Check for updates

Article

Psychometrics of the Iowa and Berlin Gambling Tasks: Unresolved Issues With Reliability and Validity for Risk Taking

Assessment
2020, Vol. 27(2) 232–245
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1073191117750470
journals.sagepub.com/home/asm
SAGE

Florian Schmitz[1], Olga Kunina-Habenicht[2], Andrea Hildebrandt[3], Klaus Oberauer[4], and Oliver Wilhelm[1]

Abstract

The Iowa Gambling Task (IGT) is one of the most prominent paradigms employed for the assessment of risk taking in the laboratory, and it was shown to distinguish between various patient groups and controls. The present study was conducted to test the psychometric characteristics of the original IGT and of a new gambling task variant for assessing individual differences. Two studies were conducted with adults of the general population (n = 220) and with adolescents (n = 389). Participants were also tested on multiple measures of working memory capacity, fluid intelligence, personality traits associated with risk-taking behavior, and self-reported risk taking in various domains. Both gambling tasks had only moderate retest reliability within the same session. Moderate relations were obtained with cognitive ability. However, card selections in the gambling tasks were not correlated with personality or risk taking. These findings point to limitations of IGT type gambling tasks for the assessment of individual differences in risky decision making.

Keywords

Iowa Gambling Task, working memory capacity, fluid intelligence, risk taking

Since its introduction in 1994, the Iowa Gambling Task (IGT; Bechara, Damasio, Damasio, & Anderson, 1994) has become one of the most popular research tools for investigating risky decision making in the laboratory. The task distinguishes various patient groups from controls (Buelow & Suhr, 2009), and a manual was published (Bechara, 2007) suggesting the use of the IGT for assessment purposes. The aim of the present study was to test the suitability of the IGT as an individual differences measure of risk taking. We also tested a task variant designed to reduce some of the conceptual challenges of the original task. Psychometric properties were tested for both task versions. This included the investigation of retest reliability and of convergent and discriminant validity with cognitive abilities, relevant personality variables, and risk-taking behavior.

Measurement Intention and Procedure

Assessing Complex Decision Making in the Laboratory

The construction of the IGT was motivated by the study of patients with lesions in the ventromedial prefrontal cortex. These patients frequently displayed major deficits in real-world decision making, showing short-sighted and risk-taking behavior (Bechara & Damasio, 2002; Bechara, Dolan, & Hindes, 2002), while neuropsychological tests revealed that their declarative memory and abstract reasoning was often in the normal range. Hence, there appeared to be a demand for a standardized laboratory paradigm that is sensitive to the observed decision-making deficits in complex and ambiguous situations. The IGT was designed to this end by mimicking core requirements of real-world decision making: Response options are probabilistically associated with gains and losses, and participants have to learn through experience sampling which options best serve long-term goals (Bechara, Damasio, & Damasio, 2000; Bechara et al., 1994). Today, the use of the IGT is no longer confined to the study of patients with ventro-medial prefrontal cortex lesion. The task has been employed to study decision making in diverse populations (Bechara, 2007; Buelow & Suhr, 2009), including persons with other focal brain lesions,

[1]Ulm University, Ulm, Germany
[2]Karlsruhe University of Education, Karlsruhe, Germany
[3]Ernst-Moritz-Arndt-Universität, Greifswald, Germany
[4]University of Zurich, Zurich, Switzerland

Corresponding Author:
Florian Schmitz, Institute of Psychology, Ulm University, Albert-Einstein-Allee 47, 89081 Ulm, Germany.
Email: florian.schmitz@uni-ulm.de

healthy older adults, individuals with substance addiction, pathological gambling, schizophrenia, obsessive-compulsive disorder, anorexia nervosa, obesity, chronic pain, ADHD, aggression disorders, affective disorders, and Huntington's disease (see Buelow & Suhr, 2009, for a review). Additionally, the role of the IGT is no longer confined to be an experimental paradigm for basic research. With the introduction of a test manual (Bechara, 2007), it is suggested that the IGT is suited for applied assessment purposes.

Procedure and Scoring of the IGT

The IGT is a card-selection paradigm in which participants draw cards from four decks (A–D) that are associated with monetary gains and, occasionally, with losses. Deck expectancies are initially unknown and have to be learned from experience (Hertwig & Erev, 2009; Mata, Josef, Samanez-Larkin, & Hertwig, 2016). In the classical version of the IGT (Bechara et al., 1994), selecting a card from Decks A or B results in a fixed gain of $100, whereas choosing a card from Decks C and D results in a fixed gain of $50. However, occasionally, card selections result in an additional loss that can outweigh the gain. The proportion of loss trials varies across decks: Decks A and C comprise losses in 5 out of 10 selections, whereas Decks B and D comprise losses in only 1 out of 10 selections. Additionally, losses differ in magnitude. Consequently, the four decks have different long-term expectancies: In the long run, Decks A and B (with the constant high gain) will lead to an average net loss of $ −25, whereas Decks C and D (with the constant low gain) will lead to an average net gain of $25. Participants are instructed that they can choose cards from any deck to maximize their long-term net gain.

Different scores, reflecting different aspects of task performance, can be computed for the IGT. The most important performance indicator is the total amount of money won across all trials, or the proportion of selections from winning decks. Additionally, the learning of contingencies across trials can be investigated by inspecting performance improvements across trials. To this end, the available trials are usually split post hoc into blocks, and the proportion of selections from advantageous decks is compared across these blocks. Normative performance data can be obtained in the IGT manual for the total number of trials as well as for five successive blocks of trials (cf. Bechara, 2007).

Psychometric Characteristics

The IGT was originally developed for comparison between patient groups, and only a few studies, some of which with an only moderate sample size, addressed the psychometric characteristics of the task from an individual differences perspective. Some previous studies report satisfactory retest reliability within one session in the magnitude of $r = .60$ (Lejuez et al., 2003, based on $N = 60$ undergraduate students), and 2-week retest reliabilities within the range of $r = .35$ to $r = .65$ (Xu, Korczykowski, Zhu, & Rao, 2013, based on $N = 40$ students). It is more difficult to evaluate the validity of the task, as this requires making explicit what the task purports to measure and testing its correlations with independent measures of these constructs. This endeavor is demanding for the IGT as the task was conceived as a "multitrait task" (Gansler, Jerram, Vannorsdall, & Schretlen, 2011), because a number of factors may contribute to decision-making deficits. This comprises both cognitive impairments and risk-taking personality traits, and these two groups of variables may affect card selections in different phases. It was argued that learning is relevant in an earlier *exploration* phase, while preferential risk taking comes into play at a later *exploitation* phase when deck-contingencies have been sufficiently discovered (Brand, Recknor, Grabenhorst, & Bechara, 2007).

Performance in learning tasks can be expected to be related with working memory capacity (WMC) and fluid intelligence. Confirming these predictions, moderate relations of IGT performance with cognitive functioning were reported in a number of studies (Beitz, Salthouse, & Davis, 2014; Cauffman et al., 2010; Toplak, Sorge, Benoit, West, & Stanovich, 2010). However, findings concerning relations with risk taking or impulsive personality traits are mixed, either supporting at least some correlations (e.g., with lack of premeditation; Zermatten, Van der Linden, d 'Acremont, Jermann, & Bechara, 2005) or disconfirming relations with any of the scales (Bayard, Raffard, & Gely-Nargeot, 2011). Some research supports the predicted relations with behavioral activation and behavioral inhibition as broad personality factors (van Honk, Hermans, Putman, Montagne, & Schutter, 2002), whereas other research supports only relations with some of the behavioral activation subscales (e.g., "drive" in Buelow & Suhr, 2013; "reward responsiveness" in Franken & Muris, 2005; "fun seeking" in Suhr & Tsanadis, 2007).

Psychometric Challenges of the IGT

Some of the documented psychometric challenges of the IGT may result from conceptual and design features (Brand et al., 2007; Buelow & Suhr, 2009; Dunn, Dalgleish, & Lawrence, 2006; Horstmann, Villringer, & Neumann, 2012; Steingroever, Wetzels, Horstmann, Neumann, & Wagenmakers, 2012; Tomb, Hauser, Deldin, & Caramazza, 2002), briefly reviewed next.

