Eliciting risk preferences: is a single item enough?

Don C. Zhang, Department of Psychology Louisiana State University,

Gino Howard, Department of Psychology, Louisiana State University

Russell A. Matthews, Department of Management, University of Alabama

Tyler Cowley, Department of Psychology, Louisiana State University

Author Note.

This paper has been accepted for publication. This is not the final copy of record. Please cite as: Zhang, D.C., Howard, G., Matthews, R.A. & Cowley, T. (in press). Eliciting risk preferences: Is a single item enough? *Journal of Risk Research*.

This work was supported, in part, by funding from the National Science Foundation (SES #2142891). We declare no conflict of interest. Correspondences should be addressed to Don C. Zhang, Ph.D., Department of Psychology, Louisiana State University. Baton Rouge, LA 70803; zhang1@lsu.edu.

Abstract

Economists and psychologists frequently use single-item measures of risk preferences despite potential limitations in reliability and criterion validity compared to their multi-item counterparts. This can be particularly problematic when individual differences in risk preferences are used to predict real-world economic, health, and financial outcomes. In this paper, we compare a popular single-item measure of risk preference, the *General Risk Question (GRQ)*, to multi-item measures of domain-general and -specific risk preference measures. In a two-wave survey study of 434 adults, we found that the GRQ had good psychometric reliability and converged with other multi-item measures of risk preferences. The GRQ also exhibited a similar pattern of associations with other personality and demographic variables as compared to multi-item measures. However, the predictive validity of the GRQ was lower than multi-item measures for most of the outcomes examined. The GRQ also explained less incremental variance for real-world outcomes over the Big Five personality traits than the multi-item counterparts. Although the GRQ is a construct-valid measure of risk preferences, researchers should nonetheless consider the trade-off between survey efficiency and predictive efficacy when deciding whether a single item is enough.

Keywords: Risk preference, individual differences, single item, predictive validity, psychometrics

Eliciting risk preferences: is a single item enough?

Social scientists have historically been skeptical of using self-report questionnaires (Nisbett & Wilson, 1977). This is also true of risk researchers, where self-report measures of risk preferences (i.e., stated risk preferences) were assumed to be inferior to behavioral elicitations in the lab (i.e., revealed risk preferences) (Harrison & Rutström, 2008). Recent research from psychology and economics, however, has highlighted that: self-report measures are more than 'cheap talk' and that they reflect genuine individual differences in risk preferences (Arslan et al., 2020; Dohmen et al., 2011; Steiner et al., 2021); and that self-report measures are often more useful than behavioral elicitations of risk preferences for predicting real-life outcomes (Charness et al., 2020; Frey et al., 2017; Kaiser & Oswald, 2022; Tasoff & Zhang, 2021).

Considering the benefits of self-report measures, several large-scale economic panel surveys include a measure of self-report risk preference (e.g., German Socio-Economic Panel (SOEP) (Dohmen et al., 2011). Given practical restraints such as survey length, these panel studies typically include only a one-item measure of risk preference. A popular item known as the *General Risk Question* (GRQ) asks respondents to indicate the degree to which they are risk-seeking (versus risk-averse). The simplicity and brevity of the GRQ also make it ideal for measuring individual differences in risk preferences in a wide range of experimental and non-experimental settings (Bran & Vaidis, 2019; Lonnqvist et al., 2015). It is not surprising that the GRQ has also become increasingly popular in both economic and psychological research. In fact, the GRQ is one of the few psychological traits measured annually in the SOEP.

Despite its practical advantages, a single-item measure of risk preference has several potential theoretical and methodological shortcomings (Fisher et al., 2016; Menkhoff & Sakha, 2017). According to psychometric theory, single-item measures of psychological constructs such

as risk preference suffer from low reliability and content coverage, which could undermine the construct validity and predictive efficacy of the measure (Matthews et al., 2022). Indeed, psychometricians have long cautioned against the use of single-item measures of individual differences (Gardner et al., 1998; Wanous & Hudy, 2001). Nevertheless, single items may be comparable, or even superior, to their multi-item counterparts (Allen et al., 2022; Bergkvist & Rossiter, 2007) and are suitable for some situations. Indeed, single-item measures remain widely used in a variety of social science disciplines such as psychology, marketing, and organizational behavior (Ang & Eisend, 2018; Matthews et al., 2022).

Although the GRQ is widely used in economic and psychological sciences as a concise measure of general risk preference, it is not clear if the GRQ meets the requirements for construct validity, especially compared to other multi-item measures of risk preference that have undergone more thorough construct validation efforts. Without establishing the validity of the GRQ, it may be difficult to accurately understand the impact of risk preferences on economic, health, and social outcomes (Bagozzi et al., 1991). Furthermore, the observed relationships between risk preference measured with the GRQ with real-world outcomes may be understated. Thus, whether the use of the GRQ in lieu of multi-item measures of risk preference is justified depends on the psychometric properties of the measure.

The purpose of this paper, therefore, is to examine the psychometric qualities of the single-item measure of risk preference (i.e., the GRQ) compared to other multi-item self-report measures of risk preferences. Specifically, we report comparisons of psychometric reliability and predictive validity of single-item and multi-item measures of risk preference. Furthermore, we examine the convergent validity of the GRQ with multi-item measures of risk preferences as well as the discriminant validity of single- vs. multi-item measures of risk preference with other

individual difference constructs (e.g., Big Five personality). Together, this paper sheds light on the construct validity of the GRQ and its predictive efficacy compared to other multi-item measures of risk preferences.

Self-report measures of risk preferences

Risk researchers have historically been skeptical of self-report measures of economic preferences such as risk preferences (i.e., stated risk preferences; Charness et al., 2013). Instead, it is assumed that an accurate assessment of risk preferences can only be obtained with incentive-compatible tasks where people must choose between risky vs. riskless options (Holt & Laury, 2002). Unlike stated risk preference measures, these lab-based elicitations (i.e., revealed risk preferences) reveal people's risk preferences based on their *behaviors*, rather than self-reports.

Although a behavioral approach has been the zeitgeist for scholars interested in measuring risk preferences, emerging research has identified several empirical and theoretical shortcomings of this approach. First, behavior *across* different elicitations of risk-taking tends to diverge, as they each capture unique variance associated with the *task,* rather than the respondent (Pedroni et al., 2017; Zhou et al., 2021). For example, Pedroni et al., (2017) found minimal convergence in individual risk preferences across six different sixteen elicitation methods. Their findings call into question the validity of behavioral elicitations as measures of stable risk preferences (c.f. Holzmeister & Stefan, 2021).

