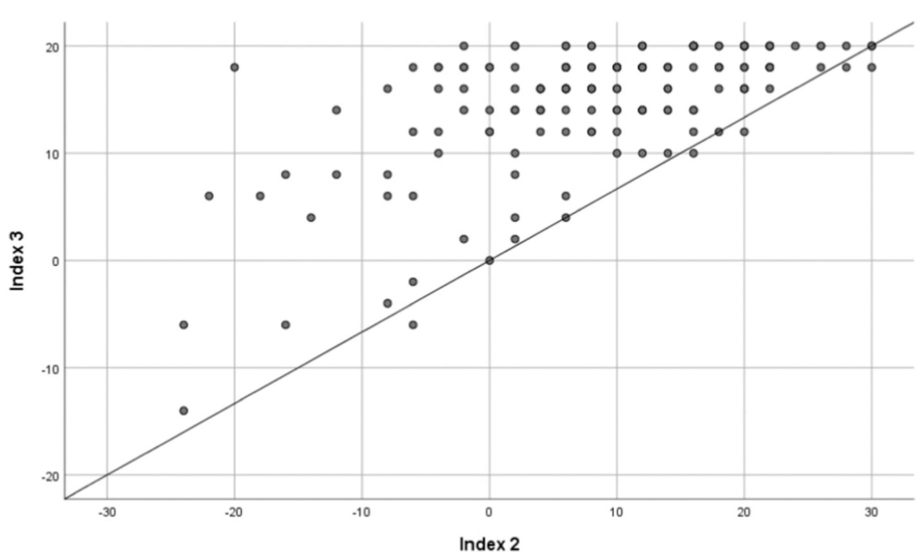

Fig. 2. Linear regression analysis between performance during Set 2 (frustrating set) and Set 3.

Table 2
Mean scores, correlations and effect sizes according to Grade.

	Sample	1st Grade	2nd Grade	3rd Grade	4th Grade	5th Grade	*F*, *r* and effect sizes
	N = 160	*n* = 23	*n* = 14	*n* = 41	*n* = 46	*n* = 36	
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	
Hits	59,87 (9,57)	50,57 (10,49)	52,07 (15,38)	59,44 (6,45)	64 (6,24)	64,06 (6,26)	(F (4,159) = 16.075; p = .000) (r = 0.514; p = .000) (r^2 = 0.264; p = .000)
Errors	17,51 (7,75)	23,04 (7,52)	24,21 (11,99)	17,85 (5,88)	14,39 (6,24)	14,94 (6,04)	(F (4,159) = 10.416; p = .000) (r = −0.417; p = .000) (r^2 = 0.173; p = .000)
Omiss[a]	2,63 (5,07)	6,39 (10,16)	3,71 (6,49)	2,71 (3,96)	1,61 (1,26)	1 (1,55)	(F (4,159) = 5.226; p = .001) (r = −0.332; p = .000) (r^2 = 0.110; p = .001)
RT[b]	202,972,59 (29,352,22)	224,736,3 (40,333,08)	208,164,79 (34,436,21)	211,269,59 (25,830,8)	195,547,91 (22,114,67)	187,086,53 (18,024,05)	(F (4,159) = 8.955; p = .001) (r = −0.417; p = .001) (r^2 = 0.173; p = .000)

[a] Abbreviations: Omiss: omissions. RT: reaction time.
[b] Reaction times: expressed in milliseconds.

Table 3
Mean scores for the three task sets according to Grade.