Task Difficulty. The IGT is very difficult, such that many healthy young participants perform at floor level (Steingroever et al., 2012). Therefore, poor IGT performance per se is not indicative of impaired decision-making abilities.

Table 1. Characteristics of the Iowa Gambling Task (IGT) and Berlin Gambling Task (BGT).

	IGT Decks				BGT Decks			
	A	B	C	D	A	B	C	D
	Bad Deck, Frequ. Losses	Bad Deck, Infr. Losses	Good Deck, Frequ. Losses	Good Deck, Infr. Losses	Bad Deck, Low Variab.	Bad Deck, High Variab.	Good Deck, High Variab.	Good Deck, Low Variab.
Const. gain/trial	100	100	50	50	—	—	—	—
% Losses	50	10	40	10	60	60	40	40
Expectancy (*M*)	−25	−25	25	25	−40	−40	40	40
Variability (*SD*)	142	395	29	79	150	301	299	151
Skewness	−0.46	−3.16	−0.54	−3.16	0.31	0.50	−0.27	0.01
Kurtosis	−1.65	10.00	−1.39	10.00	−1.28	−0.91	−1.09	−0.41

Note. Const. gain/trial = gain in each trial from which loss is subtracted in case of loss trials; Expectancy = mean gain after 10 card selections; frequ./infr. losses = frequent/infrequent losses (corresponding with % losses of the respective decks).

Confounding of Expectancies With Skew of the Outcome Distributions. Decks do not only differ in their expectancy but also in their outcome variability and the skew of the outcome distribution. These skewed outcome distributions are responsible for the weak correspondence of proportions of winning decks with deck expectancy across decks (Spearman's ρ = .24). Even university students appear to select cards based on gain frequency, thereby ignoring the actual magnitude of the gain and loss experiences (Horstmann et al., 2012; Steingroever et al., 2012).

Lack of Separation of Exploration and Exploitation Phase. It was argued that at some point participants switch from an exploration to an exploitation mode during card selection (Bechara et al., 1994; Brand et al., 2007). Individuals may differ in when they switch strategy, and there may be no clear transition point between both phases (Steingroever et al., 2012). Consequently, a card selections at any particular trial may be governed by completely different processes across participants.

Changing Expectancies Across Card Selections. Participants can only estimate deck quality based on their previous experiences (Colombetti, 2008). The large variability in the outcome distributions results in considerable fluctuations in deck expectancies over time. As the order of selected cards will differ between participants, the experiences they can use to infer the deck expectancies will differ, too. As a consequence, task difficulty (i.e., to identify the most promising deck at a particular trial) may vary considerably across participants, introducing measurement noise.

A Possible Solution: The Berlin Gambling Task

Some of the problems with the IGT could be alleviated by generating more appropriate outcome distributions while keeping the general task requirements comparable with that of the original IGT (i.e., learning the quality of response options from their probabilistic outcomes). We sought to do this by developing a task variant that avoids some of the discussed problems, the Berlin Gambling Task (BGT).

One aim of the BGT was to dissociate possible effects of the quality of decision making from the riskiness of decision making. To this end, net outcomes (i.e., long-term expectancies) and variability of the decks vary orthogonally (see Table 1), yielding two orthogonal scores, one reflecting the quality of decisions by the degree to which a person preferred winning over losing decks, and the other reflecting the riskiness of decisions by the degree to which a person preferred high-variability over low-variability decks. Another aim of the BGT was to reduce the difficulty of the gambling task to a more diagnostic level. Hence, gains and losses were integrated into one net outcome per trial to reduce the need for calculations. As net outcomes were normally distributed around the deck means, the proportions of gains corresponded with the expectancies of the decks (see Table 1, right half). Additionally, heavy fluctuations of experienced outcomes were avoided in the precompiled lists of deck outcomes.

Finally, we sought to reduce the challenges associated with individual differences in the order of card selections. Specifically, both tasks started with a learning phase of 40 trials in which the computer chose cards in a precompiled sequence. This guaranteed that all participants saw an identical sample of card selections that is representative of the deck contingencies prior to taking their first choice. Previous studies have shown that 40 trials suffice to gain sufficient insight into deck contingencies (Bowman, Evans, & Turnbull, 2005; Brand et al., 2007; Cella, Dymond, Cooper, & Turnbull, 2007; Maia & McClelland, 2004). These 40 trials were considered to be a guided exploration phase. The next 60 trials allowed exploitation, and participants could select from the four decks on their own.

Study 1: A Comparison of the Iowa Gambling Task and the Berlin Gambling Task

Aims of Study 1

Study 1 aimed to investigate the psychometric properties of both task variants. We tested whether the procedural modifications in the BGT reduced difficulty in the BGT and whether these modifications improved its psychometric characteristics. Building on the prediction that both learning and risk-taking play a role in the gambling tasks (Bechara et al., 1994; Brand et al., 2007), we tested relationships with WMC, which was previously shown to be a strong predictor of learning rate (Lewandowsky, 2011; Tamez, Myerson, & Hale, 2012). Additionally, we investigated relations with a set of broad impulsive traits and self-reported risk-taking in various real-life domains.

Method

Sample and Procedure. The gambling tasks were administered as part of a larger research project on cognitive conflicts and WMC (Wilhelm, Hildebrandt, & Oberauer, 2013). This is the first study addressing the gambling tasks. All data from the first wave of data collection were used in the current analyses, comprising $N = 220$ research volunteers (122 female). Participants were recruited by flyers and information in local print media. Their mean age was 27.4 years ($SD = 5.0$; range = 18-35 years), and they had diverse professions, ranging from skilled workers to academics. Each gambling task was administered twice as part of a larger computerized battery comprising also measures of WMC and self-report questionnaires. Data were collected in groups of up to 10 participants on two consecutive days for each person. On each day, there were two 1.5 hour sessions with eight and nine short experimental tasks in a fixed order, separated by a long break. All gambling tasks were presented on the second day. All participants signed informed consent prior to participation. They received 40 Euros as a fixed compensation. Additionally, they could win more money depending on their gains in the gambling tasks (the value of their virtual account was divided by 1,000 and remunerated in Euros).

Iowa Gambling Task. Four piles of cards were displayed on the monitor, front sides facing down. On selection by mouse-click, the gain and an occasional loss of the selected card was displayed below the deck. All participants started with a virtual account of 500 Euros, and the current state of their account was displayed above the card decks in the center of the screen. There was also a counter that informed the participants how many cards they had selected so far. The current version of the gambling task comprised a guided exploration phase of 40 trials and an active card-selection phase of 60 trials. In the exploration phase, participants could only press a "continue" button, and the program selected a card for them. Instructions for the exploration phase emphasized that participants should pay attention to the outcomes in order to detect which were the better and which were the worse card decks. The sequence of card selections strictly followed an ABCD scheme in the guided exploration phase. The gain and occasional loss outcomes followed the scheme given by Bechara et al. (1994). At the beginning of the active card selection phase, participants' virtual accounts were reset to 500 Euros. Instructions for the active selection phase emphasized to select cards from profitable decks in order to gain money, and to avoid losing money. The actual positions of the four decks in the IGT were randomly determined when the paradigm was started. Consequently, the positions of good and bad decks were likely to change from the first to the second administration of the task, and participants were explicitly informed about this fact. For the correlational analyses, we computed one outcome score for the IGT: the proportion of card choices from winning decks (i.e., decks with positive expectancy).

Berlin Gambling Task. The procedure of the BGT was largely comparable with that of the classical IGT, with the major difference that outcome sequences were more balanced to avoid large fluctuations in deck experiences across card selections. Differently from the IGT, there was only one outcome information displayed beneath the selected card deck. The outcomes were drawn from normal distributions with the means and standard deviations given in Table 1. There was also a guided exploration phase of 40 trials, followed by an active card-selection phase of 60 trials. The positions of card decks were counterbalanced and randomly determined at the beginning of each task administration. We computed two orthogonal scores for the BGT: The first reflected the proportion of card choices from winning decks, and the second reflected the proportion of card choices from risky decks (i.e., decks with high variability).

