opinion

Challenges in the Measurement of Gambling Product Risk:
A Critical Review of the ASTERIG Assessment Tool

Paul Delfabbro¹ & Jonathan Parke²

¹ School of Psychology, University of Adelaide, Adelaide, Australia
² Sophro Ltd., Manchester, England, United Kingdom & Northern Ireland

Introduction

There is a strong consensus that the structural characteristics of gambling products play a significant role in facilitating risky gambling behaviour and harm to individuals (Dow Schull, 2012; Griffiths, 1993; Parke & Griffiths, 2007). Prevalence studies suggest that problem or disordered gambling is generally higher in communities when people have ready access to continuous forms of gambling¹ that enable more rapid and repeated gambling opportunities (Storer et al., 2009). Although individual differences, including underlying co-morbidities, are recognized as risk factors for gambling disorder (Blaszczynski & Nower, 2002), it is thought that the disorder develops or is maintained as a result of the complex interaction between product characteristics and individual vulnerabilities (Delfabbro & King, 2019; Korn & Shaffer, 1999; Parke et al., 2016). In particular, features or products that attract greater expenditure or generate an experience that make it difficult to stop gambling have the potential to increase gambling harm. The importance of structural characteristics is therefore recognized in mainstream psychological research that looks for the factors that contribute to excessive gambling. At the same time, this topic has emerged to an increasingly strong extent in public health or harm-focused perspectives in which the modification of the riskier features or gambling products (e.g., bet sizes) is considered integral for harm minimization (Livingstone et al., 2017).

These arguments are generally supported by empirical evidence that shows significant differences between products in relation to their capacity to cause harm (Delfabbro et al., in press). Slot machines and electronic gaming machines (EGMs) appear to have a much stronger association with problem gambling or harm than other activities do (Delfabbro & Parke, 2020). This is borne out in multiple peer-reviewed studies that allow for comparisons between products (Afifi et al., 2010, 2014; DeFuentes-Merillas

¹A continuous product is one in which there is a short gap between outcomes and the next gamble, allowing people to gamble more quickly (e.g., a slot machine is more continuous than a draw lottery).
et al., 2003; Petry, 2003; Scalese et al., 2016), in major prevalence studies with multivariate analyses (ACIL Allen Consulting, 2014, 2018; The Allen Consulting Group, 2011; Browne et al., 2019; Orford et al., 2013; Woods et al., 2018), and in data drawn from treatment-seeking populations (Delfabbro & Wallace, 2020; GamCare, 2019; Ministry of Health, 2006). For example, Australian, New Zealand, and UK data all suggest that EGMs are mentioned as the principal cause of problems by 60%–70% of gamblers in general population and in treatment populations. Wagering is mentioned by 20%-30%, whereas lottery products are identified by less than 2% of respondents (Delfabbro & Wallace, 2020; Ministry of Health, 2006). These observations have, not surprisingly, led to requests for regulatory or public health responses that limit people’s access to activities such as EGMs (Delfabbro & Eltridge, 2008) or for modifications to the machines themselves (e.g., the introduction of lower value stakes or slowing of play speeds; Livingstone et al., 2017).

Much of this evidence is consistent with the view that the supply or regulation of gambling products (type, volume, design, availability, and accessibility) could play an important role in reducing the harms associated with gambling. As noted by the Productivity Commission (1999), regulators have the ability to determine which products are allowed onto the market, their distribution, and the mode of access (e.g., which venues, online or offline, electronic payments vs. cash). Regulators also have oversight over the evolution of gambling, including the potential effect of new product features, new classes of product, or changes in the ways that the games are promoted or made available to players. The ultimate aim of regulation is to prevent or reduce harm wherever possible, but also to do so in a way that is fair and proportionate. Effective gambling regulation, it is argued, should try to strike a balance between the recreational benefits of gambling to the majority of lower risk gamblers, while trying to target changes or interventions toward products, features, or measures that will reduce the harms for vulnerable gamblers (Productivity Commission, 1999).

One increasingly common way in which regulators have been guided in their decision making concerning the potential risk or harm associated with certain products or features has been to conduct formal risk assessments (Blanco et al., 2013). Prior to new gambling products being introduced into a market, they may be appraised for the extent to which they are likely to contribute to an increase in problem gambling or harm. For example, in some jurisdictions in Australia, this can be based on an inquisitorial-based approach in which the industry has to defend a new product or feature (its risk and potential mitigation strategies) in open hearings or by submission. Other jurisdictions (e.g., some countries in the European Union) may require the product to be assessed by using one of a small number of structured risk-assessment tools that rate products on the basis of their potential risk for harm. The demand from charities, legislators, and academics to consider and restrict potentially harmful games has also increased. Only recently, for example, the charity Gambling with Lives has argued for the introduction of a product safety “kitemark”\footnote{This point was made as part of oral evidence given to the Select Committee on the Social and Economic Impact of the Gambling Industry (https://committees.parliament.uk/oralevidence/101/html/).} to assess
the risks posed by new games before these games are introduced into the market. Likewise, in the United Kingdom, a House of Lords Select Committee conducted a review into gambling harm and recommended that the Gambling Commission, the regulator for Great Britain, establish “a system for testing all new games against a series of harm indicators, including their addictiveness and whether they will appeal to children. A game which scores too highly on the harm indicators must not be approved” (House of Lords Select Committee on the Social and Economic Impact of the Gambling Industry, 2020, p. 50).

