Toward a Better Measurement of Strategic Skills: The Multiple Choice Strategic Quotient (McSQ)

Andrea Piazzoli[1,3], Gianpietro Sgaramella[2,3], and Alan Mattiassi[3](✉)

[1] Department of Economics and Statistics, University of Siena, 53100, Siena, Italy
[2] AXES Unit, IMT School for Advanced Studies Lucca, Piazza San Francesco 19, 55100 Lucca, Italy
[3] GAME Science Research Center, IMT School for Advanced Studies Lucca, Piazza San Francesco 19, 55100 Lucca, Italy
alan.mattiassi@gmail.com

Abstract. This paper develops the strategic quotient construct and its measure, the Strategic Quotient (SQ), a test to assess actual strategic competencies. While the developmental and refinement work is still in progress, here we document the process of rethinking the foundations of the test and the concepts it is based on, in order to design a novel version of the SQ, the "multiple choice Strategic Quotient" (mcSQ).

Keywords: strategic quotient · strategic ability · strategic interaction · games

1 Introduction

1.1 Aims and Goals

In the scientific literature, the definition and conceptualization of strategic ability varies from study to study. Recently, the concept of strategic quotient has been developed and presented (Bilancini et al. 2019). The strategic quotient has been conceived as separate from general intelligence and from the intelligence quotient (Deary 2012). While certain intelligent quotient measuring tools primarily assess cognitive abilities related to problem-solving and reasoning, the strategic quotient encompasses additional dimensions such as social intelligence, theory of mind, and the ability to understand and respond to strategic interactions. Conversely, it does not assess areas such as knowledge, perception, motor or practical skills that are considered by intelligent quotient tests (Otero et al. 2022), although it could in principle be affected by language and verbal skills, memory, and attention. Inhibition and possibly other executive functions are hypothesized to be critical for the strategic quotient. Additionally, and crucially, social skills such as empathy, perspective-taking, and understanding social cues play an important role in the concept of strategic quotient, but less so in the intelligence quotient. Overall, the conceptualization of the strategic quotient emphasizes its multidimensional nature, combining cognitive and social skills to assess and respond to strategic interactions in different decision-making contexts.

M. Clayton et al. (Eds.): INTETAIN 2023, LNICST 560, pp. 3–19, 2024.
https://doi.org/10.1007/978-3-031-55722-4_1

While strategic behavior had previously been studied extensively (Camerer et al. 2015; Weyhrauch and Culbertson 2014; Gill et al. 2023), no measure of strategic behavior or ability had previously been proposed before the introduction of the strategic quotient concept in the aforementioned work (Bilancini et al. 2019). In that paper, the authors presented the validation of a novel measure of strategic abilities called "Strategic Quotient", a test that we will refer to as SQ hereon (vs. the concept strategic quotient, lowercase). The authors conducted an experiment that was designed to create a strategic context, i.e. the context in which the outcome of a choice does not depend solely on the individual choice but on the combination of choices of all participants (Stahl and Wilson 1995). Indeed, the participants played games from the behavioral economic literature (Guala and Mittone 2010) in which they were required not only to understand the structure and payoffs of the games but also to assess and predict the behavior of other participants and try to choose the optimal response given the predicted choices of others.

Here we start by observing a number of shortcomings that we found by working with the SQ since its publication and then present a number of possible solutions that allowed us to develop a new version of the test that introduces, among other novelties, multiple choice-type responses and is as such called "mcSQ" (multiple-choice Strategic Quotient). Our goal is not to present a novel iteration of the test but rather to document the refining process of the test in order to simplify it, make it more coherent, and, most importantly, we propose a number of indicators that could empower the interpretation of the measured behavior.

1.2 Conceptualization of the Strategic Quotient

While Bilancini et al. (2019) first proposed the term "Strategic Quotient" both as the concept of strategic ability and as the name of its measuring tool, there are antecedents in the scientific literature that preceded it and led to its construction (Jelenc and Swiercz 2011). The historical development and evolution of the concept of strategic quotient in literature can be traced through several key milestones and research contributions. Here are some notable and contributing developments:

1. Game Theory: The basis for the study of strategic interactions can be traced back to the emergence of game theory in the mid-20th century (Von Neumann and Morgenstern 2007). Game theory provided a formal framework for analyzing strategic decisions and understanding how individuals reason and strategize in competitive situations.
2. Social Intelligence and Theory of Mind (ToM): In the 1980s and 1990s, research on Social Intelligence and ToM gained prominence (Wellman et al. 2001). These areas of study focused on the ability of individuals to understand and interpret the thoughts, emotions, and intentions of others. The inclusion of social intelligence as a component of strategic thinking contributed to the development of the strategic quotient foundations (Parales-Quenza 2006).
3. Experimental Studies: Experimental studies began to investigate individual differences in strategic thinking and decision-making (Golman et al. 2020) (Bayer and Renou 2012). Researchers developed various experimental games and tasks to measure participants' strategic abilities, often focusing on aspects such as anticipation