Estimates of Deck Benefit. Participants were asked twice in each gambling task to enter their estimates of the net expectancies of each deck, once after the exploration phase and once after completing the selection phase. Estimates do not need to be accurate in absolute terms in order to select cards from winning decks. It is sufficient if winning decks are estimated clearly more beneficial than losing decks. Therefore, estimates of the four decks were standardized (Z transformed) within participant and measurement occasion. We calculated a score indexing the relative preference for winning decks over losing decks (i.e., positive values indicated that winning decks were rated as more beneficial) (Formula 1).

$$\textit{Preference for winning decks}: \quad \textit{PrefWin} = \Big(\big(Z(C)+Z(D)\big)-\big(Z(A)+Z(B)\big)\Big) \tag{1}$$

From the estimates of card expectancies in the BGT we could calculate a second, orthogonal score indexing the preference for variable over less variable decks (i.e., positive values indicated that variable decks were rated as more beneficial) (Formula 2).

$$\textit{Preference for risky decks}: \quad \textit{PrefRisk} = \Big(\big(Z(B)+Z(C)\big)-\big(Z(A)+Z(D)\big)\Big) \tag{2}$$

Working Memory Capacity. The battery comprised a set of typical WMC tasks, including (1) binding, (2) updating, and (3) simultaneous storage and processing tasks. All task types were administered with three different sets of materials (e.g., letters, numbers, figures, etc.). Partial credit scores (Conway et al., 2005) were computed for each task, then they were standardized and aggregated into composite scores for each task type for the psychometric analyses.

Impulsive Personality Traits and Risk Taking. Most contemporary taxonomies conceptualize impulsivity as a multifaceted construct. We assessed four broad factors identified by Whiteside and Lynam (2001), namely *urgency* (U), *lack of premeditation* (P), *lack of perseverance* (P), and *sensation seeking* (S). In the present study, we used the German adaptation of the UPPS (Keye, Wilhelm, & Oberauer, 2009). Additionally, we included the Domain-Specific Risk-Taking scales (DOSPERT; Blais & Weber, 2006; German adaptation by Johnson, Wilke, & Weber, 2004), assessing self-reported risky activities and behaviors across five general domains.

Data Analysis. The data were analyzed with R (R Core Team, 2016). All gambling data were complete, but WMC data from three participants were lost and seven self-report questionnaires were missing. These data were imputed with the R package Amelia II (Honaker, King, & Blackwell, 2011), which uses a bootstrapping-based algorithm for multiple imputation, and results were found to be highly comparable across 10 imputations. The psychometric analyses were conducted with the R package psych (Revelle, 2015). Moderated growth models across blocks of card selections were estimated with the generalized estimating equations solver geepack for R (Højsgaard, Halekoh, & Yan, 2006). Variables were centered and standardized to this end.

Results

To start with, we investigated card selections across blocks in the IGT and in the BGT (see Figure 1). In the IGT, participants showed a strong preference for the decks with infrequent losses (unfilled points) relative to decks with more frequent losses (filled points) irrespective of the long-term expectancies of the respective decks. In contrast, in the BGT, participants preferred the decks with a positive expectancy (black icons) over the decks with negative expectancy (red icons) right from the beginning. Additionally, the preference to select advantageous decks and to avoid disadvantageous decks was more pronounced for the decks with low variability, relative to the decks with highly variable outcomes.

A similar picture arose for the ratings of deck benefit. The proportions of participants rating advantageous decks as more beneficial than disadvantageous decks were 47% and 61% in the first and in the second administration of the IGT, respectively. In the BGT, these rates were 64% and 66%, respectively. Only a minority of participants (32% and 35% in the first and in the second BGT administration, respectively) estimated risky (variable) decks as more beneficial.

Next, we tested the reliability of the scores derived from the gambling tasks. As card selections in sequential learning tasks are not independent, estimates of internal consistency calculated from individual card selections (e.g., split-half reliability or Cronbach's alpha) overestimate reliability. Therefore, we used the correlation of task scores across the two task administrations within the same session (separated by ca. 1 hour) as an estimate of retest reliability. As can be seen in the diagonal of Table 2, retest reliability was only moderate. In fact, higher correlations were observed between participants' estimates of card expectancies and their actual card selections in the same task. Cross-task correlations between IGT and BGT were very moderate for the expectancy estimates as well as for actual selections. Additionally, the BGT scores capturing task performance (preference for winning decks) and riskiness (preference for variable decks) were virtually independent.

To investigate validity, standardized gambling scores were aggregated across measurement occasions to increase their reliability. Additionally, we computed scores for the first and for the second half of the selection phase to test differential correlation patterns for an *exploration* and an *exploitation* phase. Moreover, we report correlations for the proportions of selections from each deck to trace the sources of possible relations (see Table 3).

The gambling performance scores were moderately correlated with all indicators of WMC. A closer look at the individual deck selections suggests that this correlation was primarily driven by more able participants avoiding the losing Deck A, and less by them selecting the winning Deck D. Selections from the other decks were not consistently correlated with WMC. If anything, selections from winning decks tended to be more highly related with WMC in the second half of the selection phase compared with the first half. There was no relation of the gambling risk-taking

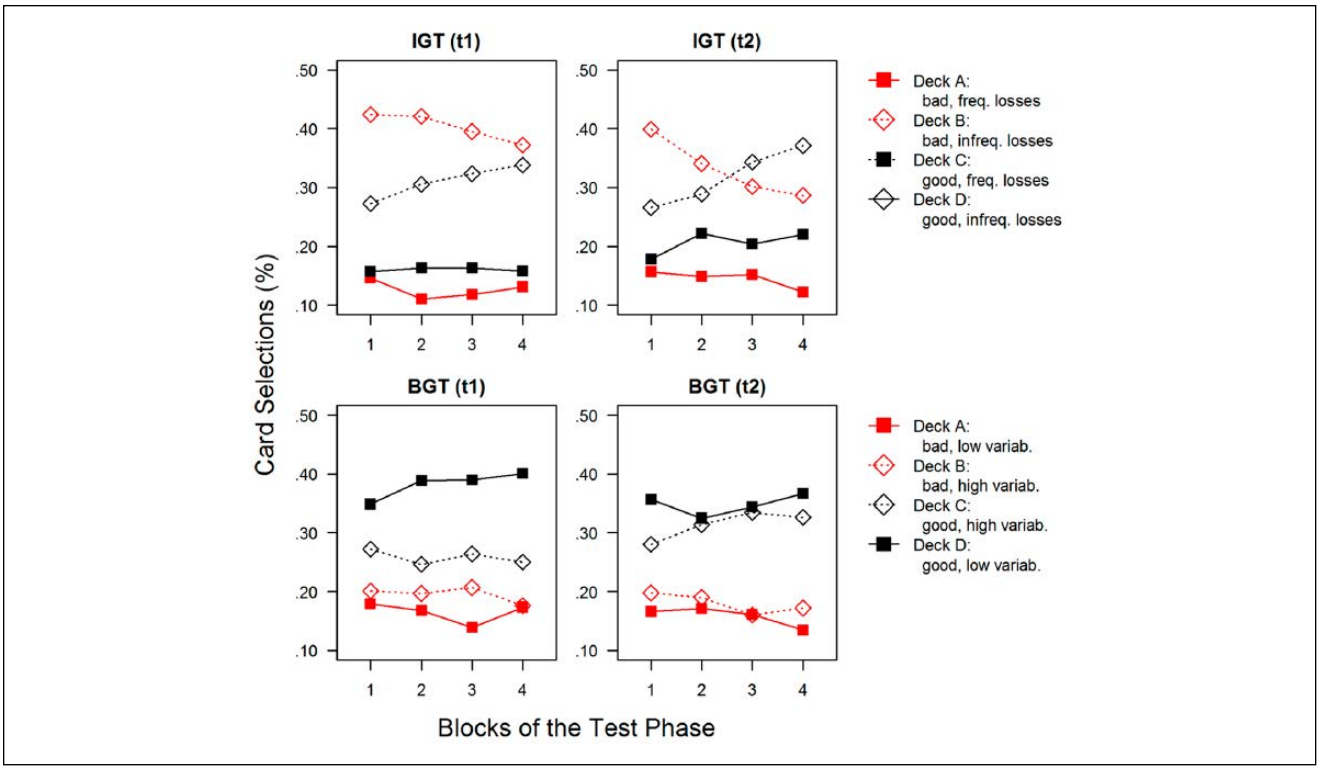

Figure 1. Study 1: Card selections in the Iowa Gambling Task (IGT) and the Berlin Gambling task (BGT).
Note. Trials of the test phase were split into four blocks, and frequencies of card selections are displayed separately for Decks A to D: Black and red lines indicate decks with a positive and negative expectancy, respectively. In the IGT, broken and continuous lines stand for decks with infrequent and frequent losses, respectively. In the BGT, broken and continuous lines stand for decks with high and low outcome variability, respectively. (t1) and (t2) indicate the first and second measurement occasion. The chance level is .25.