		Sample	1st Grade	2nd Grade	3rd Grade	4th Grade	5th Grade
		N = 160	*n* = 23	*n* = 14	*n* = 41	*n* = 46	*n* = 36
		M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
SET 1	Hits	18,32 (2,38)	16,91 (2,71)	16,29 (4,83)	18,68 (1,42)	18,74 (1,72)	19,06 (1,58)
	Errors	1,39 (1,97)	2,13 (2,03)	3 (4,02)	1,12 (1,21)	1,2 (1,69)	0,83 (1,38)
	Omissions	0,29 (0,9)	0,96 (1,69)	0,71 (1,64)	0,2 (0,46)	0,07 (0,25)	0,11 (0,32)
	Reaction time	45,376,24 (9609,34)	54,948,48 (11,773,67)	50,174,21 (12,589,35)	46,950 (7675,43)	41,941,89 (6202,24)	39,990,75 (6248,36)
SET 2	Hits	24,16 (5,62)	19,39 (5,63)	20,21 (6,29)	23,22 (4,87)	26,72 (4,37)	26,53 (4,51)
	Errors	13,84 (4,83)	16,43 (5,25)	17,43 (4,91)	14,49 (4,27)	11,78 (4,21)	12,69 (4,35)
	Omissions	2 (3,72)	4,17 (7,14)	2,36 (3,56)	2,29 (3,88)	1,5 (1,24)	0,78 (1,4)
	Reaction time	112,143,48 (16,998,57)	116,897,57 (23,450,56)	110,061,86 (19,193,06)	116,496,24 (18,284,28)	110,833,43 (14,247,8)	106,632,28 (10,715,38)
SET 3	Hits	17,39 (3,04)	14,26 (4,1)	15,57 (5,03)	17,54 (2,07)	18,54 (1,8)	18,47 (1,5)
	Errors	2,28 (2,55)	4,48 (3,4)	3,79 (4,06)	2,24 (1,92)	1,41 (1,77)	1,42 (1,44)
	Omissions	0,33 (1,13)	1,26 (2,38)	0,64 (1,6)	0,22 (0,52)	0,04 (0,21)	0,11 (0,32)
	Reaction time	45,452,88 (9073,6)	52,890,26 (12,598,9)	47,928,71 (9289,13)	47,823,34 (8010,47)	42,772,59 (6466,22)	40,463,5 (5887,83)

Table 4
Mean scores for task indexes according to Grade.

	Sample	1st Grade	2nd Grade	3rd Grade	4th Grade	5th Grade	*F*, *r* and effect sizes
	N = 160	*n* = 23	*n* = 14	*n* = 41	*n* = 46	*n* = 36	
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	
Index Set 1	16,64 (4,77)	13,83 (5,42)	12,57 (9,65)	17,37 (2,84)	17,48 (3,44)	18,11 (3,17)	(F (4,159) = 8.955; p = .000) (r = 0.332; p = .000) (r^2 = 0.110; p = .001)
Index Set 2	8,31 (11,23)	−1,22 (11,26)	0,43 (12,58)	6,44 (9,73)	13,43 (8,75)	13,06 (9,03)	(F (4,159) = 13.286; p = .000) (r = 0.479; p = .000) (r^2 = 0.229; p = .000)
Index Set 3	14,79 (6,09)	8,52 (8,21)	11,14 (10,07)	15,07 (4,15)	17,09 (3,6)	16,94 (3)	(F (4,159) = 13.280; p = .000) (r = 0.472; p = .000) (r^2 = 0.223; p = .000)
General Performance Index	39,74 (19,13)	21,13 (20,97)	24,14 (30,76)	38,88 (12,9)	48 (12,49)	48,11 (12,51)	(F (4,159) = 16.075; p = .000) (r = 0.514; p = .000) (r^2 = 0.264; p = .000)
Frustration Index	-1,85 (4,87)	−5,3 (8,48)	−1,43 (4,86)	−2,29 (3,65)	−0,39 (3,62)	−1,17 (3,22)	(F (4,159) = 4.592; p = .002) (r = 0.264; p = .001) (r^2 = 0.070; p = .001)

participants, had a strong influence on task performance. Nevertheless, factor Grade could only explain about 7% of Frustration Index scores ($r^2 = 0.070$; $p = .001$) being the smallest effect size found for this variable.

For further analysis, we computed six groups of participants based on the results given by three variables of our task: General Performance Index, Total Reaction Time and Frustration Index. We carried out this categorization to perform more in-depth analyses because they are the most representative variables of the task. The scores of these three variables were converted to z-scores to select only the participants who scored below the 15th percentile (creating the "Low Groups") and those who scored above the 85th percentile (creating the "High Groups"). Afterwards, the sample was composed of six groups: Low/High Performance, Low/High Frustration and Low/High Reaction Time.