(Gill and Prowse 2016), prediction (Otero et al. 2022), and decision-making in competitive settings (Camerer et al. 2004). These studies provided empirical evidence that can be categorized as a significant step in the development of the strategic quotient.
4. Conceptualization of the Strategic Quotient: The term 'strategic quotient' began to receive attention in 2016 with the research of Boncinelli, Bilancini, and Mattiassi, published in 2019. The researchers sought to define and conceptualize it as a distinct construct, separate from general intelligence measures such as IQ. The focus shifted from logical items to understanding decision-making, performance, and success in interactions with the sample.

1.3 Features of the SQ Test

The main feature of the SQ, which also establishes how it is different from intelligence quotient measures, is that the correctness of the answer to most items of each game depends on the distribution of choices that all participants make. By using the pool of responses given by the entire sample of participants, the authors were able to calculate a score for each game and a general score that allowed them to rank the participants in order of strategic ability, i.e. from those who won the most to those who won the least. Since winning and losing were defined by the ability to estimate the behavior of other participants, each game had no *a priori* correct answer, but it varied depending on the responses of the whole sample. Participants with higher Strategic Quotient scores, then, were the ones who chose the most advantageous option based on a better assessment of the competence of others' behavior and on the best prediction of their behavior.

The SQ is particularly interesting also because it allows for repeated administrations: since it is possible to correctly estimate the behavior of others, and since each population has its own profile of actions, a participant can play against different populations and rank differently. As an anecdotical example, during one of the first administrations of SQ at a game fair, a participant obtained a low rank when playing against the entire population of participants playing at the fair but obtained one of the top ranks when repeating the test against participants from the high IQ society Mensa. It is worth noting that other means of ability estimation lose their validity when used a second time. This peculiarity of the SQ makes it possible to use the test in a competitive context such as human resources hiring, for example, to find the best candidate or the person who can best estimate the behavior of several groups or a specific group, etc.

Additionally, and most importantly, the SQ does not measure intelligence quotient. In their paper, the SQ authors examined the possible correlation between the SQ score and the Raven APM matrices and found it to be non-significant.

The SQ test is designed to assess not only an individual's ability to understand games and to perform rational problem-solving but also the ability to evaluate the skills of others and to predict their actions. To obtain a high score in the SQ, the following skills are thought to be important:

- Critical Thinking: involves the analysis of complex information. This competency is important for evaluating available information and formulating effective strategies.
- Metacognition: is concerned with being aware of one's thinking and learning process. Metacognition helps identify one's own knowledge gaps and develop effective learning strategies.

– Self-regulation: is about the ability to monitor one's behavior and adapt it to the demands of the situation. This implies mental flexibility and the ability to modify one's strategies based on the feedback received.
– Perspective-Taking: refers to the cognitive ability to understand and adopt the point of view, thoughts, feelings, and intentions of others. It involves mentally simulating the experiences of others and taking into account their unique perspectives, beliefs, and motivations. This process enables individuals to empathize with others, make accurate attributions, and engage in prosocial behaviors. Perspective-taking plays a crucial role in social interactions, interpersonal relationships, and the development of empathy, compassion, and cooperation.

Accordingly, the SQ test has been built using three kinds of items:

– Rationality items: these items have the form of a question that uses the game rules but presents a perfectly rational competitor, and as such do not consider the answers of the rest of the participants ("What would you choose if you were to play a game against a perfectly rational robot?"). Here, a correct answer exists, and the player needs to find it to score the maximum points: as such, these items measure by design the rational skills of the participant.
– Items related to the proficiency in predicting others' rationality: these items comprise the largest part of the SQ items and have the form of the traditional question posed by the game in the behavioral economic literature ("What do you choose in this game?"). There is also a secondary form of these items that relates to the same indicator: the one in which the participant is asked to play the same game in another role, such as the proposer and the responder in the ultimatum game ("What would you choose as the other player?"). These items measure by design the ability to understand the interaction of participants' choices; however, they do not measure rationality alone nor predict skills alone but have also a component of perspective-taking.
– Items related to the proficiency in predicting the prediction of others: these items ask participants to predict what other participants think that other participants will do – or the predicting ability of the other participants. These items take the form of a question that asks what the other participants did in a previous game or in a previous item of the same game, as such considering not only the prediction of the others' performance but also the prediction of the other's prediction ("what do you think the other participants predicted?"). These items are indeed focused on metacognition and perspective-taking.

All items require critical thinking and self-regulation; however, while critical thinking is considered in the original SQ paper, we argue that self-regulation has been left on the sidelines, when it should be better emphasized.

