DSM-5 criteria for gambling disorder: Underlying structure and applicability to specific groups of gamblers

PAWEL SLECZKA[1]*, BARBARA BRAUN[1], DANIELA PIONTEK[1], GERHARD BÜHRINGER[1,2]
and LUDWIG KRAUS[1,3]

[1]IFT Institut für Therapieforschung, München, Germany
[2]Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Dresden, Germany
[3]Centre for Social Research on Alcohol and Drugs, Stockholm University, Stockholm, Sweden

Background and aims: DSM-5 provides nine diagnostic criteria for gambling disorder. All criteria have a pre-assumed equal diagnostic impact and are applied to all individuals and groups in an equal manner. The aims of the study are to analyse the structure underlying the diagnosis and to assess whether DSM-5 is equally applicable to different groups of gamblers. *Methods:* Data from the 2009 German Epidemiological Survey of Substance Abuse and from a study on slot machine gamblers were used. Item Response Theory analysis was applied to estimate discrimination and severity parameters of the criteria. With the use of Differential Item Functioning analysis, potential criterion biases were analysed. We analysed data from 107 participants from the general population sample and 376 participants from the slot machine gamblers' sample who answered a 19-item diagnostic questionnaire based on the DSM criteria for gambling disorder. *Results:* A single underlying factor, the severity of gambling disorder, was identified in both samples. In the general population sample the criteria of preoccupation and chasing were least severe and showed low discriminatory power. Bailout, withdrawal and jeopardized matters criteria had highest severity and discriminatory power. The comparison of the two samples revealed two criterion biases in the preoccupation and tolerance criteria. *Conclusions:* The structure underlying the criteria is unidimensional but the disorder is manifested differently depending on disorder severity. The assumed equal impact of each criterion lacks support in the findings. The DSM-5 criteria measure a partially different construct in slot machine gamblers than in gamblers in the general population.

Keywords: gambling disorder, DSM-5, IRT, criterion bias

INTRODUCTION

Gambling disorder is classified as a non-substance-related disorder within "Substance-Related and Addictive Disorders" in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (American Psychiatric Association, 2013). The fourth edition of the DSM (DSM-IV) (American Psychiatric Association, 1994) has been widely used for the diagnosis of pathological gambling in practice and scientific research on different continents, including Europe (Rounsaville et al., 2002; Sassen, Kraus & Bühringer, 2011). Some changes have been introduced in the DSM-5: *illegal activities to support gambling* is no longer a diagnostic criterion of a gambling disorder (GD), the severity of GD can be specified based on the number of endorsed criteria (American Psychiatric Association, 2013), and a diagnosis is given if at least four out of nine criteria are met (5 out of 10 in the DSM-IV) (American Psychiatric Association, 1994; Petry, 2010; Reilly & Smith, 2013). This paper reports the results of an Item Response Theory (IRT) analysis of the structure underlying the nine criteria and addresses the question of applicability of these criteria to different groups of gamblers.

The DSM-5 is based on knowledge that has been gained in empirical studies and expresses the consensus of the scientific community on the nature of mental disorders (Kupfer, First & Regier, 2002; Rounsaville et al., 2002). It is basic to

the development of diagnostic instruments and is often referred to as a criterion of external validity for other measures in various fields of research and practice (Rounsaville et al., 2002). Although the new manual has been developed for North America, it is also widely used in studies elsewhere.

In general, a diagnosis is a form of reduction of information which might lead to oversimplified results (MacCallum, Zhang, Preacher & Rucker, 2002). While the information on the occurrence of the disorder remains important, a more detailed analysis of the underlying structure can offer insight into the syndrome. Symptoms (fulfilled criteria) can be understood as external manifestations of the unobservable latent construct 'disorder'). Although it is assumed that more fulfilled criteria indicate a more severe GD (American Psychiatric Association, 2013), there is only little knowledge on the symptoms which are associated with different levels of severity. IRT provides a set of analytical tools that allow conclusions about the relations between single criteria and the severity of the disorder.

Two studies that used IRT methods to analyse the structure of GD according to DSM-IV were both based on

* Corresponding author: Pawel Sleczka; IFT Institut für Therapieforschung, Parzivalstr. 25, D-80804 München, Germany; Phone: +49 (89) 36 08 04 35; Fax: +49 (89) 36 08 04 46; E-mail: sleczka@ift.de

the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) (Sacco, Torres, Cunningham-Williams, Woods & Unick, 2011; Strong & Kahler, 2007). The criteria *preoccupation with gambling* and *tolerance development* were found to be associated with low levels of GD severity, whereas *withdrawal symptoms* and *jeopardizing important matters* proved to be associated with a more severe level of GD. The criterion *illegal activities* was associated with the most severe stages of GD. As this criterion was reported by individuals who already fulfilled five other criteria, Strong and Kahler (2007) concluded that it contributes only little to the diagnosis, which consequently advocated its removal in DSM-5 (Petry, 2010). Furthermore, as the criteria constituted a unidimensional scale, the study supported the idea of an aggregated index for problem severity based on the number of criteria fulfilled. In Europe, based on data from the British Gambling Prevalence Survey (BGPSS) a bidimensional (Orford, Wardle, Griffith, Sproston & Erens, 2010) and a categorical (James, O'Malley & Tunney, 2014) construct underlying the DSM-IV criteria were reported. In sum, the literature on the structure of the DSM-IV criteria for gambling disorder is not consistent. In order to address this inconsistency and to acknowledge the recent revision of the DSM, our primary goal is (1) to analyse the structure underlying the DSM-5 criteria for GD.