Table 2. Study 1: Intercorrelations of Iowa Gambling Task (IGT) and Berlin Gambling Task (BGT) Scores.

	IGT: Winning Decks		BGT: Winning Decks		BGT: Risky Decks	
	Estimates	Selections	Estimates	Selections	Estimates	Selections
IGT: Winning decks						
• Estimates	(.28**)					
• Selections	.49**	(.36**)				
BGT: Winning decks						
• Estimates	.10	.10	(.15*)			
• Selections	.14*	.17*	.40**	(.44**)		
BGT: Risky decks						
• Estimates	.09	.17**	.07	−07	(.15*)	
• Selections	.02	.24**	.09	−.08	.54**	(.10)

Note. Winning decks = proportion of cards selected from decks with a positive expectancy; Risky decks = proportion of cards selected from decks with high variability in outcomes; Estimates = participant's relative predicted outcomes; Selections = proportion of cards actually selected. All scores are compound scores averaging standardized scores of the first and second task administration. Estimates after the exploration phase and after selection phase were additionally averaged for simplicity. Retest reliability is displayed in the diagonal.
*$p \leq .05$. **$p \leq .01$.

score (i.e., preference for variable decks) from the BGT with any of the WMC indicators. Finally, there were no substantial relations of gambling scores with impulsive personality traits or of self-reported risk taking.

Given that WMC was found to be correlated with selections from winning decks, the question arises whether participants with high WMC already started at a higher level after the guided exploration phase, for comparison between

Table 3. Study 1: Correlations of IGT/BGT Scores With Working Memory Capacity, Impulsivity, and Risk Taking.

	Working Memory Capacity			Impulsivity				Risk Taking		
	BIN	UPD	CSP	URG	PRE	PER	SEN	BEHV	PERC	UTIL
Working memory capacity										
• Binding (BIN)	(.75)									
• Updating (UPD)	.70*	(.79)								
• Complex span (CSP)	.68*	.76*	(.79)							
Impulsivity										
• Urgency (URG)	−.16*	−.29*	−.26*	(.64)						
• Premeditation (−) (PRE)	.00	−.01	−.04	.26*	(.70)					
• Perseverence (−) (PER)	.11	−.06	.00	.35*	.16*	(.81)				
• Sensation seeking (SEN)	−.06	−.11	−.08	.25*	.11	.20*	(.73)			
Risk taking										
• Behavior (BEHV)	−.03	−.10	−.06	.27*	.16*	.17*	.59*	(.63)		
• Perception (−) (PERC)	.06	.02	.01	.09	.16*	.07	.23*	.49*	(.62)	
• Utility UTIL	.13*	.12	.07	.19*	.07	.17*	.34*	.57*	.33*	(.52)
Gambling tasks (IGT/BGT)										
• Winning decks										
o Estimates	.17*/.30*	.16*/.27*	.15*/.32*	−.12/−.12	−.13*/.02	.03/.10	.10/.10	.05/.12	.08/.14*	.10/.15*
o Selections total	.18*/.19*	.17*/.20*	.16*/.22*	−.10/−.11	−.12/−.03	−.03/.01	.02/.03	.02/.09	−.01/.12	.10/.20*
o Selections first half	.10/.17*	.10/.15*	.10/.16*	−.07/−.04	−.10/−.02	−.01/.07	.02/.05	.00/.09	−.03/.09	.08/.19*
o Selections second half	.21*/.18*	.21*/.21*	.18*/.24*	−.10/−.15*	−.11/−.03	−.05/−.05	.01/.00	.03/.08	.00/.12	.10/.17*
• Risky decks										
o Estimates	—/.02	—/−.12	—/−.12	—/.08	—/.10	—/.06	—/.05	—/.06	—/.16*	—/.08
o Selections total	—/.08	—/−.04	—/.01	—/.06	—/.03	—/.14*	—/.06	—/.07	—/.09	—/.13
o Selections first half	—/.02	—/−.04	—/.00	—/.04	—/.06	—/.13	—/.09	—/.05	—/.09	—/.09
o Selections second half	—/.12	—/−.03	—/.02	—/.06	—/−.01	—/.12	—/.03	—/.07	—/.08	—/.14*
• Specific decks										
o Deck A	−.32*/−.21*	−.33*/−.19*	−.34*/−.21*	.11/.13	.05/.06	−.05/−.06	.06/−.02	.03/−.04	−.03/−.09	−.14*/−.16*
o Deck B	−.03/−.10	−.02/−.12	.00/−.14*	.05/.04	.11/−.01	.05/.04	−.04/−.02	−.03/−.11	.03/−.10	−.04/−.15*
o Deck C	.04/.15*	−.06/.04	−.02/.10	.04/.03	−.20*/.04	.03/.12	.14*/.08	.21*/.14*	−.03/.16*	.06/.22*
o Deck D	.15*/.04	.21*/.14*	.17*/.10	−.14*/−.13	.03/−.05	−.06/−.10	−.08/−.05	−.13*/−.04	.01/−.04	.06/−.03

Note. IGT = Iowa Gambling Task; BGT = Berlin Gambling Task; Winning decks = proportion of cards selected from decks with a positive expectancy; Risky decks = proportion of cards selected from decks with high variability in outcomes; Estimates = participant's relative predicted outcomes; Selections = proportion of cards actually selected. Some variables were inverted (−), so that high scores in all criterion variables indicate impulsive/risk-taking tendencies. Their reliabilities (McDonald's omega) are displayed in the diagonal. All IGT/BGT scores are compound scores, averaging standardized scores of the first and second task administration. Their reliabilities and correlations are given in Table 2.
*$p \leq .05$.

patient groups, or whether they had steeper learning rates in the active selection phase, or both (see Figure 2). This was tested by conducting mixed-effects regression analyses, which allows investigating simultaneously effects in the intercept and in the slope. The probability of selecting a winning deck in each of the four blocks of the gambling tasks was entered as a dependent variable, collapsed across both measurement occasions. Predictors were the ordinal position of the block (1 to 4), a WMC composite score based on all standardized WMC indicators, and the interaction of these two variables. Residuals of the dependent variable at the four blocks were allowed to be correlated across participants, following a first-order autoregressive structure, because differential learning rates would lead to higher correlations of residuals of more proximal measurement points compared with residuals at more distant measurement points. The moderated regression for the IGT yielded a significant effect of block ($\beta = .17$, $p < .001$), indicating a general increase in task performance across blocks. There was a significant effect of WMC ($\beta = .14$, $p < .001$), revealing higher performance when WMC was high. Finally, there was an interaction of both factors ($\beta = .07$, $p < .01$), indicating a steeper increase across blocks when WMC was high. An analogous regression for the BGT yielded similar results. There was an effect of block ($\beta = .08$, $p < .001$), an effect of WMC ($\beta = .18$, $p < .001$), and an interaction effect ($\beta < .04$, $p < .05$). The interaction of WMC with block shows that participants with higher WMC learn more rapidly during the active selection phase.