2 Methods

First, we identify a number of shortcomings of the version of SQ that has been published and used since publication. Then we propose a number of improvements and discuss the constructs at the foundations of the test in order to rediscuss them to drive the development of a new version.

2.1 Identified Shortcomings

Since the Bilancini, Boncinelli, and Mattiassi paper was published, the SQ has been used in several contexts and developed in multiple iterations. Here are some shortcomings that were found and some considerations on the weaknesses that might drive the design of a novel version of the test.

Items Too Time Costly or Too Difficult to Understand. Several items have been found to be problematic because they seem to be too difficult to understand, badly worded, or just too time-consuming.

Response Modality. Participants have been found to answer in multiples of five. Indeed, any responder that tries to figure out a response is thought to accumulate evidence until a particular response is obtained; however, evidence accumulates by using multiple "anchors" that function as reference points (see for example Kvam et al. 2022).

The Arbitrary Threshold for Points Earning. In the previous version of SQ, the interval of responses that earned participants points was arbitrary. The formula with which each game was scored considered the distribution of answers of all participants and identified an interval that was considered representative of a good answer in relation to the distribution profile of answers. Unfortunately, different games elicited different distributions of responses, thus rendering them non-comparable.

Pre-test Section. The first SQ aimed at measuring the strategic quotient with a set of games very different from each other, divided into two parts: a pre-test and a test based on the pre-test. As such, participants had to switch continuously between different pages in order to solve the games and this procedure took a lot of time.

Very Different Games. The difference between the games is aimed at including all the faces of strategic thinking related to the strategic processes, but this aim is very difficult to achieve, especially through only one item for every indicator, which is not in line with the psychometric literature.

Consistency Between the Data-Driven Approach and the Constructs. A matrix of correlation to aggregate the indicators into four main constructs was used to extract four factors. Correlations, however, have weak explanatory power for the internal consistency of a construct in psychometrics. The authors tried to explain the measure with two main indicators: mentalization and rationality. While generally rationality allows individuals to grasp the structure of the game and its payoff, mentalization allows players to predict other players' behavior. One could wonder if these two indicators are sufficient in order to account for the complexity of strategic skills.

Classic Game-Theoretical Payoffs. Payoffs were calculated and presented in a classical game theory manner, but people attribute spontaneously payoffs based on their preferences about implicit goals that diverge from the optimization of gains (Gavetti 2012). However, the problem of suboptimal gain in favor of preferred behaviors is endogenous in rationality: e.g., if others' choices are perceived as not fair, it "feels" rational to punish that behavior even at a personal cost. On the contrary, if emotions or other preferences are considered exogenous to rationality (Alaoui and Penta 2016), then one should enforce a more rational interpretation of strategic ability that is not connected

to the real competencies that are needed to predict others' behavior in society: e.g., a chess master can make very solid predictions on the next actions in a game but what if he has to predict emotions that can drive other's actions (Capra 2004)? This interpretation of the typical behavioral games used in economics can be an excellent tool to explore the difference between a more descriptive approach to *rationality* in which we can consider as rational the emotions that drive a specific behavior more than the maximum gain, and *hyper-rationality*, defined as a reduction of reality only in strictly logical terms like in classical game theory (Alos-Ferrer and Buckenmaier 2018). Fairness (Suleiman 2022) and punishment (Xiao and Houser 2005) are two preferred behaviors that override the rational preference for an optimization strategy at a personal cost.

2.2 Proposed Improvements

The main possible improvements to the SQ, in our opinion, are:

- Simplification of the questions (to reduce the probability of errors in responses)
- Homogeneity of the measured constructs (to be able to limit the range of skills measured to fewer constructs with more indicators, and not more constructs with few items each)
- Progressive adjustments of the questions to stay within the psychometric validity thresholds.
- Inclusion of a dual interpretation of the results of the test (people attributing rational or hyper-rational styles to others).

One first change we introduced in the novel version of the SQ is the response modality, changing it from a continuous scale to a discrete Likert scale of twenty-one points (from 0 to 100 or 0% to 100% in multiples of five), facilitating the choice between anchors and representing them in graphical form (as a list of ordered possible choices). As such, we also increased the probability of picking the right interval to score points by attributing points to the two percentages choices near the average choice of the whole participants' sample (one full score for picking one of the two percentages nearest to the *a posteriori* correct response and a half score for picking the percentages near those). A lot of people reported a certain confusion about the questions with intensive use of percentages after the conclusion of the test. The phrasing of many items was too convoluted. Reference to other items required too many elements to elaborate, and as such we introduced a more direct approach like "What do you think is the average percentage expressed in the previous question?".