The same set of criteria is assumed to measure GD irrespectively of the kind of gambling, age or gender of the individuals, etc. However, there is only little support for the assumption of equal functioning of the diagnostic criteria in different groups. The Differential Item Functioning (DIF) analyses of the NESARC data revealed gender-related biases concerning the criteria *gambling to escape* and *preoccupation* as well as an age bias regarding the *chasing* criterion (Strong & Kahler, 2007; Sacco et al., 2011). Independent of the severity of GD, women were more likely to gamble to escape problems than men, and younger gamblers were more likely than older to play in order to win back their losses (Sacco et al., 2011). Furthermore, a race bias has been reported in the criteria *preoccupation, cessation attempts* and *suffering losses* (Sacco et al., 2011). To our knowledge, there was no investigation of criteria functioning with respect to groups engaging in different kinds of gambling. As these groups differ in relation to various aspects of gambling, also the problems developed by these groups may differ as well. Based on this premise, the second aim of this study is to (2a) confirm the general structure underlying the criteria in *naturally* derived groups *(all individuals included in the typical for the research field recruitment procedure),* in which DSM-5 is used without any further adjustments and (2b) to investigate potential criterion biases.

METHODS

Study design and procedure

For the first aim of the study, data collected in the 2009 German Epidemiological Survey of Substance Abuse (ESA) were analysed. A two-stage probability design covering the general population was implemented. The response rate was

50% and the final sample consisted of $n = 8{,}006$ participants. A detailed description of the design and the sample has been published elsewhere (Kraus & Pabst, 2010; Sassen, Kraus et al., 2011). The ESA questionnaire contained questions on all major kinds of gambling available in Germany. For the second aim of the study gamblers from the 2009 ESA survey were compared with data collected within a study on slot machine gamblers. This study was also conducted in 2009 in gamblers visiting street casinos, gambling arcades and pubs (for an overview of gambling activities provided in Germany and underlying legislation see Ludwig, Kräplin, Braun & Kraus, 2013). Gamblers were approached by trained interviewers. The response rate was 38% and the final sample consisted of $n = 591$ participants. A detailed description of the study design can be found elsewhere (Bühringer, Kraus, Höhne, Küfner & Künzel, 2010). The sample represents frequent, long-term (over 2 years) slot machine gamblers.

Participants

In both samples investigated within the current study, only gamblers fulfilling one to eight out of nine DSM-5 criteria were included in the analyses. Those fulfilling none and those meeting all criteria were excluded for two reasons: (1) they do not add any variance in the IRT analysis; and (2) retention of subjects fulfilling no criteria leads to overestimation of the unidimensionality of the data (Steppan, Piontek & Kraus, 2014). Exclusion of subjects endorsing all criteria ($n = 13$) had no effect on the results. Cases with missing values on age or gender were deleted from the sample ($n = 3$). The final sample consisted of $n = 107$ gamblers in the general population survey (GGP) and $n = 376$ from the study on slot machine gamblers (SMG). All participants were between 18 and 64 years old. The IRT analysis investigated the sample-independent structure underlying the criteria and the DIF analysis aimed at analysing the applicability of the DSM-5 to naturally differing samples. The sample of gamblers from the general population included all kinds of gamblers while the SMG sample consisted of predominantly slot machine gamblers. Consistently with the aim to compare naturally derived samples, no further sample adjustment was intended. The difference in sample size does not pose a limitation to the analyses. A description of the samples is provided in Table 1.

Measurements

DSM-5 diagnostic criteria for GD were assessed using a 19-item questionnaire (intentionally measuring DSM-IV diagnosis, adjusted for DSM-5) (Stinchfield, 2003) translated into German in both studies. With the exception of the *withdrawal* criterion that is covered by only one item, each criterion is assessed by two items. In the current study items related to *illegal activities* were excluded as this is not a DSM-5 criterion. All questions referred to the last 12 months. Subjects answered all questions on a binary scale. In both samples, only participants who had invested at least 50 EUR (about 65 USD) in gambling in the course of the past 12 months answered the questionnaire. This threshold was used for economic reasons, and it was based on previous findings