**Assessment Tool to Measure and Evaluate the Risk Potential of Gambling Products (ASTERIG): Application and Scoring**

One of the tools that has been explicitly identified in the academic literature as an option for assessing product safety and that is publicly available for use is ASTERIG. Developed initially in Germany by Peren (2010), this tool was further refined with an expert consensus or Delphi process, as described by Blanco et al. (2013). The tool comprises 10 criteria that are intended to capture the unique dimensions of risk for any gambling activity. Risk scores for each criterion are then aggregated to form an overall assessment of risk for that product. Each criterion is scored on a scale of 0 to 10 and these scores are then weighted (total weight values) to develop a final score that can range from 0 to 620 (Table 1). The different weightings are intended to capture the relative importance of the criterion in increasing risk. A total score of 1–124 = lowest risk, 125–248 = low risk, 249–372 = moderate risk, 373–496 = high risk, and 497–620 = very high risk.

Table 2 provides a summary of the criteria used in ASTERIG. The number of scaling points differ for each criterion. For example, event frequency and interval of payback have 9 anchor points, jackpot size and chance of winning a profit have 8, and availability has 6, but there are some with only 3 anchor points.
### Table 2
ASTERIG Criteria Definitions and Scaling

<table>
<thead>
<tr>
<th>Risk criteria</th>
<th>Definition provided</th>
<th>Scale range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event frequency</td>
<td>Unit of time between stake, gambling result, and next stake opportunity</td>
<td>0 = &gt; 6 days to 10 = ≤ 5 s</td>
</tr>
<tr>
<td>Interval of payback</td>
<td>Period between gambling result and notification of payment or receipt of payment</td>
<td>0 = &gt; 6 days to 10 = ≤ 5 s</td>
</tr>
<tr>
<td>Jackpot size</td>
<td>An extraordinary top prize</td>
<td>0 = none to 10 = ≤$1 million</td>
</tr>
<tr>
<td>Continuity of playing</td>
<td>Period of time when it is possible to gamble without interruption</td>
<td>0 = ≤ 5 min to 10 = &gt; 3 h of continuous gambling</td>
</tr>
<tr>
<td>Chance of winning a profit</td>
<td>The probability of realizing a profit with each game</td>
<td>0 = 0% to 10 = &gt; 25%</td>
</tr>
<tr>
<td>Availability</td>
<td>Possibility of accessing gambling opportunities</td>
<td>0 = &gt; 100-km radius to 10 = home/workplace</td>
</tr>
<tr>
<td>Multiple playing or stake</td>
<td>Opportunity to play several stakes at the same time or to take part in several gambling opportunities at the same time</td>
<td></td>
</tr>
<tr>
<td>opportunities</td>
<td></td>
<td>0 = none, 5 = multiple playing opportunities or multiple stake opportunities, 10 = multiple playing opportunities and multiple stake opportunities</td>
</tr>
<tr>
<td>Variable stake amount</td>
<td>Extent to which gamblers can choose or modify their stake amounts while playing</td>
<td>0 = fixed, 5 = variable but limited stake amount, 10 = variable but unlimited stake amount</td>
</tr>
<tr>
<td>Sensory product design</td>
<td>Auditory and visual effects</td>
<td>0 = non-existent, 5 = auditory or visual effects exist, 10 = auditory and visual effects exist</td>
</tr>
<tr>
<td>Near wins</td>
<td>Results when a gambler comes close to obtaining a desired win</td>
<td>0 = non-existent, 5 = unintentionally created or occurring by chance, 10 = intentionally created, occurring more frequently than random</td>
</tr>
</tbody>
</table>

Insights into how ASTERIG scores products can be obtained by applying it to different products that empirical evidence would generally locate at different positions on the risk continuum. The first product, positioned at the highest risk end (Delfabbro et al., in press), would be a high-intensity EGM\(^3\) that has a spin interval of around 2–3 s, a maximum stake of $10 per spin, maximum prizes of $10,000, multiple play lines, many auditory and visual effects, and an 85% return to the player; it is also accessible at a local high street or community venue (e.g., local club or hotel in Australia or New Zealand). A second product (which we would consider to be lower risk) would be a $5 fixed-stake scratch ticket with a $100,000 maximum prize and a theoretical return-to-player value (RTP) of 70%; it would take around 3 min to scratch off the panels. A third product is one in which players play an instant-win game online where the game takes under 2 min to complete. A fourth product is a weekly draw lottery with a million-dollar prize that is $5 per ticket, but where one can purchase more than one set of numbers for the same draw. As indicated earlier, empirical evidence generally shows that this a low-risk product that has little (and sometimes a negative) association with problem gambling.