To fix the homogeneity among very different games and items, we consistently introduced three kinds of items for each game (with only a few exception). Indeed, the internal coherence of the test should be favored by the constant use of three kinds of games. The items of the first kind consist of one of the typical Discrete Choice Experiments (or DCEs) economic games (Straub and Murnighan 1995): participants are asked to play the game by choosing one of the possible answers. The items of the second kind consist of a question asking the participant to estimate the average choice of the other subjects to the previous question. The items of the third kind ask the participant to estimate the average response to the second part. The first kind of item requires a multiple-choice response (in which rationality or hyper-rationality is measured) while the second

and third parts evaluate the answers given by the sample with precision intervals around the average, giving points to the responder scores with the abovementioned method.

For the third improvement, the sample of participants needs to be extended to see if the standardization in the scoring can be distributed in a normal way in some of the questions. This is a work in progress and out of the scope of the present paper.

As for the last improvement, participants can be divided with respect to the tendency towards rationality or hyper-rationality. This is, as well, out of the scope of the present paper.

2.3 Additional Construct Contributions

In the original SQ article, the authors used two different approaches to understand the data from an exploratory analysis: data-driven and theory-driven. Here, we start from the validation of that SQ version and develop a set of constructs that might improve the SQ's ability to describe the behavior of the participants, its interpretability, and its easiness of use. We changed many of the original questions to these goals, in order to also improve its construct validity.

Construct #1: The evaluation of logical competencies. This is a classical approach to the first part of every game that we have in this test. Some games in the test have a correct answer that can be only identified by using logic-related capabilities. These items are useful to evaluate logic-related competencies (Weyhrauch 2016) and consequentially to see the estimation capability that we will talk about later in this section.

- Logic-related skills

Construct #2: The difference between emotional vs rational estimation and logical competence (van Dijk and Vermunt 2000). Usually, the economics literature considers anything prioritized with respect to a strictly material gain approach as an emotion-based bias (Kahneman 2003). This way of thinking conflicts with the variants of game theory that tried to include as endogenous factors the psychology of individuals (Athanasiou et al. 2015). So, if these factors can be considered inside the system to evaluate payoffs, then "hyper-rational people" are those who exclude these psychological factors in their estimations. This reasoning works in talking about games that can have different answers based on the preferences of people (Tisserand et al. 2015) and that are not *a priori* right or wrong. If the base games had right or wrong answers, another indicator could be used based on the estimation of logic-related competencies or incompetencies (Dhir et al. 2018). Indicators could be conceived in this way:

- Rationality vs hyper-rationality
- A good vs a bad estimation of other's logic-related capabilities

Construct #3: Pattern of responses in generalization of one competence to the rest of the participants. Some stable patterns emerged in the answers of individuals. In some cases, with respect to a rational approach to the tendency toward a mean, people tended to generalize to others their competencies in all the questions, or with a tendency to attribute more extreme competencies. Thus, two indicators could be conceived as:

- Overestimation or underestimation
- Approximation or extremization

In summary, we propose five continuous indicators as mentioned before, to evaluate the answers of participants along with the general score that they provide.

3 Experimental Procedure

Here we present the results of a pilot study with a version of the mcSQ that started to integrate the proposed changes and that has been more focused on the five indicators mentioned in the previous section. The goal is not to validate the test but to show how the changes affected its psychometric properties.

3.1 Sampling Strategy and Participant Selection

The sample is composed by 160 participants, of which 47.12% were females, 46.88% were males, 1.25% were non-binaries and 3.75% preferred not to answer. The mean age is 25.06 y.o. with a standard deviation of 8.2 years.

Participants were recruited at the University of Firenze in an economic course, at PLAY: Festival del Gioco fair in Modena, and online on social networks.

3.2 Data Analysis and Interpretation Techniques

Participants took the mcSQ in different settings, ranging from in-class to online, constituting a convenience sample with no control over the setting. However, the mcSQ was implemented in Google Forms, and as such data were collected automatically. We applied the scoring that was presented in par.1.5 and performed correlation analyses on the results between all items.

3.3 Results and Findings

Descriptive Statistics

The distribution between the 4 education levels registered (from having completed the high school to having 1 or 2 degrees and finally being a PhD condidate) is: 73.12% high school (level one), 7.5% bachelor's degree (level 2), 16.25% master's degree (level 3) and 3.13% Ph.D. (level 4).

Those who came from an economics, statistics and mathematics background were further divided into four levels, being no education at all (13.75% - level 0), marginally educated (22.55% - level 1), partially educated (33.75% - level 2) and educated in the subjects previously mentioned (30% - level 3).

Table 1. and Fig. 1 represent differences between the average scores of participants pertaining to different levels of education, while Table 2. and Fig. 2 represent different levels of education specifically in ESM disciplines.