Discussion

Both gambling tasks were difficult, the IGT even more so than the BGT. In part, this may reflect that in the IGT participants were systematically misled by gain frequencies of the decks that do not correspond to long-term expectancies. As expectancy and gain frequency are confounded in the BGT, it could be speculated that participants' preference for favorable decks may actually be driven by perceived gain frequency rather than expectancy (Horstmann et al., 2012).

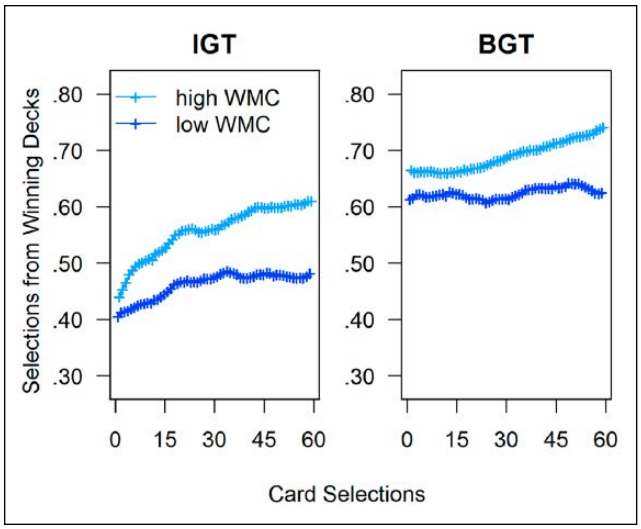

Figure 2. Study 1: Proportion of card selections from winning decks in the Iowa Gambling Task (IGT) and the Berlin Gambling task (BGT).

Note. Task performance is displayed for participants with high and with low working memory capacity (WMC) (i.e., belonging to the upper and lower third of the WMC distribution). The "+" signs indicate the actual frequency of choosing a winning deck at the respective ordinal position of the card selection, smoothed by averaging across a shifting window of 15 trials. The chance level of selecting a winning deck is .50.

However, the consistent finding that decks with larger variability were selected with intermediate frequency suggests that outcomes were not simply recoded as "gain" or "loss." This variability effect implies that the outcome quantities are somehow processed in the BGT.

The observation of correlations between estimated expectancies and actual card selections suggests that people have a subjective impression of the deck qualities and select cards in accordance with their belief. At the same time, the generally poor accuracy of these predictions and the low retest reliability suggests that deck estimates are driven by ad hoc beliefs in response to the sequence of events in each test occasion, rather than reflecting stable individual differences in the ability to learn probabilistic outcomes.

Correlations of the gambling scores with variables of the nomological network of risk taking were low, inconsistent, and mostly nonsignificant. Only WMC was moderately related with performance in the gambling tasks, confirming previous findings (Beitz et al., 2014; Cauffman et al., 2010; Toplak et al., 2010). Participants with high WMC started already at a higher performance level after the guided exploration phase, and they further improved more strongly across active card selections. Apparently, participants with higher WMC avoided card selections from the losing Deck A, and to a lesser extent they selected more cards from the winning Deck D. The expectancies of the other decks may be more difficult to learn, either because their outcomes are misleading (e.g., Deck B in the IGT, yielding only net gains for a long time) or because of their larger outcome variability (e.g., Decks B and C in the BGT). This supports the view that high task difficulty contributes to the gambling tasks' problematic validity.

Study 2: Testing More Specific Correlates in a Sample at Elevated Risk

Aim of Study 2

One striking result of Study 1 was the apparent absence of meaningful correlations between risk-taking personality variables and gambling scores. Possible explanations for this include that risk taking may have been generally low in that sample, or that personality measures were included that are simply not closely related with gambling behavior. Study 2 was designed to test these explanations.

In Study 2, we tested adolescents who we assumed to be more likely to engage in risk-taking behavior, because this developmental period is characterized by increased reward seeking and sensation seeking, while inhibition and self-control have not yet matured (Casey, Getz, & Galvan, 2008; Harden & Tucker-Drob, 2011; Steinberg, 2010). In fact, previous research using a modified version of the IGT that allows estimating approach and avoidance tendencies separately (Cauffman et al., 2010) showed that adolescents are more approach oriented and less avoidance oriented than adults. This, in turn, would result in more risk-taking behavior as a consequence.

It is assumed that learning in gambling tasks is governed by gain and loss experiences. The personality theory that appears to be most closely related with such experience learning is Reinforcement Sensitivity Theory (RST; Corr, 2008; Gray & McNaughton, 2000), and some previous research suggests RST dimensions are related to gambling behavior (Suhr & Tsanadis, 2007; van Honk et al., 2002). Accordingly, sensitivity to reward and sensitivity to punishment (Torrubia, Avila, & Caseras, 2008; Torrubia, Avila, Molto, & Caseras, 2001) were assessed in the current study. Additionally, a measure of sensation seeking (Zuckerman, 1996, 2007) was included, because high levels of *sensations seeking* may motivate participants to select cards from more arousing highly variable decks (see Tomb et al., 2002). In particular, the *disinhibition* subscale was shown to predict risk-taking behavior, such as gambling deviance and substance consumption (Greene, Krcmar, Walters, Rubin, & Hale, 2000; Roberti, 2004).

As WMC paradigms were extensively investigated in the previous study, this time, a standard measure of fluid intelligence and additionally scholastic achievement were included. Fluid intelligence was previously shown to contribute to complex problem solving (Wittmann & Suess, 1999). Scholastic achievement may be conceived as a proxy

of (cognitive) success in the real world. As a gambling paradigm, we decided to use only the BGT, because it avoids some of the conceptual problems of the IGT. Additionally, it yields independent scores for the quality and for the riskiness of card selections.

Method

Sample and Procedure. Data were collected at secondary schools through class-wise group testing. For the present analyses, all students in the critical age range of 14 to 16 years were selected from a larger student sample, yielding a total of 389 students (207 female). Academic track schools were oversampled (n = 272) relative to the intermediate-track schools (n = 40) and vocational-track schools (n = 77). Written informed consent was obtained from both parents and participants about a week prior to administration of the study. Data were collected through class-wise group testing. The multivariate test battery comprising the BGT as well as cognitive and personality measures is described in the following.

Cognitive Ability. The capacity (K) scale from the Berlin Intelligence Structure test (Jäger, Süß, & Beauducel, 1997; see Wilhelm & Schulze, 2002, for English task descriptions) was administered as a paper-and-pencil test of fluid intelligence. The K scale comprises six different subtests and takes about 30 minutes to complete. Additionally, we assessed scholastic achievement by averaging the grades obtained in that last (half-year) report in seven subjects taught at all schools according to the curriculum (i.e., German, English, Mathematics, Biology, Chemistry, History, and Religion/Ethics).

Self-Report Measures. Individual differences in RST dimensions were assessed with the *sensitivity to reward* and *sensitivity to punishment* questionnaire (SPSRQ, Torrubia et al., 2001; Torrubia et al., 2008). A short version of the questionnaire was created based on our own previous research with a German adaptation by selecting the 12 best-suited items for each scale according to the simple structure of loadings and by maximizing consistency of the scales. Sensation seeking was assessed with an analogously shortened version of Zuckerman's (1996, 2007) *Sensation Seeking Scales V* (German adaptation, Beauducel, Strobel, & Brocke, 2003), comprising 12 items in total, 3 for each of the 4 subfactors *thrill* and *adventure seeking*, *experience seeking*, *boredom susceptibility*, and *disinhibition*. General deviance was assessed by means of a list of 15 behavioral descriptors of minor deviant behavior, such as "I used public transportation without buying a ticket," "I cheated in an exam," "I spread rumors about others," and "I downloaded music or films that are copyright protected." As frequency ratings of the items were highly correlated and loaded on one general factor, they were averaged into one score indexing general deviance. Finally, participants reported their consumption of alcoholic beverages and cigarettes on an average day. All reports were standardized and averaged into a composite score of substance consumption. Note that these substances are relatively easily available for adolescents, however, their consumption is prohibited by law in the investigated age group.