The first analysis we carried out was aimed at verifying whether Grade variable produced differences in the configuration between low and high groups. Thus, we found that groups configured according to General Performance (t (49) = −6.459; $p = .000$; $\chi^2 = 24.81$; $p = .000$) and Reaction Time (t (46) = 5.064; $p = .000$; $\chi^2 = 23.23$; $p = .000$) were influenced by mean age of participants. However, there was no significantly influence of Grade in the configuration of Low/High Frustration groups (t (55) = −1.349; $p = .183$; $\chi^2 = 3.404$; $p = .493$).

According to Cohen's classification of effect sizes through Cramer's Phi statistic, the effect size of Grade in the configuration of Low/High Performance groups would be considered large ($\phi c = 0.697$; $p = .000$), as well as the effect on Low/High Reaction Time groups ($\phi c = 0.696$; $p = .000$). Thus, we can observe as we hypothesized, how Grade influences both the reaction times during the test and the general performance but does not directly affect the participant's frustration levels. This is relevant since confirms the idea of frustration scores, assessed by our task, are not significantly influenced by the age of the participants.

During data analysis, we also explored whether there might be differences in participants' scores based on gender or laterality through analysis of variance. Nevertheless, we found no differences in the performance of participants that could be explained by gender or laterality (all $ps < .05$). Additionally, we checked if the gender or hand/manual preference of participants had any influence when setting up Low/High groups. Analyses carried out did show no possible relationships between variables Gender or Laterality on the configuration of the six Low/High groups (all $ps > .05$).

We conducted additional analyses using the Low/High based groups. Tables 5–7 show data obtained from the comparison of average scores for each group among test parameters. Performance-based Low/High groups displayed significant differences in scores for every parameter of the test, including hits, errors, omissions, reaction times, and indexes (all $ps < .05$). These results are consistent with expected since better performance on the test has to be related to better scores on assessed parameters. Groups formed by their reaction time showed statistically significant scores for the number of hits (t (46) = 2397; $p = .021$) and omissions (t (46) = −4.016; $p = .000$), as well as for the Set 2 performance (t (46) = 3.40; $p = .001$), and the General Performance Index (t (46) = 2397; $p = .021$). As expected, a longer reaction time will affect both the number of correct trials and the number of missed trials. Finally, for participants in Low/High Frustration groups differences were only significant for performance of Set 1 (t (55) = 2782; $p = .007$), Set 3 (t (55) = −5.764; $p = .000$) and Frustration Index (t (55) = −16.43; $p = .000$). These results are interesting because they emphasize anew that degree of frustration seems not to be strongly related to performance accuracy during the task.

In order to explore these results in depth, we wanted to calculate the Odds Ratio (OR), an analysis traditionally applied to estimate the probability of a certain event occurring. Concerning our data, we were able to observe that the relative risk of belonging to Low Reaction Time groups is 7.52 times higher for those participants who belong to High-Performance groups than those who belong to Low-Performance groups. On the contrary, participants who were classified in High Reaction Time groups were up to 5.13 times more likely to belong to Low-Performance groups than to High-Performance groups. Finally, participants in the Low-Frustration group were up to 2.5 times more likely to belong to the Low Reaction Time group than to the High Reaction Time group. The results above confirm what was expected. Longer reaction

Table 5
Mean scores during the task for low/high performance groups.

Groups	Performance		
	Low performance	High performance	t (p-value) gl = 49
N	26	25	
Hits	43,115 (8,93)	71 (1.80)	−15,29 (p < .000)
Errors	28,884 (8,99)	8,36 (1846)	11,186 (p < .000)
Omissions	8 (10,38)	0,64 (0,81)	3532 (p = .001)
RT[a]	218,105 (43,729,92)	191,247,04 (22,852,29)	2732 (p = .009)
Index Set 1	9846 (7482)	19,44 (1781)	−6241 (p < .000)
Index Set 2	−8307 (8867)	23,12 (3789)	−16,34 (p < .000)
Index Set 3	4,69 (7714)	19,44 (0,917)	−9,49 (p < .000)
G. P. Index	6230 (17,874)	62 (3606)	−15,299 (p < .000)
Frustration Index	−5154 (8601)	0 (1826)	−2932 (p = .005)

[a] Abbreviations: RT: reaction time. G.P.: general performance.

Table 6
Mean scores during the task for low/high reaction time groups.