Both the data on education and on ESM education need to be considered descriptive only, since the scarce numerosity of the groups with second, third, and fourth levels of

Table 1. Scores for education level

Total points				
	Lev 1	Lev 2	Lev 3	Lev 4
Valid	117	12	26	5
Mean	12.333	12.333	14.231	13.600
Std. Deviation	2.586	2.103	2.804	3.362
Minimum	7.000	8.000	10.000	10.000
Maximum	21.000	15.000	21.000	19.000

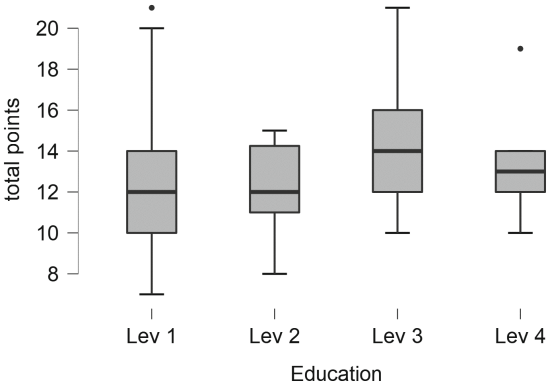

Fig. 1. Boxplot representation of scores by education level.

Table 2. Scores for ESM education levels

Total points				
	0	1	2	3
Valid	22	36	54	48
Mean	12.545	12.111	12.500	13.375
Std. Deviation	3.113	2.550	2.369	2.856
Minimum	7.000	7.000	8.000	9.000
Maximum	21.000	18.000	20.000	21.000

general education and the biased sampling method (0 levels of ESM education participants were mainly recruited online, while educated participants came from the same university courses) would render any statistical inference invalid.

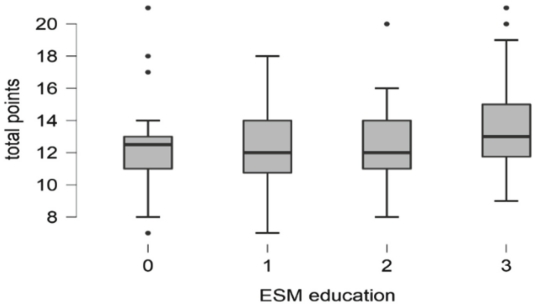

Fig. 2. Boxplot representation of scores divided by ESM education level.

The following tables (Tables 3., 4., 5., and 6.) represent descriptive statistics of each item. Many items' mean scores are approximately equal to the expected value obtained in the case of random guessing (in bold italics in the following Tables). This result suggests the possibility that many people could have answered randomly or did not clearly understand the questions. Additionally, there is some evidence in descriptive statistics that suggests a regression toward the point in the distribution that represents randomness. This highlights items that cannot be used to measure the intended construct. The problem might be related to the difficulty in understanding the questions or in the absence of any reference to the characteristics of the sample group of participants with which they must compete. Below we can see the mean points near the casuality, which for some items are 0.50 and for the Likert scales are near 0.14.

Table 3. Point means and standard deviations

	1.1	2.1	3.1	4.1	5.1	7.1	8.1	9.1	10.1	11.1	12.1
Mean	0.23	0.82	0.57	0.91	0.85	0.93	***0.21***	0.11	0.27	0.28	***0.58***
Std. Deviation	0.42	0.38	0.49	0.28	0.36	0.26	0.36	0.31	0.45	0.40	0.50

Table 4. Point means and standard deviations

	1.2	2.2	3.2	4.2	5.2	6.2	7.2	9.2	10.2	11.2	12.2
Mean	0.37	0.28	***0.19***	0.25	***0.16***	0.20	0.39	***0.19***	0.43	***0.17***	***0.47***
Std. Deviation	0.43	0.41	0.35	0.40	0.32	0.37	0.39	0.39	0.50	0.34	0.50

It is also worth noting that all the items result non-normally distributed following supporting evidence provided by the Shapiro-Wilk test (all $ps < 0.001$). The same applies to the distribution of the total scores (general SQ score) by every participant, showed below in the Fig. 3.

Table 5. Point means and standard deviations

	1.3	2.3	3.3	4.3	5.3	7.3	8.3	9.3	10.3	11.3	12.3
Mean	0.28	0.25	0.26	***0.19***	0.24	***0.18***	***0.15***	0.20	***0.53***	0.22	0.24
Std. Deviation	0.38	0.40	0.43	0.35	0.39	0.31	0.32	0.36	0.50	0.36	0.40

Table 6. Point means and standard deviations

	6.4	9.4	10.4
Mean	0.28	***0.19***	0.53
Std. Deviation	0.38	0.36	0.50

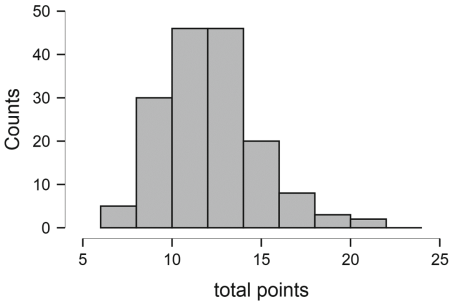

Fig. 3. Distribution of the total scores

The distribution of total scores is positively skewed, suggesting that people with greater strategic reasoning and who use that capability in a fruitful way are distributed in the right tail.