We have scored these products by using ASTERIG and summarized them in Table 3. On the whole, scoring EGMs is relatively simple, although two criteria are particularly problematic in interpreting definitions: continuity of play and chance of winning a profit. For illustrative purposes, we have chosen values that would be considered atypical or more extreme based on empirical evidence; we have assumed 3 or more hours of EGM playing, 1 hr of scratch card play, and between 5 and 30 min to fill out lottery draw tickets.\(^4\) The second, and potentially more difficult, criterion to complete is the chance of making a profit. In the ASTERIG guidelines, this is written as if based on a prize pool, but most games will be defined in terms of the RTP or the probability of winning more than the stake value. To apply ASTERIG to the EGM and scratch card, we would have assumed a value of more than 25% for any single game. For the draw lottery, we would assume that the odds are very low (a zero score).

The results in Table 3 generally show that ASTERIG is able to discriminate between lower risk and higher risk products (EGMs vs. lottery products). The EGM is classified as very high risk, whereas the draw lottery is classified as low risk. This is reassuring, but we would argue that this is hardly a discrimination that would require

---

\(^3\)A high-intensity gaming machine would be one that has a fast play speed (e.g., a 2- to 3-s spin) and that allows substantial bets to be made (e.g., $5 per bet in Australia or New Zealand would meet this definition). A Category A machine in the United Kingdom would be another example. A low-stakes machine (even if fast, e.g., < 3-s spins) would not be considered high intensity by this definition.

\(^4\)The fact that we have to assign values to populate the examples is the same situation faced by regulators or companies that are asked to rate their products against the ASTERIG criteria. What value should be assigned? No guidance is provided as to whether this should be a normative value (typical gambler), the value for a high-risk gambler or person with gambling problems, or merely the worst-case scenario. For illustrative purposes, we have included high values on the basis of examples that we have observed in studies of EGMs or in countries where ASTERIG is used and where prevalence data indicate typical and atypical levels of engagement in lottery products.
Table 3

**ASTERIG Scoring Estimates for Standard Products**

<table>
<thead>
<tr>
<th>Risk criterion</th>
<th>Electronic gaming machine</th>
<th>Online instant win</th>
<th>Scratch card</th>
<th>Draw lottery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event frequency</td>
<td>80</td>
<td>60</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Interval of payback</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Jackpot size</td>
<td>30</td>
<td>30</td>
<td>37.5</td>
<td>50</td>
</tr>
<tr>
<td>Continuity of playing</td>
<td>80(^a)</td>
<td>80(^a)</td>
<td>40(^a)</td>
<td>20(^a)</td>
</tr>
<tr>
<td>Chance of winning a profit</td>
<td>60(^a)</td>
<td>45</td>
<td>60(^a)</td>
<td>0</td>
</tr>
<tr>
<td>Availability</td>
<td>52.5</td>
<td>70</td>
<td>52.5</td>
<td>52.5</td>
</tr>
<tr>
<td>Multiple playing/stake opport.</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Variable stake amount</td>
<td>30(^b)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sensory product design</td>
<td>40</td>
<td>20</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Near wins</td>
<td>60</td>
<td>30</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>522.5</td>
<td>395</td>
<td>330</td>
<td>212.5</td>
</tr>
</tbody>
</table>

Highest risk         | High risk     | Moderate risk  | Low risk     |

Note. ASTERIG = Assessment Tool to Measure and Evaluate the Risk Potential of Gambling Products.

\(^a\)Estimate provided on the basis of a study by Delfabbro et al. (2007), which examined the lower threshold of duration of play considered to be a problem gambling indicator but not clearly defined by ASTERIG. We have assumed an extreme value for instant wins, in the absence of any research indicating the typically duration of this activity. Evidence from prevalence surveys (e.g., Department of Health, 2015) suggests that people usually purchase only one to two lottery or scratch cards, so that our estimates illustrate what might be considered a more extreme or harmful level of engagement in these activities (irrespective of whether such levels occur very often). \(^b\)Assuming no illegal playing of more than one machine.

a complex tool. Indeed, it is likely that any layperson who is familiar with both activities might make the same classification. The other classifications, however, we believe to be less satisfactory. A basic scratch card (even when we assume only an hour of play, which is likely to be an overestimate) is classified as moderate risk and toward the upper part of the range. This does not accord with our review of the empirical evidence (Delfabbro, & Parke, 2021, submitted), which identifies scratch cards as a generally low-risk activity that is rarely associated with problem gambling. Indeed, when scratch-card style games are transferred to an online environment (in the form of an instant-win game) it becomes possible to score these lottery games as high risk if one follows the ASTERIG article’s guidance that encourages scoring all items at the highest level (e.g., that a person might play instant-win games as long as they like).