Second, risk preferences are often confounded with task-specific characteristics that may require different risky behaviors to maximize incentives. Performance on incentive-compatible tasks may be influenced by the decision-makers' capacity to take *calculated risks* to maximize their winnings within the situational constraints of the tasks. Thus, revealed risk preferences may reflect individual differences in quantitative ability (e.g., numeracy), in addition to risk

preference (Lilleholt, 2019; Millroth et al., 2020). These limitations have led to diminishing confidence in laboratory-based behavioral measures (Rouder & Haaf, 2019) and a renewed appreciation for stated measures of risk preferences, such as the GRQ, in economic and psychological research (Arslan et al., 2020; Dohmen et al., 2011; Zhang et al., 2023).

Single vs. multi-item measures of risk preferences

Single-item measures are often used in longitudinal panel studies because they provide an efficient measure of economic, demographic, and health characteristics. Single-item measures can also be used to measure attitudes (e.g., life satisfaction), traits (e.g., core self-efficacy), and beliefs (e.g., political ideology). In the case of risk preference, the most popular single-item measure appears in the German Socio-Economic Panel (SOEP) where respondents are asked to respond to the item: "Are you generally a person who is fully prepared to take risks, or do you try to avoid taking risks?" using a 10-point Likert scale ranging from 1: 'not at all willing to take risks' to 10: 'very willing to take risks'" (Dohmen et al., 2011). This item, which has been labeled the *General Risk Question* (Arslan et al., 2020), is also used in other panel surveys such as the Household, Income and Labour Dynamics in Australia (HILDA) and the British Household Panel Survey (BHPS).

Benefits of the GRQ

There are several good reasons to use a single-item measure of risk preference. First, single-item measures can be completed very quickly and impose minimal burden on survey takers. This benefit is particularly important in survey research, where participant motivation can affect their attentiveness to survey items. Relatedly, multi-item measures often contain similar items that give the appearance of redundancy, which may frustrate survey takers and negatively affect survey completion rates (Fuchs & Diamantopoulos, 2009). Collectively then, it has been

suggested that the motivational and emotional toll of long and repetitive surveys may ultimately undermine the validity of the measures (Rogelberg et al., 2002). Applied social scientists (e.g., economists) are also limited by the amount of time that the target population has for their research. Therefore, it is critical to maximize the efficiency by which psychological constructs are measured. Second, a well-constructed single-item measure is better suited when measuring 'doubly concrete' constructs, where the object and attribute of the measurement are unambiguous for the respondent (Drolet & Morrison, 2001). For example, the predictive validity of single-item measures for consumer attitudes about brands is comparable to its multi-item counterparts (Bergkvist & Rossiter, 2007). Regardless of the possible benefits of using single-item measures, as discussed next, psychometricians have noted several potential shortcomings of single-item measures.

Psychometric reliability of the GRQ

Increasingly, risk researchers have begun to recognize the importance of psychometric properties for measures of risk preferences such as reliability and validity (Mata et al., 2018). Single-item measures are often criticized on the grounds of reliability because, unlike multi-item measures, the reliability of single-item measures is harder to compute in a typical empirical investigation; that is, internal consistency estimates of reliability, the most commonly used measure of reliability, cannot be calculated for single-item measures. For this reason, the reliability of single-item measures in empirical studies is (incorrectly) presumed to be unknown. Even if computed based on established methods (Wanous & Hudy, 2001), single-item measures tend to have *lower pseudo-internal consistency* reliability than multi-item measures of the same construct (Allen et al., 2022). As discussed by Matthews et al. (2022) though, this requires

scholars to consider other types of reliability. That is, several methods exist for empirically examining the psychometric reliability of single-item measures, which we describe below.

As noted, traditional approaches for assessing psychometric reliability rely on internal consistency, which is obtained by computing the average inter-item correlations (Cho & Kim, 2015). This approach, however, does not allow for the estimation of reliability for single-item measures, such as the GRQ. Two methods exist for estimating the reliability of single-item measures. The first method for estimating the psychometric reliability for single-item measures is by using factor analysis whereby single-item measures are loaded onto a factor that is composed of multi-item measures of the same construct (Wanous & Hudy, 2001). A key part of this approach is the assumption that the communality of a measure is a conservative estimate of reliability therefore the total variance is equivalent to the sum of communality, specificity, and unreliability, and when the variance is unknown or unspecified, the communality is equal to the reliability (Wanous & Hudy, 2001).

The communality of a measure is calculated by squaring the factor loadings and summing them based on the number of overall extracted components. In other words, when considering the total variance, communalities are the proportion of variance explained specifically by the factors. A key issue with this approach though is that if the multi-item measure demonstrates weaker psychometric characteristics, this will directly affect the interpretation of the single-item measure's reliability (Matthews et al. 2022). The reliability of single-item measures can also be observed through test-retest reliability. For single items, test-retest reliability can be estimated by administering the GRQ twice and observing the Pearson's correlation between the two administrations. Alternatively, test-retest reliability can also be ascertained using the intra-class correlation (ICC).

Criterion validity of the GRQ

Criterion validity reflects the degree to which it predicts meaningful outcomes of interest (Cronbach & Meehl, 1955). The criterion validity of a predictor variable is limited by its reliability. Thus, conventional wisdom amongst psychometricians is that single-item measures tend to have weaker predictive validity than multi-item measures because single-item measures are assumed to have more random measurement error, which undermines its reliability (Novick, 1966). In theory, however, there is no reason why single-item measures should exhibit weaker predictive validity or reliability than multi-item counterparts (also see Allen et al., 2022). Multi-item measures are also more prone to criterion contamination, whereby some of the items may reflect characteristics that are irrelevant to the focal construct of interest (Drolet & Morrison, 2001). Said another way, multi-item measures of risk preference can include items that reflect other constructs such as recklessness or assertiveness, which may introduce measurement error.

Existing research on the criterion validity of single measures revealed inconsistent findings. Matthews et al. (2022)'s examination of 91 organizational constructs revealed that single-item measures – when properly developed – are equally reliable and predictive than multi-item measures of the same constructs. On average, the authors observed a degradation of only $r = .02$ in criterion validity between a single-item and multi-item measure of the same construct. Similarly, a meta-analytic investigation of 189 advertising studies found that single-item measures had almost identical criterion validity as multi-item measures for predicting consumer attitudes (Ang & Eisend, 2018). Thus, it remains an empirical question whether the GRQ – a single-item measure of risk preference – is more or less predictive than multi-item measures.