Results

Study 2 findings were highly comparable with those obtained in Study 1. On average, the adolescents selected somewhat less cards from winning decks (M = 0.57, SD = 0.12) than the adults (M = 0.64, SD = 0.19) in the first administration of the BGT in Study 1 (g_{Hedges} = .47 for the group difference). There was a moderate correlation between the ratings of deck expectancies and the actual card selections (r = .27 for the correlation of the winning-deck estimate and the winning-deck choice score, and r = .32 for the correlation of risky-deck estimate and risky-deck choice scores).

Correlations of gambling scores to criterion variables are provided in Table 4. Cognitive ability was moderately related with a preference for the winning decks, both in terms of deck estimates as well as actual selections. Again, this reflected that participants with high ability avoided the losing Deck A and preferred the winning Deck D. Again, selections from decks with high outcome variability (B and C) were not systematically related with cognitive ability. Personality traits were not related with card selections. The only exception here was the narrow cluster of self-reports of deviant behavior, comprising disinhibited behavior, (minor) criminal offences, and substance consumption, which revealed consistent but arguably low correlations in the predicted direction.

Discussion

Confirming predictions, task performance of the adolescents was somewhat lower compared with that of adults on average. However, results concerning validity of the scores were comparable with those obtained in Study 1: Cognitive ability was weakly related with a preference for wining decks. In contrast, there were virtually no relations with personality, and relations with self-reported deviant behavior were so weak that meaningful predictions are not warranted.

General Discussion

This study was conducted to test the psychometric suitability of the IGT as an individual differences measure of risk taking. Additionally, the BGT was developed and tested as a task variant that seeks to reduce some of the conceptual problems of the original task version.

Table 4. Study 2: Correlations of Berlin Gambling Task (BGT) Scores With Ability, Personality, and Risk Taking.

	Ability		Personality				Risk Taking	
	GF	ACH	PUN	REW	SEN	DIS	DEV	SUB
Correlates								
• Fluid intelligence (GF)	(.78)							
• Scholastic achievement (ACH)	.43**	(.83)						
• Punishment sensitivity (PUN)	−.01	.01	(.68)					
• Reward sensitivity (REW)	.06	−.06	−.03	(.70)				
• Sensation seeking (Total) (SEN)	.04	−.11*	−.35**	.31**	(.63)			
• Disinhibition (DIS)	−.09	−.17**	−.23**	.34**	.70**	(.66)		
• General deviance (DEV)	−.05	−.32**	−.19**	.44**	.36**	.46**	(.81)	
• Substance consumption (SUB)	−.22**	−.17**	−.07	.16**	.25**	.31**	.26**	(.66)
Berlin Gambling Task								
• Winning decks								
o Estimates	.10*	.17**	−.01	−.05	−.07	−.11*	−.12*	−.12*
o Selections total	.23**	.18**	.07	−.01	−.08	−.16**	−.04	−.12*
o Selections first half	.16**	.15**	.11*	−.07	−.15**	−.21**	−.07	−.10*
o Selections second half	.24**	.17**	.01	.03	−.01	−.08	−.01	−.11*
• Risky decks								
o Estimates	.02	−.03	.06	.03	−.10*	−.09	−.01	.09
o Selections total	−.06	−.14**	.01	.10	.03	.04	.12*	.10
o Selections first half	.01	−.07	.00	.05	.02	.02	.07	.02
o Selections second half	−.12*	−.18**	.02	.12*	.03	.05	.14**	.15**
• Specific decks								
o Deck A (bad deck, low variability)	−.23**	−.16**	−.06	−.06	.06	.11*	−.01	.06
o Deck B (bad deck, high variability)	−.09	−.10*	−.04	.09	.06	.12*	.07	.11*
o Deck C (good deck, high variability)	−.01	−.09	.04	.04	−.01	−.04	.08	.03
o Deck D (good deck, low variability)	.20**	.22**	.03	−.04	−.06	−.11*	−.10	−.12*

Note. Winning decks = proportion of cards selected from decks with a positive expectancy; Risky decks = proportion of cards selected from decks with high variability in outcomes; Estimates = participant's relative predicted outcomes; Selections = proportion of cards actually selected. Reliabilities (McDonald's omega) of the working memory and personality measures are displayed in the diagonal.
*$p \leq .05$. **$p \leq .01$.

Replicating previous findings with the IGT (Horstmann et al., 2012), participants of the general population with an above-average educational level selected cards from decks with infrequent losses, apparently ignoring the magnitude of the actual outcomes. This finding implies that low performance in the IGT alone should not be considered as pathological (Steingroever et al., 2012). As intended, task performance was better in the BGT.

Retest reliability was only moderate for the scores derived from both gambling tasks, even though both task administrations were separated by only 1 hour. Somewhat higher correlations were observed between estimates of deck benefit and actual card selections within each measurement occasion. This pattern suggests that actual card selections are driven by ad hoc beliefs of deck quality. However, the accuracy of participants' estimates of deck quality was generally poor. The low retest reliability additionally implies that any stable individual differences contributing to the learning of probabilistic outcomes could have only a moderate effect on perceived deck benefits and card selections at each task administration. A large part of this problem may result from individual differences in the sequences of card selections. A single, largely random, event can lead the entire learning and choice process into a different trajectory. Reliability estimates of the gambling scores were somehow lower in this study compared with those reported in previous research with undergraduate students and samples of moderate size (Lejuez et al., 2003; Xu et al., 2013). However, the other measures included in the current studies with sufficiently large and heterogeneous samples appeared to possess satisfactory reliabilities (see Tables 3 and 4), suggesting that data in the current studies were generally of satisfactory quality.

Gambling scores were found to be task specific, as shown by the fact that IGT–BGT cross-task correlations were considerably lower than the retest correlations within each task variant. This task specificity may have attenuated validity of the gambling tasks as general measures of risk-related decision making (Bechara, 2007; Bechara et al., 1994). Relations with WMC and fluid intelligence were weak, confirming previous

findings for the IGT (Beitz et al., 2014; Cauffman et al., 2010; Toplak et al., 2010). Somewhat stronger relations were expected given that WMC has been shown to be closely related to learning (Kaufman, DeYoung, Gray, Brown, & Mackintosh, 2009; Lewandowsky, 2011; Tamez et al., 2012), and the learning of probabilistic deck outcomes should contribute to preferential selection of winning cards. However, the generally poor task performance suggests that the validity is compromised by floor effects of task performance. Somewhat higher relations of cognitive ability with selections from winning decks in the second half compared with the first half of the selection phase suggest that far more trials may be needed to measure differential learning. Possibly, validity could be substantially improved by adding many trials (Buelow, Okdie, & Blaine, 2013; Xu et al., 2013). This could be a feasible way for future research. At the same time, this would further decrease test economy for applied assessment purposes.

None of the scores derived from the gambling tasks was consistently related with self-report variables reflecting risk taking. This was found for impulsive personality traits (i.e., nonreflective decision making), for reinforcement sensitivity (i.e., learning from reward and from punishment), for sensation seeking, and for reported risk taking in different domains. None of the investigated relations was consistent and large enough to indicate validity for the purpose of assessing individual differences. This lack of meaningful correlations was not an artifact of unreliable self-report data, as the self-reported criterion variables revealed a pattern of correlations with each other in congruence with theories of risk taking (Steinberg, 2010; Torrubia et al., 2008; Zuckerman & Kuhlman, 2000).

Taken together, both the IGT and the BGT showed insufficient reliability, and we obtained no evidence for their validity either. Therefore, neither task can be recommended for the assessment of individual differences. Validity of the gambling tasks needs to be discussed in light of findings that the relations between scores from laboratory paradigms and self-report measures are generally weak at best (Cyders & Coskunpinar, 2011; Friedman & Miyake, 2004; Keye, Wilhelm, Oberauer, & van Ravenzwaaij, 2009; Ortner, Proyer, & Kubinger, 2006; Stahl et al., 2014). However, compared with other laboratory paradigms (see below), one major issue with the gambling tasks is their low retest reliability after only 1 hour. In the absence of adequate retest reliability, the task does not qualify as a measure of a stable personality trait. It could be argued that laboratory tasks are generally more state sensitive (as compared with self-report scales that are supposed to assess traits; cf. Cyders & Coskunpinar, 2011). This would suggest that a higher level of aggregation across multiple task administrations is required to arrive at more stable trait-like estimates of individual differences. This may be an avenue for basic research, but would arguably reduce feasibility from a pragmatic, applied assessment perspective.