**Difficulties With Applying ASTERIG**

We believe that ASTERIG suffers from five principal limitations. Four of these (Limitations 1–4) have a direct impact on scoring and may explain some of the difficulties in reliability and validity that we have encountered in our own application of the criteria. A complete summary of these limitations is provided in Table 4, but they are elucidated in the sections that follow. Table 4 also provides some additional discussion of issues and examples to elaborate on the points outlined in the text.
### Table 4

**Critical Appraisal of ASTERIG Criteria**

<table>
<thead>
<tr>
<th>Risk criterion</th>
<th>ASTERIG definition</th>
<th>Validity of construct/risk related (Concern 1)</th>
<th>Clarity of definition (Concern 2)</th>
<th>Scoring and calibration (Concern 3)</th>
<th>Collinearity (Concern 4)</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event frequency</td>
<td>Unit of time between stake, gambling result, and next stake opportunity</td>
<td>Strong</td>
<td>Inadequate</td>
<td>Inadequate</td>
<td>Continuity; interval of payback</td>
<td>Fast games with short intervals between games are likely to increase risks because (a) money can be lost faster; (b) they provide better conditions for mood modification and dissociation (Parke et al., 2019; Stewart &amp; Zack, 2008); and (c) they can impair decision making and/or self-control (Corr &amp; Thompson, 2014; Thompson &amp; Corr, 2013). Notably, the interdependence of speed and continuity as a risk factor should require their combination as one dimension of risk.</td>
</tr>
<tr>
<td>Continuity of playing</td>
<td>Period of time when it is possible to gamble without interruption</td>
<td>Strong</td>
<td>Inadequate</td>
<td>Inadequate</td>
<td>Event frequency; interval of payback</td>
<td></td>
</tr>
<tr>
<td>Interval of payback</td>
<td>Period between gambling result and notification of payment or receipt of payment</td>
<td>Weak</td>
<td>N/A</td>
<td>Event frequency; continuity</td>
<td>Event frequency; continuity</td>
<td>Where there may be a delay in payment following notification of winning, gambling may continue because people have access to other money to continue gambling until receiving payment. Where payment delays do result in an unwanted break in gambling, this should be</td>
</tr>
</tbody>
</table>
### Table 4 Continued.

<table>
<thead>
<tr>
<th>Risk criterion</th>
<th>ASTERIG definition</th>
<th>Validity of construct/risk related (Concern 1)</th>
<th>Clarity of definition (Concern 2)</th>
<th>Scoring and calibration (Concern 3)</th>
<th>Collinearity (Concern 4)</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chance of winning a profit</td>
<td>The probability of realizing a profit with each game</td>
<td>Weak</td>
<td>N/A</td>
<td>Jackpot size</td>
<td></td>
<td>accounted for within the “continuity” risk criterion to avoid any duplication when accounting for risk.</td>
</tr>
<tr>
<td>Jackpot size</td>
<td>An extraordinary top prize typically in the form of a</td>
<td>Weak</td>
<td>N/A</td>
<td>Chance of winning a profit</td>
<td></td>
<td>This concept is inadequately defined, rendering its relationship to risk unclear. Although the experience of winning is likely to be related to risk, it may be better considered over a series of outcomes or a session rather than by each bet or game. The relationship between winning and risk will vary depend on a number of factors, including the type of game, previous outcomes, and individual experience and involvement. The reference to profit rather than winning seems inaccurate given that in the overwhelming majority of circumstances, players will not make a profit but are considered to be paying for entertainment. Size of profit is not considered. Risk may be more related to chasing losses and financial reparation than to generating profit.</td>
</tr>
</tbody>
</table>

Larger prizes may pose greater risks because they are more exciting (Wulfert et al., 2005, 2008), may
Table 4 Continued.

<table>
<thead>
<tr>
<th>Risk criterion</th>
<th>ASTERIG definition</th>
<th>Validity of construct/risk related (Concern 1)</th>
<th>Clarity of definition (Concern 2)</th>
<th>Scoring and calibration (Concern 3)</th>
<th>Collinearity (Concern 4)</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability</td>
<td>Possibility of accessing gambling opportunities</td>
<td>Strong</td>
<td>Inadequate</td>
<td>Inadequate</td>
<td></td>
<td>prompt higher stakes (Mentzoni et al., 2012), and may justify loss-chasing behaviour (Parke et al., 2016). However, this may not be a linear relationship, because of the need to account for win probability. A moderate prize of 100–500 times the stake may be sufficient to pose significant risks by facilitating excitement or chasing. For example, a 1 in 14,000,000 chance of winning an extremely large lottery jackpot may not provide the maximum excitement or the strongest incentive to chase losses. Although greater availability and accessibility has been established as a potential risk factor (Ofori Dei et al., 2020; Sharman et al., 2019), only one dimension of risk is considered by ASTERIG: distance to travel to the gambling opportunity. Other dimensions should be included, for example, opening hours and ease of use (Abbott, 2007), privacy (Hing &amp; Haw, 2010), or unintended exposure to uninitiated or vulnerable populations (Smith et al., 2019).</td>
</tr>
<tr>
<td>Risk criterion</td>
<td>ASTERIG definition</td>
<td>Validity of construct/risk related (Concern 1)</td>
<td>Clarity of definition (Concern 2)</td>
<td>Scoring and calibration (Concern 3)</td>
<td>Collinearity (Concern 4)</td>
<td>Explanation</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------</td>
<td>-------------------------------------</td>
<td>-------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Multiple playing/stake opportunities</td>
<td>Opportunity to play several stakes at the same time or to take part in several gambling opportunities at the same time</td>
<td>Moderate</td>
<td>Inadequate</td>
<td>Inadequate</td>
<td></td>
<td>Theoretical grounds for gambling concurrently on multiple gambling activities might include the potential increase in cost of play, potential increase in excitement, and possible negative implications for decision making. However, the ability to play games concurrently is limited to only a few products (e.g., poker, EGMs) and constitutes only one dimension of cost of play, the most significant of which (stake size) is not considered.</td>
</tr>
<tr>
<td>Variable stake amount</td>
<td>Extent to which gamblers can choose or modify their stake amounts while playing</td>
<td>Moderate</td>
<td>Inadequate</td>
<td>Inadequate</td>
<td></td>
<td>Staking variability has been shown to be indicative of disordered gambling (Braverman et al., 2013; Xuan &amp; Shaffer, 2009); however, it constitutes only one dimension of cost of play, the most significant of which (stake size) is not considered.</td>
</tr>
<tr>
<td>Sensory product design</td>
<td>Auditory and visual effects</td>
<td>Moderate</td>
<td>Inadequate</td>
<td>Inadequate</td>
<td></td>
<td>Scoring in ASTERIG is determined based on the presence of sensory design features without consideration of their context. Sensory effects are “secondary effects.” They support other game design features by providing emphasis (winning, bonus games, and free spin features). Risk could be estimated in a more meaningful</td>
</tr>
</tbody>
</table>
Table 4 Continued.