Scores and Measures of Strategic Quotient and Related Constructs
Based on a theory-driven approach we divided the indicators as follows:

- LC = Logic related competence (1.1, 2.1, 3.1, 4.1, 5.1, 7.1, 8.3, 9.1, 10.2)
- hRT = Hyper-rational type (1.1, 3.1, 6.2, 6.4, 9.2)
- eLC = Estimation of LC (1.2, 2.2, 3.2, 4.2, 5.2, 7.2, 8.1, 9.3, 10.1, 11.1, 12.1)
- eeC = Estimation of estimations capabilities (1.3, 2.3, 3.3, 4.3, 5.3, 7.3, 8.1, 9.4, 10.4, 11.2, 11.3, 12.2, 12.3)

3.4 Exploratory Factor Analysis

Since some games were slightly changed along with the underlying dimensions, we chose to explore the latent structure of data with an exploratory factor analysis. The Bartlett test p-value ($X2 = 924.13$, $df = 703$, $p < 0.001$) allowed us to have supporting evidence for the feasibility of the factor analysis.

Two factors emerged using the parallel analysis ($X2 = 714.67$, $df = 628$, $p < 0.009$), differently from what we theoretically assumed. It is worth noting that almost every item has a very high uniqueness (Table 8.), with some that are very close to one and correlating with no factor at all (in bold), leading to several considerations. The first is that some of the items could be subject to elimination in future versions, considering both uniqueness and average response near random guessing. On the other hand, since strategy can be composed of many different competencies and the games presented could require different combinations of them, a model with more factors or considering many items as measuring different abilities might be a possibility that we will try to test in the future if supported by data. This seems to us a more reasonable option for most of the games considered.

The logic-related ability (Carpenter et al., 2013) seems to be somewhat distinguishable from the other factor, but the difference should be better understood. These two are substantially uncorrelated (Table 7.), which confirms that logic-related ability and the other competences (which might be considered estimation skills) pertain to different latent factors. Since these are higher-order skills, it is likely to find items related to multiple factors.

Table 7. Factor Correlations

	Factor 1	Factor 2
Factor 1	1.000	0.095
Factor 2	0.095	1.000

Furthermore, typical questions with respect to risk aversion and trust refer to preferences and components of social order, so they are not just pure "calculation" skills (Chen et al. 2017). This is another point that will need to be developed further to reconcile theory and empirical results. To date, the two-factor model looks the best when looking at the data.

Collectively, these findings show that the design of the mcSQ needs additional iterations to satisfy psychometric validity.

Table 8. Factor Loadings over a threshold of 0.15.

	Factor 1	Factor 2	Uniqueness
V0.1		0.390	0.846
V0.2		0.284	0.919
V1.1	0.314		0.891
V1.2	0.275	0.175	0.889
V1.3			**0.973**
V2.1		0.191	**0.961**
V2.2			**0.990**
V2.3	0.160		**0.970**
V3.1	0.373		0.861
V3.2			**0.975**
V3.3		−0.166	0.971
V4.1		0.253	0.914
V4.2	−0.284	0.488	0.694
V4.3			**0.989**
V5.1	0.176	0.405	0.798
V5.2	0.160	0.160	0.946
V5.3			**0.996**
V6.2	0.236		0.941
V6.4			**0.990**
V7.1		0.214	0.945
V7.2		0.468	0.778
V7.3			**0.972**
V8.1			**0.998**
V8.3	0.204	0.310	0.857
V9.1	0.548		0.678
V9.2	0.656		0.557
V9.3			**0.975**
V9.4		0.243	0.941
V10.1	0.224	−0.203	0.913
V10.2			**0.990**
V10.3	0.235		0.930
V10.4			**0.987**
V11.1			**0.990**
V11.2		−0.280	0.911
V11.3			**1.000**
V12.1	−0.266		0.923
V12.2	−0.227		0.948
V12.3	0.151		0.977

Note. Applied rotation method is promax

4 Discussion and Conclusion

The latest SQ version, the mcSQ, is a test that requires multiple-choice responses to estimate the behavior of others. It builds upon the SQ and the strategic quotient concept in evaluating an individual's rationality, ability to estimate others' rationality, and ability to predict others' predictions. Thus, they are designed to measure an individual's ability to analyze and evaluate information, formulate and implement strategies, solve problems, and adapt to changing situations. However, the particular emphasis on skill estimation and action prediction makes them a unique and useful kind of test for those seeking to assess leadership potential in highly interacting contexts. The tests are particularly useful for assessing the ability to evaluate the competencies that are important in leadership or management roles. They can also be used to identify an individual's areas of strength and weakness and develop a personalized development plan and used to evaluate the effectiveness of training and development programs. While certainly the development and the refinement of the mcSQ is still a work in progress, the considerations that have been made on its strengths and weaknesses shed important light on the constructs of the test and its complexity.