Groups	Reaction time		
	Low RT	High RT	t (p-value) gl = 46
N	24	24	
Hits	59,75 (11,96)	51,25 (12,595)	2397 (p = .021)
Errors	19,458 (11,436)	19,375 (7917)	0,029 (p > .05)
Omissions	0,79 (1215)	9375 (10,4)	−4016 (p < .000)
RT[a]	162,363,375 (10,595,75)	250,802,458 (17,546,39)	−21,137 (p < .000)
Index Set 1	16,083 (7162)	14,083 (6171)	1036 (p > .05)
Index Set 2	10,416 (11,97)	−1916 (13,128)	3,40 (p = .001)
Index Set 3	13 (7,59)	10,33 (9168)	1097 (p > .05)
G. P. Index	39,5 (23,92)	22,5 (25,19)	2397 (p = .021)
Frustration Index	−3083 (5748)	−3,75 (7326)	0,35 (p > .05)

[a] Abbreviations: RT: reaction time. G.P.: general performance.

Table 7
Mean scores during the task for low/high frustration groups.

Groups	Frustration		
	Low tolerance	High tolerance	t (p-value) gl = 55
N	26	31	
Hits	51,61 (11,10)	57 (11,35)	−1,8 (p > .05)
Errors	23,34 (7,73)	19,16 (9,68)	1,77 (p > .05)
Omissions	5,04 (9425)	3,84 (5496)	0,60 (p > .05)
RT[a]	212,678,07 (41,430,94)	211,653,06 (31,189,98)	0,106 (p > .05)
Index Set 1	16,15 (5,15)	11,80 (6415)	2782 (p = .007)
Index Set 2	1,23 (11,45)	6,38 (12,89)	−1581 (p > .05)
Index Set 3	5,85 (7,39)	15,81 (5,64)	−5764 (p < .000)
G. P. Index	23,23 (22,20)	34 (22,70)	−1801 (p > .05)
Frustration Index	−10,30 (4,10)	4 (2,36)	−16,43 (p < .000)

[a] Abbreviations: RT: reaction time. G.P.: general performance.

Table 8
Different low/high groups combinations and participants frequencies.

Frustration	Performance	Reaction time	(*n*)
High	Low	Low	3
High	Low	High	6
High	High	Low	0
High	High	High	0
Low	Low	Low	1
Low	Low	High	3
Low	High	Low	1
Low	High	High	0

times were related to poorer performance and a higher probability of experiencing frustration during the course of the task.

Table 8 provides frequencies of participants, according to their belonging to different Low/High groups. It can be seen that there were up to three group combinations with no participants, High Frustration + High Performance + Low Reaction Time, High Frustration + High Performance + High Reaction Time and Low Frustration + High Performance + High Reaction Time. It is important to note that, as we expected, the largest group of participants consists of the combination formed by High Frustration + High Performance + Low Reaction Time, with a frequency of six participants.

4. Discussion and conclusions

The psychometric analysis of the Beach Ball task has shown a good internal consistency, allowing us to discriminate between participants with high/low levels of frustration tolerance through the Frustration Index. The content validity was also satisfactory according to judges; however, further studies will allow us to know the convergent validity of the test with respect to other tests.

After the analyses were carried out, we could infer that the construction of our task has been able to produce a change in participants performance compatible with a frustration response. This is confirmed by several results. On the one hand, we verified that the general performance of participants is significantly reduced during Set 3, despite being completely identical to Set 1. In addition, we have confirmed a strong influence from Set 2 (frustration) to the performance of Set 3, being able to explain up to 39% of the variability found in the performance of the last set. Is possible this influence could be explained due to the increased cognitive load required by the second set, even though this is partly what we were looking for in test design, the increased difficulty of the second set would affect the performance of the third set, despite both being exactly identical.

Additional analyses were conducted to test whether frustration scores could be explained beyond a fatigue effect during the task. We were able to verify that decrease in performance during Set 3 could hardly be explained by the effect of fatigue during the task since we could observe that participants, in general, significantly improved their performance during the second half of Set 3. In addition, frustrated and non-frustrated participants showed the same results, significantly improving their performance during the second subset of Set 3, which would be difficult to find if fatigue was occurring during the task.