<table>
<thead>
<tr>
<th>Risk criterion</th>
<th>ASTERIG definition</th>
<th>ASTERIG</th>
<th>Clarity of definition (Concern 1)</th>
<th>Scoring and calibration (Concern 3)</th>
<th>Collinearity (Concern 4)</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Near wins</td>
<td>Results when a gambler comes close to obtaining a desired win (to miss the profit narrowly; a near win)</td>
<td>Moderate</td>
<td>Inadequate</td>
<td>Inadequate</td>
<td></td>
<td>way by considering risk in relation to how sound can increase risks by supporting features such as a &quot;loss disguised as a win&quot; (Dixon et al., 2015) or a near win. Near wins may increase risk because they are exciting (Clark et al., 2012; Dixon et al., 2010), may modify mood (Shao et al., 2013), and lead to persistence because of frustration (Amiel 1958; Bossuyt et al., 2014) or uncertainty (Zack et al., 2020). However, evidence supporting this link to harmful outcomes is inconsistent, with several studies failing to find strong support for its impact (Pislak et al., 2019; Sundali et al., 2012; Witts et al., 2015). Notably, if one assumes that near wins are risky, then the current definition and scoring are inadequate because risk is assessed based on intention rather than impact.</td>
</tr>
</tbody>
</table>

Note. ASTERIG = Assessment Tool to Measure and Evaluate the Risk Potential of Gambling Products; EGM = electronic gaming machine; N/A = Not available.
Limitation 1: Relevance and Applicability of ASTERIG Risk Criteria

Of the 10 criteria, only a few were considered to be valid indicators of potential risk (event frequency, continuity, and availability). Other ASTERIG concepts, such as the chance to win a profit, are not intuitively related to risk based on how they currently conceptualized. For example, having a greater chance to win might encourage some people to gamble for longer periods (greater risk), but having less chance of a win might mean that people lose their money more quickly (also greater risk). Thus, it is not clear how scoring in either direction demonstrates higher or lower risk. Chance of profit is likely to be difficult to understand in percentage terms and is probably better understood by using a more “frequentist” or numerical form\(^5\) (see Gigerenzer & Hoffrage, 1995); for most gambling activities, it would be considered over a gambling session rather than a single gamble (e.g., one spin of an EGM). Profit is also conceptually different from winning in that initial outlay must be considered by calculating how much people have spent versus how much they have won. Even in the event that individuals focus on profit rather than size and frequency of wins, the time frame for calculating a profit is also ambiguous. Is the profit determined over the course of playing one game, over a session, over a day, or over a month? Finally, risky gambling behaviour (defined as a higher stake and/or less controlled gambling) is often linked to chasing losses or the attempt to break even (Lesieur, 1984; O’Connor & Dickerson, 2003). These considerations add further doubt over whether profit is the most suitable parameter for assessing risk rather than the probability and size of a win in relation to the size of the stake. In other words, the implication of this confusion is that products will be assigned greater risk if the return is more profitable, but less risk if the activity is less profitable (players are losing more), which does not seem logical unless the intention is to calculate the “incentive to play the game.” However, incentive could also relate to the size of potential wins rather than merely whether the game is more profitable or has a better return to players.

Other criteria, which could be considered at least moderately associated with risk, were otherwise limited by definitions being too narrow or imprecise. For example, sensory product design is defined only by its presence in a game (0, 5, or 10 points, which are then weighted). This means that, if a product comprises lights and sounds of any description, it must be assigned the highest risk score (40/40). There is no guidance regarding how or why visual and auditory effects can influence risk. Accordingly, any digital game will necessarily be assessed as posing greatest risk regardless of the frequency or intensity of these sensory factors, and, more important, the precise role that they play in the gambling experience.