Bandwidth vs. Fidelity

One factor that may influence the criterion validity of the GRQ is the conceptual correspondence between the predictor and outcome (Hogan & Roberts, 1996). Research has shown that domain-general measures are superior for prediction when the criterion variable is broad and covers multiple domains of risk-taking. In their examination of the general risk factor, for example, Highhouse et al., (2017) found that the general risk factor was more predictive than risk attitude in any single domain (e.g., health) for predicting conceptually broad outcomes such as workplace deviance, which entails multiple risky domains (e.g., social, ethical, financial). Similarly, Zhang et al., (2019) found that general risk propensity was more predictive of broad outcomes such as entrepreneurial intentions than domain-specific measures of DOSPERT. In contrast, domain-specific measures of risk preference (e.g., health risks) are more predictive of outcomes in corresponding domains (e.g., long-term health problems) (Also see Charness et al., 2020). Because the GRQ is a domain-general measure of risk preference, we expect that it will be more predictive of broad outcomes (e.g., work deviance) than domain-specific risk preference measures whereas domain-specific measures will be more predictive of domain-matched outcomes.

Discriminant and convergent validity of the GRQ

Convergent validity refers to the degree to which a construct converges (i.e., correlates) with measures of similar constructs. Convergent validity is often the first step in establishing the construct validity of new measures. Evidence of convergent validity of the GRQ can be acquired by examining the degree to which the GRQ correlates with other measures of general risk preference. Here, correlations greater than $r = 0.70$ are considered sufficient evidence of convergent validity (Carlson & Herdman, 2012). Discriminant validity refers to the degree to

which a construct differs from other theoretically distinct constructs (Rönkkö & Cho, 2021). Discriminant validity is particularly important in the social sciences where new constructs and measures are often introduced to the literature without evidence for their uniqueness from existing ones (i.e., old wine, new bottle) (Shaffer et al., 2016). One common criterion for establishing the uniqueness of personality constructs is by demonstrating its uniqueness from the Big Five (Goldberg & Saucier, 1998), the most popular and comprehensive model of personality. Considering recent meta-analytic findings showing the uniqueness of general risk propensity from the Big Five (Highhouse et al. 2022), we expect the GRQ to be distinct from the Big Five.

Method

Sample and Procedure

We gathered data from two time points using Prolific.co, a crowd-sourcing platform used for social science research[1]. Data were collected from a total of 608 participants at Time 1 and 520 participants at Time 2 four weeks later. The GRQ was administered at both time points. We also included four multi-item risk measures split between two surveys such that each survey had one domain-general and one domain-specific measure. Other individual difference measures (e.g., personality) were administered at Time 1 and all outcome variables are measured at Time 2 to reduce common method variance. We included four attention check questions (e.g., "If you are paying attention, please select strongly disagree") across the two surveys. We removed participants who missed more than one out of four attention check questions. The final sample size of attentive participants that participated in both surveys was 434. The average age was 37 years old (SD = 12.8), 50% male, 86% Caucasian, 69% were employed part-time or full-time, and 62% had at least an associate degree or higher.

[1] This research was approved by the Louisiana State University Institutional Review Board (#11920)

Measures

General Risk Question. General Risk Question (GRQ) is a single-item measure of general risk preference. The item asks respondents: "How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please tick a box on the scale, where 0 = 'not at all willing to take risks' and 10 = 'very willing to take risks'.

Domain-General Risk Measures

GRiPS (Time 1). General Risk Propensity Scale (GRiPS) is an eight-item, self-report scale (Zhang et al., 2019). The GRiPS questions are on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). An example item from this scale is ''Taking risks makes life more fun". The internal consistency of the GRiPS is 0.93.

RPS (Time 2). The Risk Propensity Scale (RPS) is a seven-item, self-report scale (Meertens & Lion, 2008). The RPS questions are on a 9-point Likert scale (1 = totally disagree to 9 = totally agree). Responses with higher scores indicate greater risk propensity. An example item from this scale is "I take risks regularly". The internal consistency of the RPS is 0.83.

Domain-Specific Risk Measures

DOSPERT (Time 1). Domain-Specific Risk-Taking Scale (DOSPERT) is a 30-item self-report measure of domain-specific risk-taking propensity (Blais & Weber, 2006). It measures individual risk intentions/behavioral intentions in five different domains (Blais and Weber, 2006). The DOSPERT questions are on a 7-point Likert scale (1 = extremely unlikely to 7 = extremely likely). An example item from this scale is ''Having an affair with a married man/woman" (social). The internal consistency of the DOSPERT range from 0.64 to 0.87. The overall internal consistency of the summated DOSPERT score is 0.88.

RTI (Time 2). The Risk-Taking Index (RTI) is a 12-item measure of an individual's engagement across 6 risk-taking domains (Nicholson et al., 2005). The RTI questions are on a 5-point Likert scale (1 = never to 5 = very often). Items are separated by risk-taking behaviors now versus in the past. An example item from this scale is ‘‘recreational risks”. The internal consistency of the RTI ranges from 0.64 to 0.78 across the five domains. The overall internal consistency of the summated RTI score is 0.76.

Other Individual Differences

Big Five Personality. IPIP-NEO-60 is a 60-item self-report measure of the Big Five personality (Maples-Keller et al., 2019). The IPIP-NEO-60 questions are on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). An example item from this scale is ‘‘Lose my temper”. The internal consistency ranges from 0.74 to 0.86.

Dark Triad. Short Dark Triad Scale (SD3) is a 27-item, self-report scale (Jones & Paulhus, 2014). The SD3 questions are on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). Example items from this scale are Narcissism ‘‘I have been compared to famous people’’, Machiavellianism ‘‘It’s not wise to tell your secrets’’, and Subclinical Psychopathy ‘‘Payback needs to be quick and nasty’’. The internal consistency ranges from 0.73 to 0.76.

Subjective Numeracy. The Subjective Numeracy Scale (SNS) is an eight-item, self-report scale (Fagerlin et al., 2007). The SNS has four questions that ask participants to assess their numerical ability across multiple contexts. Additionally, it has four questions that ask participants to state their preferences for the presentation of numerical and probabilistic information. The SNS questions are on a 6-point Likert scale (1 = not at all good to 6 = extremely good). An example item from this scale is ‘‘How good are you at working with

fractions?". The alpha score for the overall scale demonstrates good reliability with an alpha of 0.91.

Broad Outcomes

Workplace Deviance. We measured workplace deviance using a nine-item, self-report scale developed by Robinson & O'Leary-Kelly, (1998). Participants responded to each statement on a 5-point Likert scale (1 = Never to 5 = Almost Always). An example item from this scale is ''Deliberately bent or broke rules". The internal consistency of the scale is 0.81.