Regarding the influence of the mean age of participants, even when the variable Grade could explain up to 26% of general performance or 17% of reaction times, it could only explain 7% of the variability of frustration scores. These results suggest that the cognitive development of participants affects both, the ability to discriminate stimuli and the processing speed during the test, but it does not considerably impact the measure of frustration we have created. Results were also confirmed by the analysis of Grade influence on the configuration of Low/High groups, where the performance groups and the reaction time groups had a strong influence from this variable with very large effect sizes. However, groups configured according to their frustration levels were not significantly influenced by the effect of this variable. This is relevant as it confirms the previous findings (Deater-Deckard et al., 2007; Oliva et al., 2011), the degree of frustration assessed by our task is not significantly influenced by the age of the participants.

Concerning the influence of gender on frustration tolerance, in line with Sebastian (2013) or Deater-Deckard et al. (2007), we found no differences in task performance that could be explained by the influence of the participant's gender. None of the parameters controlled during our task was shown to be sensitive to the influence of participants' gender. Similarly, regarding manual preference, the laterality of participants appears to not influence their ability to tolerate frustration in children.

In addition to the results described and in line with previous studies (Ihme et al., 2018; Posner & Rothbart, 2000; Seymour et al., 2020), the measure of frustration tolerance provided by this test, in a prior study of our group, was significantly correlated with the performance on a task assessing impulse control (Jiménez-Soto, 2020). This data suggests that the relationship between cognitive control and frustration tolerance could be important and deserves to be explored in future research.

Finally, according to expected when conducting this research, using this task participants will obtain higher frustration scores when they display longer reaction times and lower overall performance. This is very significant, as worse performance will be determined by a poorer ability to discriminate visual stimuli, making the second set more complicated and frustrating. In addition, the longer the reaction time, the greater the likelihood of committing trial omissions, especially in the second set where the size differences between balls are lower.

One of the most important limitations of this study is about sample size ($n = 160$) that does not have enough participants to provide standardized scores. However, one of the strengths of our task is that allows us to measure frustration tolerance through the comparison of a participant's performance with itself, therefore, it is not essential to have standardized scores. Either way, a systematic evaluation with a representative sample of this population could provide standard metrics for the test in the future.

To conclude, due to its configuration and difficulty, this test is not suitable for use with elder children. Increasing the difficulty of stimuli discrimination by decreasing relative size between balls, reducing time interval to respond or increasing the number of trials, could adjust this measure to be used in youth or young adults.

Declaration of competing interest

None.

Acknowledgements and funding

This research project has been funded by Gobierno de Espana Ministerio de Ciencia Innovacion y Universidades: PID2019-110739GB-I00/AEI/10.13039/501100011033 and Junta de Andalucía (HUM-646)

References

Amsel, A. (1992). *Frustration theory: An analysis of dispositional learning and memory*. Cambridge: Cambridge University Press.

Anastopoulos, A. D., Smith, T. F., Garrett, M. E., Morrissey-Kane, E., Schatz, N. K., Sommer, J. L., et al. (2011). Self-regulation of emotion, functional impairment, and comorbidity among children with AD/HD. *Journal of Attention Disorders, 15*(7), 583–592. https://doi.org/10.1177/1087054710370567

BarOn, R., & Parker, J. (2001). *EQi: YV BarOn emotional quotient inventory: Youth version. Technical manual*. Toronto: Multi-Health Systems Inc.

Berkowitz, L. (2012). A cognitive-neo association theory of aggression. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 99–117). Thousand Oaks: Sage.

Brown, T. E. (2013). *A new understanding of ADHD in children and adults: Executive function impairments*. Washington DC: Routledge.