\(^5\)Statements of probability are difficult to understand in pure percentage terms. For example, having a 50% chance of obtaining an outcome is misleading because the event either occurs or does not (0 = no, 1% or 100% = yes). It makes more sense to frame probability statements as the expectancy to spend $10 and get back 50% if 50% is an RTP measure, or as the expectancy to obtain a win in 5 of 10 or 50 of 100 gambling events.
Limitation 2: Scope and Ambiguity of Criterion Definitions

Concerns regarding definitions apply to all ASTERIG risk criteria to varying degrees. For example, ASTERIG’s two staking criteria, *multiple stake* and *variable stake*, could be considered relevant because each may contribute to losing more money (i.e., the financial cost to an individual for engaging in gambling activity). Critically, however, neither criteria account for the *size of stake*, which is a critical determinant of the cost to play. They also do not account for how game speed or the RTP affects financial loss. Moreover, the elements of these criteria are applicable to only a few product categories (e.g., playing multiple poker tables or EGMs concurrently), raising the question of whether concurrent and variable staking should be subsumed into a broader cost-of-play risk dimension.

Similarly, the imprecise characterization of potential risks arising from a *near-win* event weakens the validity of this criterion. ASTERIG rates near-win risks on the basis of design intent rather than strength or the nature of impact. This approach is problematic because near wins may be more influential in determining harmful behaviour in cases in which they have not been deliberately designed into a game (e.g., consider the result of a sporting event being determined by one play made in the final seconds of a game). Equally, engineering a near win is no guarantee of effectiveness. Indeed, artificial near wins may become predictable and less effective over time with repeated exposure (Reid, 1986) or if they are too numerous (Pisklak et al., 2019).

In the case of one of the most important dimensions of risk captured by ASTERIG, speed and continuity of play, there are inaccuracies and duplication in how criteria such as event frequency, continuity, and interval of payback are defined. For example, ASTERIG defines event frequency as the “unit of time between stake, gambling result and the stake opportunity.” However, this definition extends beyond event frequency and strays into continuity because it includes the time interval between the result of the previous gamble and the opportunity to initiate the subsequent gamble: The longer the interval, the less continuous the game. We discuss concerns regarding duplication in risk scoring in more detail later, but this example also raises questions about how the continuity risk criterion is defined: “the period of time during which it is possible to gamble without interruption.” In this definition, interruption is not defined and qualified. What constitutes an interruption? How long is an interruption? Corr and Thompson (2014) reported preliminary evidence that a 5-s interval between gambles is sufficient to reduce risky play by allowing a “period of time for reflection and preparation for a more appropriate response for the next bet opportunity” (p. 897). ASTERIG’s continuity definition may be misinterpreted to imply session duration (the time from the start of the first gamble to the end of the last gamble). To avoid such misapplication, we suggest that the time between the result of the previous bet and the opportunity to place a subsequent bet is a more accurate and useful definition: The shorter the interval, the more continuous the game and the greater the risk score.
**Limitation 3: Inadequate Calibration and Sensitivity of Rating Scales**

For a number of the risk criteria, although they are likely to be associated with risk in principle, they may not adequately detect differences in risk because of how the rating scales are constructed. For example, an event frequency of <5 s assigns the highest risk score of 80 points (for example, an EGM), but any game lasting 3 min is assigned a risk score of 50 points (for example, a scratch card). Thus, an EGM (assuming a 3-s reel spin) is 60 times faster than a scratch card but is only rated as 1.6 times riskier by ASTERIG. Differences in game speeds as brief as a few seconds (Corr & Thompson, 2014) or, as has been shown in some cases, milliseconds (Mentzoni et al., 2012), may have implications for altering the level of risk associated with a game. Accordingly, this criterion needs to be more widely calibrated, with greater sensitivity at the upper end of the scale.

Restrictive scoring boundaries increase the chance of ceiling effects (where risk scores cannot exceed a certain value) and floor effects (where risk scores cannot drop below a certain value). For example, if the upper bound of risk for event frequency is 5 s, it treats the potential risks posed in games lasting 0.5 s as being equivalent to games lasting 4.9 s: This could be described as a ceiling effect. As it is currently scored, there is a danger that risks posed by games that take a few minutes are not adequately distinguished from games lasting a few seconds or less. Distinguishing between a 5-s and a 1-s event frequency may initially seem arbitrary, as both could be viewed as extremely fast. However, consider just one implication of this distinction: The equivalent game operating at the latter speed would incur losses per hour 5 times greater in magnitude.

In addition, risk scoring for some of the criteria use categorical rating scales that, when broken down, offer binary scoring options without the ability to detect nuances in the levels of risk. This applies to near wins (exist: yes or no? intentional: yes or no?), sensory product design (visual or auditory effects exist: yes or no? both exist: yes or no?), variable staking (exists: yes or no? limited or unlimited?), multi-stakes and concurrent play (either exist: yes or no? both exist: yes or no?).

For these reasons, ASTERIG appears to be limited to differentiating between broad categories of activity (between product comparisons; e.g., slot machines vs. lottery products). It is much less useful in being able to make *within* product discriminations of risk. This occurs because the scaling is often simplistic and not all criteria are clearly defined and because of the questionable validity of some criteria. Thus, in

---

6As pointed out by an anonymous reviewer, the extent to which intentional versus non-intentional design of near misses plays a role in gambling has not been a feature of much research (if any). We would argue that the salience of near wins is the most important feature, but that intentionality relates more strongly to the ethics of product design. An activity could have stronger or more salient near-miss effects without any intention (e.g., a three of four match requirements on a slot machine will arguably be a stronger near miss than one involving two of three symbols). If the reels are weighted to make this more common (intentionally), the extent to which this effects behaviour would be based on the greater frequency of the near win, not just because it was intentional.
relation to examples of near wins, we believe that the risk definition should be revised to consider impact rather than intent, and then revised again to capture the extent of impact by using ordinal measurement (e.g., strength of near win: 1–5), rather than the current three nominal categories.