Entrepreneurial Intentions. The Entrepreneurial Intention Questionnaire (EIQ) is a five-item, self-report scale (Linan et al., 2011). The EIQ questions are on a 7-point Likert scale (1 = total disagreement to 5 = total agreement). An example item from this scale is ''A career as an entrepreneur is attractive for me". The internal consistency of the scale is 0.93.

Narrow Outcomes

Narrow outcomes are characterized by specific behaviors that fall within a single domain of risk-taking, rather than outcomes that reflect risk-taking across multiple domains. We included eleven outcome variables (e.g., job change, car accidents, etc.) matched to a specific dimension of risk-taking. The full list of outcomes, items, and relevant risk dimensions is presented in Table 1.

Analytical Plan

We examine the validity of the GRQ in several ways. First, we computed and compared the reliability of the GRQ using three methods: 1) test-retest reliability, 2) communality, and 3) ICC (Wanous & Hudy, 2001). Second, we examined the convergent and discriminant validity of the GRQ by examining its correlation with other multi-item domain-general measures such as the GRiPS and RPS as well as domain-specific measures such as the DOSPERT and RTI.

Finally, we examine the discriminant validity of the GRQ from the Big Five personality traits. We used multiple regression analyses to examine the proportion of variance in the GRQ explained by the combined Big Five traits to demonstrate the uniqueness of GRQ from the Big Five. We also include subjective numeracy as an additional individual difference variable, for exploratory purposes. Finally, we examine the criterion validity of the GRQ by examining its correlation with a wide range of broad (e.g., workplace deviance) and narrow outcomes (e.g., number of broken bones).

Results

Table 2 contains the means, standard deviations, and internal consistencies of the study's risk-taking measures, as well as demographic characteristics. Consistent with past research, we found both sex[2] (rs = 0.14 to 0.27) and age (rs = -0.10 to -0.24) were significantly correlated with risk preference across different measures. We did not, however, find any significant correlations between risk preference and employment, education, or income level.

Psychometric reliability

Table 3 contains the reliability of the GRQ for both survey administrations as well as the reliability of multi-item measures of risk preferences. Consistent with recent research on examining the psychometric characteristics of single-item measures (Matthews et al., 2022), test-retest reliability was assessed using traditional Pearson's correlation coefficient and the intra-item correlation (ICC), which has been suggested as an alternative approach to calculating test-retest reliability. We found that the GRQ had acceptable test-retest reliability (r = 0.70, ICC_{mixed} = 0.71[95% C.I. = 0.66, 0.76]) albeit slightly lower than the test-retest reliability of multi-item measures (e.g., GRiPS, $r_{3\ months}$ = 0.80, Zhang et al., 2019). In addition to the test-retest

[2] Males were more risk-seeking

reliability, reliability for the single-item GRQ was also obtained using the factor analysis methods described by (Wanous & Hudy, 2001.) To do so, we computed the communality of the GRQ with each of the two multi-item measures (GRiPS and RPS). Using this method, we found the GRQ exhibited good reliability (0.76 to 0.93).

Convergent validity

We first examined the convergent validity for single vs. multi-item of general risk preference measures. The GRQ measured at both Time 1 and Time 2 was significantly correlated with both the domain-general measures of risk preference (GRiPS; $r = 0.69$ to 0.79; RPS; $r = 0.65$ to 0.74) as well as the summated risk preference score using domain-specific measures (Summated DOSPERT: $r = 0.52$ to 0.57; Summated RTI: $r = 0.41$ to 0.49). The magnitude of correlations suggests that the GRQ had better convergent validity with domain-general measures, compared to summated scores using domain-specific scales.

We next examined the convergence between the GRQ with the domain-specific risk scores obtained in the DOSPERT (Time 1) and the RTI (Time 2). The GRQ was moderately correlated with each of the five DOSPERT dimensions: social ($r = 0.33$), recreation ($r = 0.52$), finance ($r = 0.46$), health ($r = 0.36$), and ethics ($r = 0.27$). Likewise, the GRQ was also moderately correlated with five RTI dimensions: recreation ($r = 0.40$), career ($r = 0.24$), financial ($r = 0.43$), safety ($r = 0.33$), and social ($r = 0.32$). Interestingly, the GRQ was not significantly correlated with the health dimension of the RTI ($r = 0.07$). These results suggest that the GRQ better captures one's general preference for risks, rather than risk preferences in any single domain.

Discriminant validity

We first examined the degree to which risk preferences measured using single- vs. multi-item measures are empirically distinct from the Big Five, the dominant model of personality. Table 4 contains the results of multiple regression analysis. Overall, the Big Five explained 25% and 34% of the variance in two multi-item measures of general risk propensity (RPS and GRiPS) respectively. The Big Five also explained 18% and 36% of the variance in the two domain-specific measures of risk preference (RTI and DOSPERT) respectively. As for the single item, we found that the Big Five explained 17% and 27% of the variance for the GRQ measured at Time 1 and Time 2 respectively. Collectively, these results suggest that the GRQ – like other measures of risk preference – is relatively distinct from the Big Five model of personality.

In addition to the discriminant validity from the Big Five, we also examined the degree to which risk preference measures are distinct from the Dark Triad model. Table 5 contains the results. First, the dark triad traits explained 21.37% of the variance in the GRQ while explaining 17.68% and 33.18% of the variance in the RTI and DOSPERT respectively. The dark triad traits explained between 24.06% and 34.24% of the variance for the domain-general risk preference measures (GRiPS, RPS).

For exploratory purposes, we also examined the association between risk preference and subjective numeracy. We found a weak relationship between subjective numeracy and risk preferences. Observed bivariate correlations range from $r = 0.08$ to $r = 0.17$.

Criterion and incremental validity

GRQ vs. general risk preference measures. We compared the criterion validity of the GRQ with multi-item measures of general risk preference measures for both broad and specific outcomes. Table 5 contains the bivariate correlations between each risk measure and both broad and narrow outcomes. We present the results separately for cross-sectional predictions where the

predictors and criterion were both measured at the same time (concurrent validity) versus time-lagged predictions where the predictors and criterion were measured at separate time points (predictive validity).

We found that compared to multi-item measures of risk preference, the GRQ was less predictive of workplace deviance. However, the GRQ had comparable and slightly stronger predictions for entrepreneurial intentions than other multi-item measures. As for narrow outcomes, we found the GRQ to be a weaker predictor across most outcome variables when compared with multi-item measures. The difference in the magnitude of prediction was most pronounced for speeding frequency, number of romantic relationships, frequency of cheating in relationships, and frequency of job change.