Naturally, specificity of the assessment methods will contribute to the dissociation of behavioral tasks and self-reports. Research in the related domain of impulsivity suggests that different self-report scales are highly correlated, while laboratory tasks purported to assess the same underlying factor are only weakly related with each other, and often they are virtually independent of self-report measures (e.g., Cyders & Coskunpinar, 2012; Keye, Wilhelm, Oberauer, & van Ravenzwaaij, 2009; Stahl et al., 2014). Characteristics of both the laboratory tasks and of the self-report instruments may contribute to this dissociation. On the side of the laboratory paradigms, task impurity, insufficient reliability, and task-specificity have been identified as challenges (Friedman & Miyake, 2004). However, self-report measures are also plagued by a plethora of validity issues; most important, the presuppositions that participants are able and willing to introspect and report their subjective assessments (Greenwald et al., 2002). Additionally, some of the investigated variables of risk taking, substance consumption, and deviant behavior might generally be difficult to predict (e.g., Simons, Dvorak, & Batien, 2008; Townsend, Flisher, Gilreath, & King, 2006). Still, the sound psychometric properties of self-report measures clearly indicate them as the preferred method for the assessment of individual differences.

Nevertheless, some laboratory tasks for investigating risky decision making were shown to reveal moderate relations of self-reported risk taking, for example, the balloon analogue risk task (Lauriola, Panno, Levin, & Lejuez, 2014; Lejuez et al., 2002), the angling risk task (Pleskac, 2008), and the game of dice (Brand et al., 2002). The relatively higher validity of these tasks compared to gambling paradigms possibly reflects the well-known distinction between description tasks and experience sampling tasks (Hertwig & Erev, 2009; Mata et al., 2016). In description tasks, the riskiness of the choice alternatives is directly communicated. For instance, the riskiness of a gamble can be easily computed in the game of dice. Similarly, the angling risk task has a condition in which the probability of a loss can be inferred from an intuitive visual display of fishes in a pond. The balloon analogue risk task was actually conceptualized as an experience sampling task, but the riskiness of pumping a balloon can be easily learned. Additionally, validity in terms of relations with risk-taking indicators could be further increased by making the riskiness of pumping even more explicit (Schmitz, Manske, Preckel, & Wilhelm, 2016). One feature shared by these tasks is their intuitive appeal and their low requirements on learning. Consequently, these laboratory paradigms seem to offer purer measures of preferential risk taking.

We believe that laboratory tasks have their importance in basic research, but more research is necessary to understand under which circumstances they qualify for assessment of individual differences. This should also include other

objective measures and relevant life-event data as validation criteria, thereby circumventing the problems of self-report data.

Conclusion

WMC and fluid intelligence were weakly positively related with learning rates in both gambling tasks. However, no consistent relations of a meaningful size were found with risk-taking personality traits and self-reported risk taking. Taken together, gambling tasks are not suggested for measuring individual differences in risk taking for assessment purposes.

Authors' Note

Olga Kunina-Habenicht is currently affiliated to Karlsruhe University of Education, Karlsruhe, Germany.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported in this article was supported by Grant WI2667-4 from the German Research Foundation (DFG) accorded to OW and KO.

References

Bayard, S., Raffard, S., & Gely-Nargeot, M.-C. (2011). Do facets of self-reported impulsivity predict decision-making under ambiguity and risk? Evidence from a community sample. *Psychiatry Research, 190*, 322-326.

Beauducel, A., Strobel, A., & Brocke, B. (2003). Psychometric properties and norms of a German version of the Sensation Seeking Scales, Form V. *Diagnostica, 49*(2), 61-72.

Bechara, A. (2007). *Iowa gambling task (IGT) professional manual*. Lutz, FL: Psychological Assessment Resources.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*(1-3), 7-15.

Bechara, A., & Damasio, H. (2002). Decision-making and addiction (part I): Impaired activation of somatic states in substance dependent individuals when pondering decisions with negative future consequences. *Neuropsychologia, 40*, 1675-1689.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex, 10*, 295-307.

Bechara, A., Dolan, S., & Hindes, A. (2002). Decision-making and addiction (part II): Myopia for the future or hypersensitivity to reward? *Neuropsychologia, 40*, 1690-1705.

Beitz, K. M., Salthouse, T. A., & Davis, H. P. (2014). Performance on the Iowa Gambling Task: From 5 to 89 years of age. *Journal of Experimental Psychology: General, 143*, 1677-1689.

Blais, A.-R., & Weber, E. U. (2006). A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making, 1*(1), 33-47.

Bowman, C. H., Evans, C. E., & Turnbull, O. H. (2005). Artificial time constraints on the Iowa Gambling Task: The effects on behavioural performance and subjective experience. *Brain and Cognition, 57*(1), 21-25.

Brand, M., Greco, R., Schuster, A., Kalbe, E., Fujiwara, E., Markowitsch, H. J., & Kessler, J. (2002). The game of dice: A new test for the assessment of risk-taking behavior. *Neurorehabilitation & Neural Repair, 16*, 142-143.

Brand, M., Recknor, E. C., Grabenhorst, F., & Bechara, A. (2007). Decisions under ambiguity and decisions under risk: Correlations with executive functions and comparisons of two different gambling tasks with implicit and explicit rules. *Journal of Clinical and Experimental Neuropsychology, 29*(1), 86-99.

Buelow, M. T., Okdie, B. M., & Blaine, A. L. (2013). Seeing the forest through the trees: Improving decision making on the Iowa Gambling Task by shifting focus from short- to long-term outcomes. *Frontiers in Psychology, 4*, ArtID 773.

Buelow, M. T., & Suhr, J. A. (2009). Construct validity of the Iowa Gambling Task. *Neuropsychology Review, 19*(1), 102-114.

Buelow, M. T., & Suhr, J. A. (2013). Personality characteristics and state mood influence individual deck selections on the Iowa Gambling Task. *Personality and Individual Differences, 54*, 593-597.

Casey, B., Getz, S., & Galvan, A. (2008). The adolescent brain. *Developmental Review, 28*(1), 62-77.

Cauffman, E., Shulman, E. P., Steinberg, L., Claus, E., Banich, M. T., Graham, S., & Woolard, J. (2010). Age differences in affective decision making as indexed by performance on the Iowa Gambling Task. *Developmental Psychology, 46*(1), 193-207.

Cella, M., Dymond, S., Cooper, A., & Turnbull, O. (2007). Effects of decision-phase time constraints on emotion-based learning in the Iowa Gambling Task. *Brain and Cognition, 64*, 164-169.

Colombetti, G. (2008). The somatic marker hypotheses, and what the Iowa Gambling Task does and does not show. *British Journal of Philosophy of Science, 59*, 51-71.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769-786.

Corr, P. J. (Ed.). (2008). *The reinforcement sensitivity theory of personality*. New York, NY: Cambridge University Press.

Cyders, M. A., & Coskunpinar, A. (2011). Measurement of constructs using self-report and behavioral lab tasks: Is there overlap in nomothetic span and construct representation for impulsivity? *Clinical Psychology Review, 31*, 965-982.

Cyders, M. A., & Coskunpinar, A. (2012). The relationship between self-report and lab task conceptualizations of impulsivity. *Journal of Research in Personality, 46*(1), 121-124.

Dunn, B. D., Dalgleish, T., & Lawrence, A. D. (2006). The somatic marker hypothesis: A critical evaluation. *Neuroscience & Biobehavioral Reviews, 30*, 239-271.

Franken, I. H., & Muris, P. (2005). Individual differences in decision-making. *Personality and Individual Differences, 39*, 991-998.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*, 101-135.

Gansler, D. A., Jerram, M. W., Vannorsdall, T. D., & Schretlen, D. J. (2011). Does the Iowa Gambling Task measure executive function? *Archives of Clinical Neuropsychology*, *26*, 706-717.

Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety: An inquiry into the functions of the septo-hippocampal system*. Oxford, England: Oxford University Press.

Greene, K., Krcmar, M., Walters, L. H., Rubin, D. L., & Hale, J. L. (2000). Targeting adolescent risk-taking behaviors: The contribution of egocentrism and sensation-seeking. *Journal of Adolescence*, *23*, 439-461.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*(1), 3-25.

Harden, K., & Tucker-Drob, E. M. (2011). Individual differences in the development of sensation seeking and impulsivity during adolescence: Further evidence for a dual systems model. *Developmental Psychology*, *47*, 739-746.

Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, *13*, 517-523.

Højsgaard, S., Halekoh, U., & Yan, J. (2006). The R Package geepack for Generalized Estimating Equations. *Journal of Statistical Software*, *15*(2), 1-11.

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, *45*(7), 1-47.

Horstmann, A., Villringer, A., & Neumann, J. (2012). Iowa gambling task: There is more to consider than long-term outcome. Using a linear equation model to disentangle the impact of outcome and frequency of gains and losses. *Frontiers in Neuroscience*, *6*, 61.

Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test: Form 4* [Berlin Intelligence Structure Test]. Göttingen, Germany: Hogrefe.

Johnson, J. G., Wilke, A., & Weber, E. U. (2004). A domain-specific scale measuring risk perceptions, expected benefits, and perceived-risk attitude in German-speaking populations. *Polish Psychological Bulletin*, *35*, 153-163.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Brown, J., & Mackintosh, N. (2009). Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence*, *37*, 374-382.

Keye, D., Wilhelm, O., & Oberauer, K. (2009). Structure and correlates of the German version of the Brief UPPS impulsive behavior scales. *European Journal of Psychological Assessment*, *25*, 175-185.

Keye, D., Wilhelm, O., Oberauer, K., & van Ravenzwaaij, D. (2009). Individual differences in conflict-monitoring: Testing means and covariance hypothesis about the Simon and the Eriksen Flanker task. *Psychological Research/Psychologische Forschung*, *73*, 762-776.

Lauriola, M., Panno, A., Levin, I. P., & Lejuez, C. W. (2014). Individual differences in risky decision making: A meta-analysis of sensation seeking and impulsivity with the balloon analogue risk task. *Journal of Behavioral Decision Making*, *27*, 20-36.

Lejuez, C., Aklin, W. M., Jones, H. A., Richards, J. B., Strong, D. R., Kahler, C. W., & Read, J. P. (2003). The balloon analogue risk task (BART) differentiates smokers and nonsmokers. *Experimental and Clinical Psychopharmacology*, *11*, 26-33.

Lejuez, C., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., . . . Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, *8*(2), 75-84.

Lewandowsky, S. (2011). Working memory capacity and categorization: Individual differences and modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 720-738.

Maia, T. V., & McClelland, J. L. (2004). A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa Gambling Task. *Proceedings of the National Academy for Science U S A*, *101*, 16075-16080.

Mata, R., Josef, A. K., Samanez-Larkin, G. R., & Hertwig, R. (2016). Age differences in risky choice: A meta-analysis. *Annals of the New York Academy of Sciences*, *1235*, 18-29. doi:10.1111/j.1749-6632.2011.06200.x

Ortner, T. M., Proyer, R. T., & Kubinger, K. D. (2006). *Theorie und Praxis Objektiver Persönlichkeitstests* [Theory and application of objective persoanliyt tests]. Bern, Switzerland: Huber.

Pleskac, T. J. (2008). Decision making and learning while taking sequential risks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 167-185.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Revelle, W. (2015). *Psych: Procedures for personality and psychological research* (Version 1.5.1.). Evanston, IL: Northwestern University. Retrieved from http://CRAN.R-project.org/package=psych

Roberti, J. W. (2004). A review of behavioral and biological correlates of sensation seeking. *Journal of Research in Personality*, *38*, 256-279.

Schmitz, F., Manske, K., Preckel, F., & Wilhelm, O. (2016). The multiple faces of risk taking: Scoring alternatives for the balloon analogue risk task. *European Journal of Psychological Assessment*, *32*, 17-38. doi: 10.1027/1015-5759/a000335

Simons, J. S., Dvorak, R. D., & Batien, B. D. (2008). Methamphetamine use in a rural college population: Associations with marijuana use, sensitivity to punishment, and sensitivity to reward. *Psychology of Addictive Behaviors*, *22*, 444-449.

Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K., & Klauer, K. C. (2014). Behavioral components of impulsivity. *Journal of Experimental Psychology: General*, *143*, 850-886. doi:10.1037/a0033981

Steinberg, L. (2010). A dual systems model of adolescent risk-taking. *Developmental Psychobiology*, *52*, 216-224.

Steingroever, H., Wetzels, R., Horstmann, A., Neumann, J., & Wagenmakers, E.-J. (2012). Performance of healthy participants on the Iowa Gambling Task. *Psychological Assessment*, *25*, 180-193.

Suhr, J. A., & Tsanadis, J. (2007). Affect and personality correlates of the Iowa Gambling Task. *Personality and Individual Differences*, *43*, 27-36.

Tamez, E., Myerson, J., & Hale, S. (2012). Contributions of associative learning to age and individual differences in fluid intelligence. *Intelligence*, *40*, 518-529.

Tomb, I., Hauser, M., Deldin, P., & Caramazza, A. (2002). Do somatic markers mediate decisions on the gambling task? [Letter]. *Nature Neuroscience*, *5*, 1103-1104.

Toplak, M. E., Sorge, G. B., Benoit, A., West, R. F., & Stanovich, K. E. (2010). Decision-making and cognitive abilities: A review of associations between Iowa Gambling Task performance, executive functions, and intelligence. *Clinical Psychology Review*, *30*, 562-581.

Torrubia, R., Avila, C., & Caseras, X. (2008). Reinforcement sensitivity scales. In P. J. Corr (Ed.), *The reinforcement sensitivity theory of personality* (pp. 188-227). New York, NY: Cambridge University Press.

Torrubia, R., Avila, C., Molto, J., & Caseras, X. (2001). The Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. *Personality and Individual Differences*, *31*, 837-862.

Townsend, L., Flisher, A. J., Gilreath, T., & King, G. (2006). A systematic literature review of tobacco use among adults 15 years and older in sub-Saharan Africa. *Drug and Alcohol Dependence*, *84*, 14-27.

van Honk, J., Hermans, E. J., Putman, P., Montagne, B., & Schutter, D. J. (2002). Defective somatic markers in sub-clinical psychopathy. *NeuroReport*, *13*, 1025-1027.

Whiteside, S. P., & Lynam, D. R. (2001). The Five Factor Model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, *30*, 669-689.

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*, 433. doi:10.3389/fpsyg.2013.00433

Wilhelm, O., & Schulze, R. (2002). The relation of speeded and unspeeded reasoning with mental speed. *Intelligence*, *30*, 537-554.

Wittmann, W. W., & Suess, H.-M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem-solving performances via Brunswik symmetry. In P. L. Ackerman, P. C. Kyllonen & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 77-108). Washington, DC: American Psychological Association.

Xu, S., Korczykowski, M., Zhu, S., & Rao, H. (2013). Risk-taking and impulsive behaviors: A comparative assessment of three tasks. *Social Behavior and Personality*, *41*, 477-486.

Zermatten, A., Van der Linden, M., d 'Acremont, M., Jermann, F., & Bechara, A. (2005). Impulsivity and decision making. *Journal of Nervous and Mental Disease*, *193*, 647-650.

Zuckerman, M. (1996). Item revisions in the Sensation Seeking Scale Form V (SSS-V). *Personality and Individual Differences*, *20*, 515-515.

Zuckerman, M. (2007). The Sensation Seeking Scale V (SSS-V): Still reliable and valid. *Personality and Individual Differences*, *43*, 1303-1305.

Zuckerman, M., & Kuhlman, D. (2000). Personality and risk-taking: Common biosocial factors. *Journal of Personality*, *68*, 999-1029.