**Limitation 4: The Relationship Between Risk Criteria Not Adequately Considered**

We also have concerns about how each risk criterion is aggregated to provide an overall measure of risk for a product. ASTERIG encourages the independent scoring of attributes, but this strategy can create problems that are generally considered unacceptable and invalid in standard methodological approaches to measuring consumer choice (e.g., conjoint analysis). For example, if a conjoint model were developed to rate cars and this produced a car example that had both high seating capacity and maximum “sportiness,” it would be considered invalid because these qualities would usually be seen to be negatively correlated (Hair et al., 1995). However, ASTERIG encourages scoring of risk criteria independently despite some existing interrelationships between game parameters. For example, the awarding of very large prizes on some games (e.g., lottery tickets) can often be subject to restrictions on how, when, and where prizes can be claimed for verification and security purposes. However, lower tier prizes can usually be claimed more conveniently and expediently. How, then, should a lottery ticket be scored on ASTERIG? In the case of a larger prize, should risk be scored higher on the jackpot criterion but lower on the interval of payback criterion, and, vice versa, for a lower value prize, would jackpot risk be scored lower and payback interval risk higher?

The problem of aggregating ASTERIG risk criteria extends beyond the interrelationships of certain game parameters. As we outlined earlier, there is also a degree of commonality between some criterion definitions, and, in some cases (e.g., event frequency and continuity), clear duplication. As we have noted, ASTERIG identifies continuity and event frequency as separate risk criteria. However, event frequency, as defined by ASTERIG, incorporates a measure of continuity (i.e., the time from the result to the next gambling opportunity) in its definition. Consequently, this dimension of product risk, which has already been given the highest weighting of any criterion, can be inadvertently accounted for twice. For example, when rating a 2- to 3-min-duration scratch card, it is possible for this value to be used for event frequency and continuity without differentiating the constructs. A danger of such double scoring occurs because of ambiguity in criterion definitions. Another example applies to bet type. A game such as blackjack permits multi-games (playing multiple hands concurrently), multi-stakes (placing more than one bet on each hand), and variable stakes (a different size stake for each hand). However, buying multiple sets of numbers in a lottery draw can usually be classed only as multi-stake, not multi-game (it involves only one number draw) or variable stake (usually the price of a lottery ticket is fixed). In our professional experience, failing to accurately make such distinctions can cause confusion and may lead to risk assessment errors in real-world applications.
Limitation 5: Important Dimensions of Risk Omitted

Any risk assessment protocol needs to find the balance between capturing the most relevant dimensions of risk while avoiding the complexity and impracticality of being all-inclusive. However, ASTERIG still disregards critical risk factors, among which the potential for financial loss by accounting for cost of play is arguably one of the most important (e.g., price point or direct cost of play; e.g., a $100 maximum bet slot machine would be not rated differently from one with a $0.50 cost). Moreover, determining the theoretical cost of play is more straightforward when calculating relative risk to other risk criteria. Although ASTERIG does include risk scores regarding the variability and concurrence of staking, stake size is omitted. However, even stake size measured in isolation would be an unreliable indicator of the cost of gambling over any given period. Ideally, a composite measure is required that accounts for all relevant variables (e.g., stake size, speed of play, and RTP; see Parke et al., 2016) and determines how much money an individual can lose playing a game over a particular period.

Discussion and Future Directions

Mitigating product risks is a multi-stage process in which risk assessment is the first step. In making critical regulatory or commercial decisions within safer gambling strategies that focus on product design, it is essential that risks posed by products or its features are identified with an acceptable level of validity. By “validity,” we mean that the risk factor has good face and construct validity (e.g., speed of play appears to be a clear risk factor for greater harm, but less clear is whether having a higher or lower payback percentage is as clearly related to an increased risk of harm). Moreover, it is important that the protocols used to assess risk are subject to critical and ongoing review and empirical testing against independent evidence of product risk (e.g., the association with problem gambling or variations in gambling-related harm). Limitations in protocol design could lead to product risks being over-estimated (often as a result of floor effects) or underestimated (often as a result of ceiling effects).

We argue that the best strategy for assessing product risk involves the development of risk assessment protocols that are validated against empirical research. Such research should be appropriately powered and ecologically valid (i.e., controlled laboratory or experimental studies backed up with data obtained from experienced gamblers playing for real money in genuine gambling environments). Such work may require analysis of industry-sourced data or field experiments in which self-report data collected from individual gamblers (e.g., level of risk or harm) are analysed in relation to objective patterns of behaviour observed for products with varying characteristics. Examples might include slot games with varying cost of play, play speeds, betting options, or levels of product volatility (the distribution or win types; see Turner, 2011). We note the existence of examples of innovative research designs in which industry data have been used to examine the relationships between variations in responsible gambling features and gambling behaviour in online
gambling environments (Auer & Griffiths, 2013, 2019; Auer et al., 2018, 2019, et al., 2020; Nelson et al., 2008). At the same time, we see ongoing merit in controlled studies of specific gambling features (e.g., Dixon et al., 2010, 2015, 2018), as well as epidemiological self-report studies that examine the relationship between variations in product participation and gambling risk or harm (e.g., Afifi et al., 2010; Brosowski et al., 2020).