Although the difference was smaller for other outcomes, the GRQ did not 'out-predict' any other multi-item measures of general risk preference for any of the narrow outcomes or vice versa. The average criterion validity for the GRQ across 13 outcomes was r_{mean}= .14 and r_{mean} = .16 for concurrent and predictive validities respectively whereas the criterion validity for the two multi-item general measures were r_{mean} = .18 (GRiPS) and r_{mean} = .20 (RPS) and the criterion validity for the two multi-item domain-specific measures were r_{mean} = .22 (DOSPERT) and r_{mean} = .23 (RTI).

GRQ vs. domain-specific risk preference measures. We next compared the criterion validity of the GRQ with domain-specific measures (e.g., DOSPERT). Table 6 contains the bivariate correlations between each risk measure and both broad and narrow outcomes. For broad outcomes, we found the GRQ to better predict deviance than some of the risk domains (e.g., recreation) but not others (e.g., ethical). Interestingly, we found that the GRQ better predicted entrepreneurial intent than any of the specific domains of risk from both the DOSPERT and RTI.

For narrow outcomes, a domain-relevant measure of risk preference always outperformed the GRQ in terms of predictive validity. Speeding was better predicted by recreation and safety risk attitudes; the number of romantic relationships was better predicted by financial and recreation risk attitudes; frequency of cheating in romantic relationships was better predicted by ethical risk attitudes; frequency of moving and job change were better predicted by social risk attitudes; and number of car accidents was better predicted by safety risks. Overall, the prediction of narrow outcomes was significantly better if domain-relevant measures of risk-taking were used.

Incremental validity of the GRQ vs. multi-item risk measures

We next examined the incremental predictive validity of the GRQ over the Big Five, as well as comparisons of incremental validity of the GRQ vs. multi-item risk preference measures (Table 7). After controlling for the Big Five, the GRQ (Time 1 and Time 2) added incremental prediction for entrepreneurial intent. The GRQ also added incremental prediction for deviance, but only when it was measured at the same time as the outcome variable. In contrast, the multi-item measures explained more incremental variance over the Big Five for workplace deviance. Interestingly, the GRQ contributed more incremental prediction for entrepreneurial intent than all but one multi-item measure (summated RTI).

As far as specific outcomes, the GRQ measured at Time 1 only explained incremental variance over the Big Five for three outcomes (speeding, broken bones, frequency of moving). In contrast, the eight-item GRiPS also measured at Time 1 explained incremental variance beyond the Big Five for six outcomes. Of the narrow outcomes where both GRQ and GRiPS explained incremental variance, the GRiPS explained more unique variance than the GRQ for speeding (4.4% vs. 1.0%); number of broken bones (2.1% vs. 1.4%), but not frequency of moving (1.4%

vs. 1.7%). The incremental prediction was even greater for summated scores using a domain-specific measure (e.g., DOSPERT).

Together, the predictive utility of the GRQ was generally inferior to that of multi-item measures of general risk preference. The eight-item GRiPS and the seven-item RPS both exhibited stronger overall predictive utility as well as incremental prediction over the Big Five. However, the GRQ demonstrated a surprisingly strong prediction of entrepreneurial intent compared to multi-item measures. Overall, general risk preference obtained by summing across domains (e.g., DOSPERT and RTI) resulted in the best overall prediction, though at the cost of a much longer scale.

Discussion

Single-item measures of risk preferences are frequently used in longitudinal survey studies and laboratory experiments due to their economic efficiency. This paper is the first to examine the psychometric qualities of the GRQ, a popular single-item measure of general risk preferences in terms of reliability, convergent validity, discriminant validity, and criterion validity relative to multi-item measures of risk preferences. In a two-wave survey study, we compared the psychometric qualities (reliability, convergent validity, discriminant validity, and criterion validity) of the GRQ with four well-established measures of risk preferences.

Reliability

The most mentioned limitation of single-item measures is psychometric reliability, an argument that has been put into question based on recent research on the application of single-item measures (Matthews et al, 2022). Consistent with recommendations by Matthews et al. (2002) that it is necessary to empirically examine the possible pros and cons of using single-item measures (relative to multi-item measures), results from the current program of research suggest

that both the test-retest reliability and internal consistency of the GRQ is adequate and comparable to longer multi-item measures.

Convergent Validity

The GRQ converged with existing multi-item measures of domain-general risk preferences. This is not surprising, as the multi-item measures consist of items similar to that of the GRQ (e.g., "I enjoy taking risks in most aspects of my life; GRiPS). The correlation between the GRQ and summated scores from domain-specific measures (e.g., DOSPERT) was slightly lower than with domain-general measures (e.g., GRiPS). One explanation is that domain-specific measures such as the DOSPERT do not cover all possible risk domains that are sampled by individuals in their response to the GRQ. Put simply, domain-specific measures may not capture all types of risks that people think about when they answer questions in the GRQ, which results in content deficiency of the DOSPERT, and reduced convergence between the GRQ.

Discriminant Validity

The GRQ also inhabited a similar position within the nomological network as multi-item measures of risk preferences. Specifically, the GRQ and multi-item risk measures share a similar pattern of relationships with the Big Five and Dark Triad. Specifically, the GRQ and multi-item risk measures share a similar pattern of relationships with the Big Five and Dark Triad. Also consistent with past research, the GRQ appears to be distinct from both the Big Five and Dark Triad (Highhouse et al., 2022; Joseph & Zhang, 2021). The Big Five only accounted for between 17% and 26% of the variance in the GRQ whereas the Dark Triad accounted for between 23% and 27%. Interestingly, the Big Five accounted for more variance in multi-item measures of risk preferences. This could be attributed to the measurement error associated with single-item measures. Alternatively, these findings may suggest greater contamination of multi-item

measures. Nevertheless, our findings show that risk preference appears to be distinct from the Big Five, regardless of the method of measurement.

Criterion Validity

Our results suggest that the GRQ was inferior to multi-item measures of risk preferences for predicting real-world outcomes. It also explained less incremental variance over the Big Five, despite less shared variance with the Big Five. Interestingly, the GRQ was equally or slightly more predictive of entrepreneurial intentions compared to multi-item measures of general risk preferences as well as domain-specific measures (specific domain scores and summated scores). These results suggest that the GRQ may be sufficient in studies of entrepreneurial activities. However, the GRQ was noticeably worse at predicting negative outcomes such as workplace deviance, excessive speeding, cheating in romantic relationships, number of broken bones, and number of car accidents. These results suggest that the GRQ may be better suited for predicting more industrious aspects of risk-taking than reckless and unethical risky behaviors. These results are similar to meta-analytic findings showing that general risk preference was a better predictor of adaptive (vs. maladaptive) outcomes (Highhouse et al. 2022).