Carlson, G. A., Hellander, M., Kowatch, R., Kusumakar, V., Laughren, T., Leibenluft, E., et al. (2003). Methodological issues and controversies in clinical trials with child and adolescent patients with bipolar disorder: Report of a consensus conference. *Journal of Child Adolesc Psychopharmacology, 13*(1), 13–27. https://doi.org/10.1089/104454603321666162

Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, J., et al. (2015). Basic psychological need satisfaction, need frustration, and need strength across four cultures. *Motivation and Emotion, 39*(2), 216–236. https://doi.org/10.1007/s11031-014-9450-1

Deater-Deckard, K., Petrill, S. A., & Thompson, L. A. (2007). Anger/frustration, task persistence, and conduct problems in childhood: A behavioral genetic analysis. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 48*(1), 80–87. https://doi.org/10.1111/j.1469-7610.2006.01653.x

Deveney, C. M., Connolly, M. E., Haring, C. T., Bones, B. L., Reynolds, R. C., Kim, P., et al. (2013). Neural mechanisms of frustration in chronically irritable children. *The American Journal of Psychiatry, 170*(10), 1186–1194. https://doi.org/10.1176/appi.ajp.2013.12070917

Fives, C. J., Kong, G., Fuller, J. R., & DiGiuseppe, R. (2011). Anger, aggression, and irrational beliefs in adolescents. *Cognitive Therapy Research, 35*(3), 199–208. https://doi.org/10.1007/s10608-009-9293-3

Flores, J. C. (2009). Características de comorbilidad en los diferentes subtipos de trastorno por déficit de atención con hiperactividad. *Psicothema, 21*(4), 592–597.

Harrington, N. (2005). The frustration discomfort scale: Development and psychometric properties. *Clinical Psychology & Psychotherapy, 12*(5), 374–387. https://doi.org/10.1002/cpp.465

Ihme, K., Unni, A., Rieger, J. W., & Jipp, M. (2018). Assessing driver frustration using functional near-infrared spectroscopy (fNIRS). In H. Ayaz, & F. Dehais (Eds.), *Neuroergonomics: The brain at work and in everyday life* (pp. 215–216). London: Academic Press. https://doi.org/10.1016/B978-0-12-811926-6.00042-7.

Jeronimus, B. F., Riese, H., Oldehinkel, A. J., & Ormel, J. (2017). Why does frustration predict psychopathology? Multiple prospective pathways over adolescence: A TRAILS study. *European Journal of Personality, 31*(1), 85–103. https://doi.org/10.1002/per.2086

Jiménez-Soto, A. (2020). *Batería Informatizada para la Evaluación Neuropsicológica de las Afectaciones Relacionadas con el TDAH: BIENART*. Sevilla: Universidad de Sevilla. Ph. D dissertation.

Jiménez-Soto, A., Vargas, J. P., Díaz, E., & López, J. C. (2020). Traditional scales diagnosis and endophenotypes in attentional deficits disorders: Are we on the right track? In H. G. Kumperscak (Ed.), *ADHD*. London: Intech Open. https://doi.org/10.5772/intechopen.94507.

Johnstone, B., & Stonnington, H. H. (2009). *Rehabilitation of neuropsychological disorders: A practical guide for rehabilitation professionals* (2nd ed.). London: Psychology Press.

Korzenev, A., Lebedev, A., & Abritalin, E. (2012). P-790 - neural correlates of frustration in depressed patients and healthy controls: An fMRI study. *European Psychiatry, 27* (S1), 1. https://doi.org/10.1016/s0924-9338(12)74957-x

Leibenluft, E. (2011). Severe mood dysregulation, irritability, and the diagnostic boundaries of bipolar disorder in youths. *American Journal of Psychiatry, 168*(2), 129–142. https://doi.org/10.1176/appi.ajp.2010.10050766

Maedgen, J. W., & Carlson, C. L. (2000). Social functioning and emotional regulation in the attention deficit hyperactivity disorder subtypes. *Journal of Clinical Child and Adolescent Psychology, 29*(1), 30–42. https://doi.org/10.1207/s15374424jccp2901_4

McElroy, E. M., & Rodriguez, C. M. (2008). Mothers of children with externalizing behavior problems: Cognitive risk factors for abuse potential and discipline style and practices. *Child Abuse & Neglect<icepaste/>, 32*(8), 774–784. https://doi.org/10.1016/j.chiabu.2008.01.002

Moreno, A. L., Hernández, J. M., García Leal, O., & Santacreu, J. (2000). Un test informatizado Para la evaluación de la tolerancia a la frustración. *Anales de<span/>Psicología, 16*(2), 143–155.