Firstly, the underlying constructs of the SQ might be reconsidered. By re-designing the test with a theory-driven focus, one could expect to better measure strategic skills. However, the psychometric properties of the mcSQ are not yet acceptable, so further development needs to be implemented. Items need to be modified through subsequent evidence-based design-test cycles to improve internal coherence, item uniqueness and obtain better factor loadings.

Secondly, the mcSQ introduces a discrete multiple-choice response mode, that should help participants in anchoring the evidence accumulation for each alternative and choose the identified answer in a more time-efficient way and possibly with lower cognitive cost.

Thirdly, both the wording and the very heterogeneous structure of the SQ test might be improved. We designed mcSQ for this purpose by using the same three-part structure for the items of all games and simplified the wording on many items.

As mentioned, the development and refining process is still in progress and requires further research.

4.1 Limitations and Future Directions for Research

While the concept of strategic quotient has garnered attention and research interest, there are several criticisms, limitations, and gaps in the literature. These include:

Lack of Consensus. Different studies employ varying definitions of strategic ability and methodologies and measures to measure it, making it challenging to compare findings across studies and establish a standardized framework for the concept.

Contextual Specificity. Many studies examining strategic skills focus on specific games, such as the "Dictator Game". However, it remains unclear whether the findings and conclusions derived from these specific contexts can be generalized to different or real-world strategic decision-making scenarios.

Limited External Validity. Most research on the strategic quotient has been conducted in voluntary samples, which may not fully capture the complexity and nuances of real-life strategic interactions. This raises concerns about the external validity and applicability of findings to real-world situations.

Lack of Longitudinal Studies. The existing literature primarily relies on cross-sectional data, providing limited insight into the stability, development, and potential changes in strategic quotient over time. Longitudinal studies tracking individuals' strategic quotient across different stages of life or in response to interventions are needed to better understand its dynamics.

Relationship with Other Constructs. We need to explore the relationship between strategic quotient and other cognitive or social constructs, such as intelligence or social intelligence, the nature of these relationships is still not well understood. Further research is needed to clarify the interplay between strategic quotient and related constructs.

Cultural and Individual Differences. Principal strategic quotient research has been conducted in Western cultural contexts with WEIRD samples, raising questions about its universality and applicability across diverse cultural backgrounds. In some literature, there is evidence that different countries use different strategic approaches, so exploration of cultural effects is due.

Practical Applications. Although there is potential for strategic quotient to have practical applications in various domains, such as education, business, and policymaking, the literature has not yet fully explored this opportunity.

Addressing these criticisms, limitations, and gaps in the literature will be crucial for further advancing the understanding of strategic quotient and its implications in various fields.

References

Alaoui, L., Penta, A.: Endogenous depth of reasoning. Rev. Econ. Stud. **83**(4), 1297–1333 (2016). https://doi.org/10.1093/restud/rdv052

Alos-Ferrer, C., Buckenmaier, J.: Cognitive sophistication and deliberation times. SSRN Electron. J. (2018). https://doi.org/10.2139/ssrn.3218928

Athanasiou, E., London, A.J., Zollman, K.J.S.: Dignity and the value of rejecting profitable but insulting offers. Mind **124**(494), 409–448 (2015). https://doi.org/10.1093/mind/fzu186

Bayer, R.C., Renou, L.: Logical abilities and behavior in strategic-form games. SSRN Electron. J. (2012). https://doi.org/10.2139/ssrn.1743515

Bilancini, E., Boncinelli, L., Mattiassi, A.: Assessing actual strategic behavior to construct a measure of strategic ability. Front. Psychol. **9**, 2750 (2019). https://doi.org/10.3389/fpsyg.2018.02750

Camerer, C.F., Ho, T.-H., Chong, J.K.: A psychological approach to strategic thinking in games. Curr. Opin. Behav. Sci. **3**, 157–162 (2015). https://doi.org/10.1016/j.cobeha.2015.04.005

Camerer, C.F., Ho, T.-H., Chong, J.-K.: A cognitive hierarchy model of games. Q. J. Econ. **119**(3), 861–898 (2004). https://doi.org/10.1162/0033553041502225

Capra, C.M.: Mood-driven behavior in strategic interactions. Am. Econ. Rev. **94**(2), 367–372 (2004). https://doi.org/10.1257/0002828041301885