Implications

Our findings suggest that the use of GRQ in longitudinal panel studies such as the SOEP is likely justified due to the benefit of economic efficiency without significant sacrifices to reliability. We also find that the GRQ, for the most part, exhibits similar psychometric qualities as multi-item counterparts due to its similar pattern of correlations with other personality measures such as the Big Five. Thus, the position of the GRQ within the nomological network of personality traits is likely similar to that of multi-item measures of risk preferences. Overall, these findings suggest that the GRQ is a valid measure of general risk preferences.

Although we found that the predictive efficacy of the GRQ was weaker for some of the outcomes measured in this study, the observed reductions were modest. Therefore, studies with sufficient statistical power (e.g., large sample size) should be able to detect true relationships between risk preference and real-world outcomes as long as the researchers accept that the magnitude of that association may be attenuated. For this reason, the GRQ is not advised when expected effects are weak or when the study is potentially underpowered. In such cases, we advise researchers to consider longer measures of general risk preference (e.g., GRiPS) if space allows and domain-specific measures (e.g., DOSPERT) when the study context and outcomes of interests are well-aligned with a specific domain (e.g., health). Nevertheless, single measures of risk preference may be suitable if the researchers are not concerned with specific predictions. For example, the GRQ may be fine to use as a statistical control of general risk preferences.

Our findings differ from what was observed by Matthews et al. (2022), where the criterion validity of their single-item measures performed much better than the GRQ. One explanation is that the GRQ was not developed to maximize construct validity. Thus, our results do not necessarily speak to the shortcoming of using single items as a measure of all risk preferences. However, it does point to the possibility that the GRQ may need revision to be on equal psychometric footing with its multi-item counterparts. In sum, researchers should carefully consider the context of the research and weigh the trade-off between survey efficiency and predictive accuracy when deciding whether to use the GRQ.

References

Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single Item Measures in Psychological Science: A Call to Action. *European Journal of Psychological Assessment*, *38*(1), 1–5. https://doi.org/10.1027/1015-5759/a000699

Ang, L., & Eisend, M. (2018). Single versus Multiple Measurement of Attitudes: A Meta-Analysis of Advertising Studies Validates the Single-Item Measure Approach. *Journal of Advertising Research*, *58*(2), 218–227. https://doi.org/10.2501/JAR-2017-001

Arslan, R. C., Brümmer, M., Dohmen, T., Drewelies, J., Hertwig, R., & Wagner, G. G. (2020). How people know their risk preference. *Scientific Reports*, *10*(1), 1–14.

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing Construct Validity in Organizational Research. *Administrative Science Quarterly*, *36*(3), 421. https://doi.org/10.2307/2393203

Bergkvist, L., & Rossiter, J. R. (2007). The Predictive Validity of Multiple-Item versus Single-Item Measures of the Same Constructs. *Journal of Marketing Research*, *44*(2), 175–184. https://doi.org/10.1509/jmkr.44.2.175

Blais, A.-R., & Weber, E. U. (2006). A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, *1*(1), 15.

Bran, A., & Vaidis, D. C. (2019). Assessing risk-taking: What to measure and how to measure it. *Journal of Risk Research*, 1–14. https://doi.org/10.1080/13669877.2019.1591489

Carlson, K. D., & Herdman, A. O. (2012). Understanding the Impact of Convergent Validity on Research Results. *Organizational Research Methods*, *15*(1), 17–32. https://doi.org/10.1177/1094428110392383

Charness, G., Garcia, T., Offerman, T., & Villeval, M. C. (2020). Do measures of risk attitude in the laboratory predict behavior under risk in and outside of the laboratory? *Journal of Risk and Uncertainty*, *60*(2), 99–123. https://doi.org/10.1007/s11166-020-09325-6

Charness, G., Gneezy, U., & Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization*, *87*, 43–51.

Cho, E., & Kim, S. (2015). Cronbach's Coefficient Alpha: Well Known but Poorly Understood. *Organizational Research Methods*, *18*(2), 207–230. https://doi.org/10.1177/1094428114555994

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, *9*(3), 522–550. https://doi.org/10.1111/j.1542-4774.2011.01015.x

Drolet, A. L., & Morrison, D. G. (2001). Do we really need multiple-item measures in service research? *Journal of Service Research*, *3*(3), 196–204.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring Numeracy without a Math Test: Development of the Subjective Numeracy Scale. *Medical Decision Making*, *27*(5), 672–680. https://doi.org/10.1177/0272989X07304449

Fisher, G. G., Matthews, R. A., & Gibbons, A. M. (2016). Developing and investigating the use of single-item measures in organizational research. *Journal of Occupational Health Psychology*, *21*(1), 3–23. https://doi.org/10.1037/a0039139

Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science Advances*, *3*(10). https://doi.org/10.1126/sciadv.1701381

Fuchs, C., & Diamantopoulos, A. (2009). Using single-item measures for construct measurement in management research: Conceptual issues and application guidelines. *Die Betriebswirtschaft*, *69*(2), 195.

Gardner, D. G., Cummings, L. L., Dunham, R. B., & Pierce, J. L. (1998). Single-item versus multiple-item measurement scales: An empirical comparison. *Educational and Psychological Measurement*, *58*(6), 898–915.

Goldberg, L. R., & Saucier, G. (1998). What is beyond the Big Five? *Journal of Personality*, *66*(4), 495–524.

Harrison, G. W., & Rutström, E. E. (2008). Chapter 81 Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods. In C. R. Plott & V. L. Smith (Eds.), *Handbook of Experimental Economics Results* (Vol. 1, pp. 752–767). Elsevier. https://doi.org/10.1016/S1574-0722(07)00081-9

Highhouse, S., Wang, Y., & Zhang, D. C. (2022). Is Risk Propensity Unique from the Big Five Factors of Personality? A Meta-Analytic Investigation. *Journal of Research in Personality*, 104206. https://doi.org/10.1016/j.jrp.2022.104206

Hogan, J., & Roberts, B. (1996). Issues and non-issues in the fidelity-bandwidth trade-off. *Journal of Organizational Behavior*, *17*(6), 627–637.

Holzmeister, F., & Stefan, M. (2021). The risk elicitation puzzle revisited: Across-methods (in)consistency? *Experimental Economics*, *24*(2), 593–616. https://doi.org/10.1007/s10683-020-09674-8

Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A Brief Measure of Dark Personality Traits. *Assessment*, *21*(1), 28–41. https://doi.org/10.1177/1073191113514105

Joseph, E., & Zhang, D. C. (2021). Personality Profile of Risk Takers: An Examination of the Big Five Facets. *Journal of Individual Differences*.