Netter, P., Baars, M., Weber, J., & Abuhalil, A. (2014). Are anxiety and impulsiveness differently related to frustrations from '"positive"' and '"negative"' punishment? in alcohol dependent patients and controls? *Pers Individ Dif., 60*, S25–S47. https://doi.org/10.1016/j.paid.2013.07.032

Oliva, A., Antolín, L., Pertegal, M. A., Ríos, M., Parra, A., Hernando, A., & Reina, M. C. (2011). *Instrumentos Para la evaluación de la salud mental y el desarrollo positivo adolescente y los activos que lo promueve*. Consejería de Salud: Junta de Andalucía.

Papini, M. R., Penagos-Corzo, J. C., & Pérez-Acosta, A. M. (2019). Avian emotions: Comparative perspectives on fear and frustration. *Frontiers in Psychology, 9*, 1–14. https://doi.org/10.3389/fpsyg.2018.02707

Perlman, S. B., Luna, B., Hein, T. C., & Huppert, T. J. (2014). FNIRS evidence of prefrontal regulation of frustration in early childhood. *NeuroImage, 85*(1), 326–334. https://doi.org/10.1016/j.neuroimage.2013.04.057

Posner, M. I., & Rothbart, M. K. (2000). Developing mechanisms of self-regulation. *Development and Psychopathology, 12*(3), 427–441. https://doi.org/10.1017/S0954579400003096

Ramirez-Castillo, D., Garcia-Roda, C., Guell, F., Fernandez-Montalvo, J., Bernacer, J., & Morón, I. (2019). Frustration tolerance and personality traits in patients with substance use disorders. *Frontiers in Psychiatry, 10*(421), 1–12. https://doi.org/10.3389/fpsyt.2019.00421

Reich, D. B., & Zanarini, M. C. (2001). Developmental aspects of borderline personality disorder. *Harvard Review of Psychiatry, 9*(6), 294–301. https://doi.org/10.1080/10673220127909

Rodríguez, C. M., Russa, M. B., & Kircher, J. C. (2015). Analog assessment of frustration tolerance: Association with self-reported child abuse risk and physiological reactivity. *Child Abuse & Neglect, 46*, 121–131. https://doi.org/10.1016/j.chiabu.2015.02.017

Sebastian, V. (2013). Gender differences in the attitudes towards work among young students. Cognitive and motivational features. *Procedia - socBehavioral Science, 78*, 551–555. https://doi.org/10.1016/j.sbspro.2013.04.349

Seymour, K. E., Macatee, R., & Chronis-Tuscano, A. (2019). Frustration tolerance in youth with ADHD. *Journal of Attention Disorders, 23*(11), 1229–1239. https://doi.org/10.1177/1087054716653216

Seymour, K. E., Rosch, K. S., Tiedemann, A., & Mostofsky, S. H. (2020). The validity of a frustration paradigm to assess the effect of frustration on cognitive control in school-age children. *Behavior Therapy, 51*(2), 268–282. https://doi.org/10.1016/j.beth.2019.06.009

Trip, S., & Bora, C. (2012). Psychometric properties of low frustration tolerance scale for students. *Procedia - Social and Behavioral Sciences, 33*, 578–582. https://doi.org/10.1016/j.sbspro.2012.01.187

Tseng, W. L., Moroney, E., Machlin, L., Roberson-Nay, R., Hettema, J. M., Carney, D., et al. (2017). Test-retest reliability and validity of a frustration paradigm and irritability measures. *Journal of Affective Disorders, 212*, 38–45. https://doi.org/10.1016/j.jad.2017.01.024

Vargas, J. P., Díaz, E., Portavella, M., & López, J. C. (2016). Animal models of maladaptive traits: Disorders in sensorimotor gating and attentional quantifiable responses as possible endophenotypes. *Frontiers in Psychology, 7*(206), 1–9. https://doi.org/10.3389/fpsyg.2016.00206

Weiner, A. S., & Adams, W. V. (1974). The effect of failure and frustration on reflective and impulsive children. *Journal of Experimental Child Psychology, 17*(2), 353–359.

Yu, R., Mobbs, D., Seymour, B., Rowe, J. B., & Calder, A. J. (2014). The neural signature of escalating frustration in humans. *Cortex, 54*(1), 165–178. https://doi.org/10.1016/j.cortex.2014.02.013