In this article, we have argued that ASTERIG does not assess product risks with an adequate level of accuracy. By examining each risk criterion and how criteria are combined to provide an aggregate risk score for a product, we have identified shortcomings relating to their validity, designation, and aggregation to yield an overall level of product risk. In advancing such criticisms, however, we recognize that developing a product risk assessment protocol is a complex, resource-intensive, and contentious task. Numerous challenges exist in developing a rigorous and discriminative risk assessment protocol. For example, one principal challenge is that relevant academic literature does not often keep pace with a dynamic industry in which new products and structural features are being continually introduced. It may also be challenging to develop a one-size-fits-all solution that captures endless variations in games and their configurations.

Accordingly, despite the suggested limitations of the ASTERIG risk assessment protocol, we want to also emphasize the important and foundational work that the ASTERIG protocol and associated project work has made to the field. We hope that this critical appraisal of ASTERIG will assist in the refinement of the risk-assessment process and the development of revised and updated tools that address the issues that we have raised in this paper. In our view, there are several guiding principles that will facilitate the development of more valid instruments. Valid risk assessment will be enhanced by (a) a clear empirical, or at least theoretical, basis for including each risk criterion; (b) the use of clear definitions; and (c) the use of calibration methods that are capable of detecting subtle variations in risk. For example, with reference to (a), we believe that ASTERIG’s current strategy of having one risk dimension that classifies staking as variable or fixed and another that categorizes stake as concurrent or sequential without any reference to stake size is one problematic area. In relation to (b), we believe that more information on the form of operational definitions for each risk criterion should be provided. The construction and meaning of each dimension of risk should be thoroughly explained, providing empirical support where possible or theoretical justification where it is not. Circumstances that fall outside the operational definition should also be discussed to reduce the risks of misinterpretation or misapplication when risk scoring. For example, ASTERIG might state clearly that “multi-game” refers to playing two games concurrently but does not encompass playing multiple games in sequence.

Within this endeavour, we recognize that one of the greatest challenges in developing an effective and efficient approach to measuring product risk is in maximizing its accuracy while minimizing its complexity. This involves trying to assess risk with as few dimensions as possible to widen and improve its application through simplicity.
and functionality. We also recognize that developing and maintaining an accurate risk scoring protocol will be an ongoing and iterative process. This is true not only because the products in question are routinely evolving, but also because some development needs will be borne out by its use in applied settings through a process of trial and error. In some cases, deficiencies in risk protocols might arise only when gambling companies and regulators attempt to apply them in practice. For these reasons, further research in the area of game design in relation to gambling harm will be a critical requirement for improving the accuracy with which risks posed by gambling products can be assessed. The most basic of these research needs is the development of a more precise understanding of how structural characteristics may lead to gambling harm. Even for those characteristics for which there is a strong case to suspect risk, such as game speed and continuity, considerable gaps in knowledge remain that hamper safer gambling applications with any reasonable degree of confidence. Thus, although we feel confident that fast continuous games increase risk of harm, the thresholds at which risk significantly increases are not well understood. More also needs to be known about how structural risks vary in relation to individual differences and needs (e.g., the impacts on normative gambling as opposed to higher risk gambling). Moreover, we recognize that there may be interactive effects whereby the distinct features of games and their platforms may interact to generate a different risk profile from what might apply if these features were considered in isolation (additively).

In conclusion, although we recognize the many challenges, we do not believe that the task of developing valid methods to assess product risk is insurmountable. What we encourage is further theoretical and conceptual analysis of the dimensions of product risk, greater validation of risk instruments against different product types and ranges, and, most important, further empirical research into the objective nature of gambling behaviour and whether theorized dimensions of risk account for apparent variations in the levels of harm associated with different products.

References


Hing, N., & Haw, J. (2010). The influence of venue characteristics on a player's decision to attend a gambling venue. Centre for Gambling Education & Research, Southern Cross University.


Peren, F.W. (2010). Assessment tool to measure and evaluate the risk potential of gambling products asTeriG. Research Institute for Gambling and Gaming.


to concentration of electronic gaming machines. *International Gambling Studies, 9*, 225–244. https://doi.org/10.1080/14459790903257981


Submitted October 2, 2020; accepted December 18, 2020. This opinion paper was peer reviewed by members of the JGI editorial board. All URLs were available at the time of submission.
For correspondence: Paul Delfabbro, Ph.D., Department of Psychology, University of Adelaide, Nth. Tce. Adelaide, South Australia, 5005, Australia. E-mail: paul.delfabbro@adelaide.edu.au

Competing interests: Both authors conduct consultancy research for both regulators, industry groups and government in relation the assessment of product risk.

Ethics approval: None required.

Acknowledgements/Funding Sources: Neither author received funding for this project.