Kaiser, C., & Oswald, A. J. (2022). The scientific value of numerical measures of human feelings. *Proceedings of the National Academy of Sciences*, *119*(42), e2210412119. https://doi.org/10.1073/pnas.2210412119

Lilleholt, L. (2019). Cognitive ability and risk aversion: A systematic review and meta analysis. *Judgment & Decision Making*, *14*(3).

Lonnqvist, J.-E., Verkasalo, M., Walkowitz, G., & Wichardt, P. C. (2015). Measuring individual risk attitudes in the lab: Task or ask? An empirical comparison. *JOURNAL OF ECONOMIC BEHAVIOR & ORGANIZATION*, *119*, 254–266. https://doi.org/10.1016/j.jebo.2015.08.003

Maples-Keller, J. L., Williamson, R. L., Sleep, C. E., Carter, N. T., Campbell, W. K., & Miller, J. D. (2019). Using Item Response Theory to Develop a 60-Item Representation of the NEO PI–R Using the International Personality Item Pool: Development of the IPIP–NEO–60. *Journal of Personality Assessment*, *101*(1), 4–15. https://doi.org/10.1080/00223891.2017.1381968

Mata, R., Frey, R., Richter, D., Schupp, J., & Hertwig, R. (2018). Risk Preference: A View from Psychology. *Journal of Economic Perspectives*, *32*(2), 155–172. https://doi.org/10.1257/jep.32.2.155

Matthews, R. A., Pineault, L., & Hong, Y.-H. (2022). Normalizing the use of single-item measures: Validation of the single-item compendium for organizational psychology. *Journal of Business and Psychology*, 1–36.

Meertens, R. M., & Lion, R. (2008). Measuring an Individual's Tendency to Take Risks: The Risk Propensity Scale. *Journal of Applied Social Psychology*, *38*(6), 1506–1520. https://doi.org/10.1111/j.1559-1816.2008.00357.x

Menkhoff, L., & Sakha, S. (2017). Estimating risky behavior with multiple-item risk measures. *Journal of Economic Psychology*, *59*, 59–86.

Millroth, P., Juslin, P., Winman, A., Nilsson, H., & Lindskog, M. (2020). Preference or ability: Exploring the relations between risk preference, personality, and cognitive abilities. *Journal of Behavioral Decision Making*, *n/a*(n/a). https://doi.org/10.1002/bdm.2171

Nicholson, N., Soane, E., Fenton-O'Creevy, M., & Willman, P. (2005). Personality and domain-specific risk taking. *Journal of Risk Research*, *8*(2), 157–176. https://doi.org/10.1080/1366987032000123856

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, *3*(1), 1–18. https://doi.org/10.1016/0022-2496(66)90002-2

Pedroni, A., Frey, R., Bruhin, A., Dutilh, G., Hertwig, R., & Rieskamp, J. (2017). The risk elicitation puzzle. *Nature Human Behaviour*, *1*(11), 803–809. https://doi.org/10.1038/s41562-017-0219-x

Rogelberg, S. G., Church, A. H., Waclawski, J., & Stanton, J. M. (2002). Organizational survey research. *Handbook of Research Methods in Industrial and Organizational Psychology*, 140–160.

Rönkkö, M., & Cho, E. (2021). An Updated Guideline for Assessing Discriminant Validity. *Organizational Research Methods*, 42.

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*(2), 452–467. https://doi.org/10.3758/s13423-018-1558-y

Shaffer, J. A., DeGeest, D., & Li, A. (2016). Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs. *Organizational Research Methods*, *19*(1), 80–110.

Steiner, M. D., Seitz, F. I., & Frey, R. (2021). Through the window of my mind: Mapping information integration and the cognitive representations underlying self-reported risk preference. *Decision*, *8*(2), 97–122. https://doi.org/10.1037/dec0000127

Tasoff, J., & Zhang, W. (2021). The Performance of Time-Preference and Risk-Preference Measures in Surveys. *Management Science*, mnsc.2020.3939. https://doi.org/10.1287/mnsc.2020.3939

Wanous, J. P., & Hudy, M. J. (2001). Single-Item Reliability: A Replication and Extension. *Organizational Research Methods*, *4*(4), 361–375. https://doi.org/10.1177/109442810144003

Zhang, D. C., Barratt, C. L., & Smith, R. W. (2023). The Bright, Dark, and Gray Sides of Risk Takers at Work: Criterion Validity of Risk Propensity for Contextual Work Performance. *Journal of Business and Psychology*. https://doi.org/10.1007/s10869-023-09872-0

Zhang, D. C., Highhouse, S., & Nye, C. D. (2019). Development and validation of the general risk propensity scale (GRiPS). *Journal of Behavioral Decision Making*, *32*(2), 152–167.

Zhou, R., Myung, J. I., Mathews, C. A., & Pitt, M. A. (2021). Assessing the validity of three tasks of risk-taking propensity. *Journal of Behavioral Decision Making*, *34*(4), 555–567.

Table 1. List of specific outcomes matched to relevant risk domain

Specific Outcomes	Item	DOSPERT Domain	RTI Domain
Cigarettes	"About how many cigarettes do you smoke on an average day? "	Health	Health
Speeding	"How often do you drive above the speed limit by more than 10 miles per hour?"	Recreation	Safety
Tickets	"How many speeding tickets have you received in the past 3 years?"	Health	Safety
Credit Card Debt	"About how much credit card debt (In US. Dollars) do you carry from month to month"	Financial	Financial
Broken Bones	"How many broken bones do you have?"	Health	Safety
Relationships	"How many romantic relationships longer than 3 months have you been involved in during the past 5 years?"	Social	Social
Cheating	"In your lifetime, approximately how many times have you cheated on your partner in a monogamous relationship"	Ethical	Social
Moving	"How many times have you moved residence in the past 10 years?"	Social	Career
Job Change	"How many times have you changed jobs in the past 10 years?"	Social	Career
Shoplift	"How many times have you shoplifted in the past 5 years?"	Ethical	Social
Car Accidents	"In the last 5 years, in how many car accidents have you been involved?"	Health	Safety

Table 2. Means, Standard Deviations, and Demographic Correlates of Risk-Taking Measures

			Demographic Correlates			
	Mean	SD	Sex	Age	Education	Income
GRQ (Time 1)	5.35	2.16	.16**	-.20**	.02	.07
DOSPERT	3.11	0.82	.27**	-.24**	.05	.07
GRiPS	2.53	0.93	.23**	-.23**	.07	.09
GRQ (Time 2)	5.00	2.11	.14**	-.20**	-.05	.06
RTI	2.21	0.62	.19**	-.10*	.09	.10*
RPS	3.80	1.38	.18**	-.18**	.02	.08

Table 3. Reliability of Single vs. Multi-Item Measures of Risk Preferences

	Reliability
General Risk Question	
Test-Retest Reliability (4 weeks, Pearson's *r*)	0.70
Test-Retest Reliability (4 weeks, ICC)	0.71
Communality with GRiPS	0.70
Communality with RPS	0.72
Multi-Item Scales	
GRiPS	0.93
RPS	0.83
DOSPERT Summated	0.88
RTI Summated	0.76

Notes. Reliability indices for multi-item scales are calculated with the Cronbach's alpha.