Carpenter, J., Graham, M., Wolf, J.: Cognitive ability and strategic sophistication. Games Econom. Behav. **80**, 115–130 (2013). https://doi.org/10.1016/j.geb.2013.02.012
Chen, Y.-H., Chen, Y.-C., Kuo, W.-J., Kan, K., Yang, C.C., Yen, N.-S.: Strategic motives drive proposers to offer fairly in ultimatum games: an fMRI study. Sci. Rep. **7**(1), 527 (2017). https://doi.org/10.1038/s41598-017-00608-8
Dhir, S., Dhir, S., Samanta, P.: Defining and developing a scale to measure strategic thinking. Foresight **20**(3), 271–288 (2018). https://doi.org/10.1108/FS-10-2017-0059
Gavetti, G.: PERSPECTIVE—toward a behavioral theory of strategy. Organ. Sci. **23**(1), 267–285 (2012). https://doi.org/10.1287/orsc.1110.0644
Gill, D., Knepper, Z., Prowse, V.L., Zhou, J.: How cognitive skills affect strategic behavior: cognitive ability, fluid intelligence and judgment. SSRN Elect. J. (2023). https://doi.org/10.2139/ssrn.4465561
Gill, D., Prowse, V.: Cognitive ability, character skills, and learning to play equilibrium: a level-k analysis. J. Polit. Econ. **124**(6), 1619–1676 (2016). https://doi.org/10.1086/688849
Golman, R., Bhatia, S., Kane, P.B.: The dual accumulator model of strategic deliberation and decision making. Psychol. Rev. **127**(4), 477–504 (2020). https://doi.org/10.1037/rev0000176
Guala, F., Mittone, L.: Paradigmatic experiments: the dictator game. J. Socio-Econ. **39**(5), 578–584 (2010). https://doi.org/10.1016/j.socec.2009.05.007
Deary, I.J.: Intelligence. Annu. Rev. Psychol. **63**(1), 453–482 (2012). https://doi.org/10.1146/annurev-psych-120710-100353
Jelenc, L., Swiercz, P.M.: Strategic thinking capability: conceptualization and measurement. SSRN Electron. J. (2011). https://doi.org/10.2139/ssrn.2747927
Kahneman, D.: Maps of bounded rationality: psychology for behavioral economics. Am. Econ. Rev. **93**(5), 1449–1475 (2003). https://doi.org/10.1257/000282803322655392
Kvam, P.D., Marley, A.A.J., Heathcote, A.: A unified theory of discrete and continuous responding. Psychol. Rev. **130**, 368 (2022)
Otero, I., Salgado, J.F., Moscoso, S.: Cognitive reflection, cognitive intelligence, and cognitive abilities: A meta-analysis. Intelligence **90**, 101614 (2022). https://doi.org/10.1016/j.intell.2021.101614
Parales-Quenza, C.J.: Astuteness, trust, and social intelligence. J. Theory Soc. Behav. **36**(1), 39–56 (2006). https://doi.org/10.1111/j.1468-5914.2006.00295.x
Stahl, D.O., Wilson, P.W.: On players′ models of other players: theory and experimental evidence. Games Econom. Behav. **10**(1), 218–254 (1995). https://doi.org/10.1006/game.1995.1031
Straub, P.G., Murnighan, J.K.: An experimental investigation of ultimatum games: information, fairness, expectations, and lowest acceptable offers. J. Econ. Behav. Organ. **27**(3), 345–364 (1995). https://doi.org/10.1016/0167-2681(94)00072-M
Suleiman, R.: Economic harmony—a rational theory of fairness and cooperation in strategic interactions. Games **13**(3), 34 (2022). https://doi.org/10.3390/g13030034
Tisserand, J.-C., Cochard, F., Gallo, J.L.: Altruistic or strategic considerations: a meta-analysis on the ultimatum and dictator games ∗ (2015)
Van Dijk, E., Vermunt, R.: Strategy and fairness in social decision making: sometimes it pays to be powerless. J. Exp. Soc. Psychol. **36**(1), 1–25 (2000). https://doi.org/10.1006/jesp.1999.1392
von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition): Princeton University Press (2007). https://doi.org/10.1515/9781400829460
Wellman, H.M., Cross, D., Watson, J.: Meta-analysis of theory-of-mind development: the truth about false belief. Child Dev. **72**(3), 655–684 (2001). https://doi.org/10.1111/1467-8624.00304
Weyhrauch, W.S.: A mindset for strategic thinking: Developing a concept and measure (2016)

Weyhrauch, W.S., Culbertson, S.S.: A mindset for strategic thinking: conceptual synthesis of the capacity for strategic insight. Acad. Manag. Proc. **2014**(1), 12988 (2014). https://doi.org/10.5465/ambpp.2014.12988abstract

Xiao, E., Houser, D.: Emotion expression in human punishment behavior. Proc. Natl. Acad. Sci. **102**(20), 7398–7401 (2005). https://doi.org/10.1073/pnas.0502399102