Table 4. Regression Results between Personality and General Risk Preferences

	Time 1			Time 2		
	GRQ	GRiPS	DOSPERT	GRQ	RPS	RTI
Big Five Traits	β	β	β	β	β	β
Neuroticism	-0.14	-0.08	0.02	-0.15	-0.24*	0.04
Extraversion	1.36**	0.58**	0.39**	0.95**	0.54**	0.19**
Openness	0.72**	0.38**	0.49**	0.64**	0.66**	0.26**
Agreeableness	-0.53**	-0.42**	-0.52**	-0.57**	-0.53**	-0.25**
Conscientiousness	-0.54**	-0.36**	-0.29**	-0.56**	-0.65**	-0.22**
Multiple R^2	26%	34%	36%	17%	25%	18%
Dark Triad Traits						
Narcissism	1.36**	0.54**	0.34**	0.95**	0.53**	0.11*
Psychopathy	1.51**	0.88**	0.91**	1.54**	1.28**	0.57**
Machiavellianism	-0.70**	-0.22**	-0.16*	-0.53**	-0.46**	-0.21**
Multiple R^2	27%	40%	41%	23%	27%	20%

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

Table 5. Criterion Validity of General Risk Preference Measures

	Concurrent Validity			Predictive Validity		
	GRQ	GRiPS	DOSPERT	GRQ	RPS	RTI
Broad Outcomes						
Workplace Deviance	.16**	.26**	.40**	.21**	.30**	.37**
Entrepreneurial Intent	.39**	.36**	.36**	.40**	.33**	.31**
Narrow Outcomes						
Cigarette Use	.00	.01	.05	.02	.06	.13**
Speeding	.18**	.29**	.38**	.25**	.31**	.33**
Traffic Tickets	.06	.09	.13**	.12*	.12*	.16**
Credit Card Debt	.04	.05	.08	.11*	.10*	.11*
Broken Bones	.15	.16**	.19**	.11*	.17**	.22**
Relationships	.18**	.25**	.30**	.20**	.26**	.22**
Cheating	.13**	.19**	.23**	.14**	.19**	.25**
Moving	.19**	.19**	.23**	.18**	.22**	.25**
Job Change	.14**	.20**	.24**	.17**	.22**	.24**
Shoplift	.06	.11*	.13**	.08	.13**	.20**
Car Accident	.12*	.15**	.18**	.15**	.15**	.21**

Notes. * $p < .05$; ** $p < .01$

Table 6. Criterion Validity of Domain-Specific Risk Preferences

	GRQ		DOSPERT					RTI					
	Time 1	Time 2	Social	Recreation	Financial	Health	Ethical	Recreation	Health	Career	Financial	Safety	Social
Broad Outcomes													
Workplace Deviance	.16**	.21**	.23**	.18**	.20**	.39**	.46**	.11*	.19**	.22**	.25**	.32**	.25**
Entrepreneurial Intent	.39**	.40**	.26**	.28**	.32**	.25**	.15**	.00	.26**	.32**	.16**	.16**	.07
Narrow Outcomes													
Cigarette Use	.00	.02	.05	-.06	.01	.15**	.08	-.06	.36**	.06	.08	.06	-.05
Speeding	.18**	.25**	.08	.36**	.26**	.36**	.24**	.28**	.06	.04	.22**	.47**	.10*
Traffic Tickets	.06	.12*	.02	.10*	.09	.11*	.14**	.12**	.04	.04	.13**	.15**	.10*
Credit Card Debt	.04	.11*	.11*	.03	-.02	.09	.10*	.02	.10*	.09	.10*	.06	.02
Broken Bones	.15	.11*	.14**	.19**	.05	.21**	.05	.20**	.08	.11*	.06	.22**	.09*
Relationships	.18**	.20**	.15**	.23**	.23**	.22**	.19**	.12*	.04	.08	.21**	.19**	.15**
Cheating	.13**	.14**	.13**	.08	.13**	.23**	.28**	.10*	.14**	.07	.19**	.22**	.14**
Moving	.19**	.18**	.25**	.21**	.12*	.18**	.05	.22**	.08	.19**	.09	.13**	.17**
Job Change	.14**	.17**	.25**	.20**	.15**	.16**	.10*	.15**	.09	.28**	.04	.18**	.09
Shoplift	.06	.08	.11*	.02	.06	.09	.20**	.10*	.12*	.09	.12*	.13**	.18**
Car Accident	.12*	.15**	.08	.13**	.13**	.15**	.13**	.15**	.06	.08	.12*	.18**	.17**

Notes. * $p < .05$; ** $p < .01$

Table 7. Incremental Validity of Risk Preference Over Big Five

		Delta R^2					
	R^2 - Big Five	GRQ (Time 1)	GRQ (Time 2)	GRiPS	DOSPERT	RTI	RPS
Broad Outcomes							
Deviance	.188	.006	.015**	.017**	.059**	.058**	.031**
Entrepreneur	.101	.072**	.086**	.053**	.063**	.145**	.045**
Specific Outcomes							
Cigarettes	.034	.000	.000	.001	.005	.021**	.005
Speeding	.061	.010*	.035**	.044**	.110**	.085**	.070**
Ticket	.022	.000	.007	.001	.007	.017**	.005
CC Debt	.013	.000	.007	.000	.002	.007	.006
Broken Bones	.018	.014*	.006	.021**	.033**	.042**	.027**
Relationships	.102	.001	.007	.009*	.029**	.017**	.022**
Cheating	.008	.003	.004	.011*	.021**	.024**	.015**
Move	.064	.017**	.015**	.014*	.022**	.031**	.022**
Job Change	.068	.003	.009*	.011*	.017**	.022**	.018**
Shoplift	.024	.000	.001	.003	.004	.025**	.005
Car Accident	.028	.002	.010*	.007	.017**	.029**	.009*

Notes. * $p < .05$; ** $p < .01$