Table 1. Sample description

	Gamblers in the general population (GGP) ($n = 107$)	Slot machine gamblers (SMG) ($n = 376$)	Difference test
Gender			
Female	15 (14.0%)	71 (18.9%)	
Male	92 (86.0%)	305 (81.1%)	$\chi^2(1) = 1.38; p > 0.05$
Age; Mean (SD)	33 (13.1)	39 (11.5)	$t(481) = -4.36; p < 0.001$
<30	60 (56.1%)	89 (23.7%)	
30–49	30 (28.0%)	216 (57.5%)	
>49	17 (15.9%)	71 (18.9%)	
Marital status			
Single	64 (59.8%)	182 (48.4%)	
Married	37 (34.6%)	114 (30.3%)	
Widowed	0 (0.0%)	7 (1.9%)	
Divorced	6 (5.6%)	73 (19.4%)	$\chi^2(3) = 14.31; p < 0.01$
Nationality			
German	85 (79.4%)	300 (80.7%)	
Other	22 (20.6%)	72 (19.6%)	$\chi^2(1) = 0.078; p > 0.5$
Gambling on slot machines in the last 12 months	36 (33.6%)	373 (100.0%)*	$\chi^2(1) = 285.487; p < 0.001$
Gambling on slot machines at least once a week	16 (15.0%)	306 (81.4%)*	$\chi^2(1) = 169.458; p < 0.001$
Slot machines as favourite kind of gambling	15 (14.0%)	326 (86.7%)	$\chi^2(1) = 212.0; p < 0.001$

*$n = 3$ answers missing

indicating that individuals betting lower amounts of money are at no or at only a limited risk of fulfilling any DSM-5 criteria for GD (Currie et al., 2006; National Research Council, 1999). Cronbach's α was 0.83 in the GGP sample and 0.82 in the SMG sample.

Statistical analyses

Item Response Theory analysis. With regard to the first aim of the study, a two parameter logistic IRT model (2PL-IRT model) was run using Mplus (Version 5: Muthén & Muthén, 2007) software. The 2PL-IRT model was chosen over the Rasch model because of its ability to estimate both the discrimination and the threshold parameters for each criterion. The threshold parameter (b parameter) indicates criterion severity – criteria with higher thresholds are met only on a higher level of severity. The discrimination parameter (a parameter) is related to the precision with which an item distinguishes between respondents below and above the threshold. Based on the a and b parameters of all criteria the total information function can be estimated, giving information on how well all criteria discriminate between different levels of severity of GD. The adequacy of the IRT model fit was verified using the Tucker Lewis index (TLI), the comparative fit index (CFI) and the root mean square error of approximation (RMSEA) in both data sets. Hu and Bentler (1999) suggested a maximum value of 0.06 for RMSEA and a minimum value of 0.95 for either TLI or CFI as cut-offs indicating a good model fit.

Differential item functioning. To identify potential criterion biases, Differential Item Functioning (DIF) analysis was conducted to compare gamblers from the general population with slot machine gamblers. Comparisons between groups differing in size or severity of gambling problems are legitimate because of the specificity of IRT models, in which the underlying structure is estimated independently of these factors. It is assumed that a criterion functions differently when persons with the same GD severity differ in their responses because of other characteristics. The criterion bias is detected when either a or b, or both parameters differ significantly between the two groups. The DIF analysis was conducted using the IRTLRDIF (Version 2.0b; Thissen, 2001) software. The procedure is described in detail elsewhere (Thissen, 2001). The analysis comprises three steps. First, those criteria that function equally in both groups were identified and used as a set of anchor-items for testing the other items. Based on these items, the underlying F-parameter (in this case GD severity) was calculated for both groups. Second, the remaining items were tested for significant DIF, and those that tested negative were included in the set of anchor-items in the third step. The remaining items were again tested for significant DIF with the F-parameter computed based on the anchor-items.

Assumptions of the 2PL-IRT model (and the DIF analysis). In order to verify the applicability of the 2PL-IRT model for the data at hand, the following preconditions were analysed using the two data sets individually and combined:

Unidimensionality. While some IRT models deal with multidimensionality, the 2PL-IRT model posits that all items (criteria) measure a single dimension, in this case the severity of GD. Exploratory factor analyses (EFA) using tetrachoric correlations were run using *Mplus* to assess the dimensionality of GD criteria in both samples (Muthén & Muthén, 2007). The factor extraction was based on the scree plot assessment (Costello & Osborne, 2005), the ratio of the first-to-second eigenvalue (Solcum-Gori & Zumbo, 2010) and, finally, the theoretical accuracy of the factors. In both samples, the scree plot evaluation and the ratio of first-to-second eigenvalue exceeding 3 supported the one-factor model (Solcum-Gori & Zumbo, 2010). In both samples the first factor's eigenvalue was significantly larger than the

second (5.01 vs. 1.26, and 4.04 vs. 1.08 for GGP and SMG, respectively). In the two-factor models, the first factor loaded on all but the 6th criterion in the GGP sample and on all but the 7th criterion in the SMG sample. Two-factor models were therefore considered theoretically unsupported. The factor loadings for the one-factor models are presented in Table 2. Some of the criteria had low factor loadings, one falling below 0.4 in the SMG sample.

The *Monotonicity* assumption is met when the probability of endorsing each criterion rises analogously to the probability of endorsing others. The assessment of monotonicity was done by graphical evaluation of multiple graphs illustrating the total sum of endorsed criteria on the X-axis and the ratio of people endorsing the criterion on the Y-axis. The assumption was confirmed for the analysed data sets.

The residual correlation matrix and model indices were analysed to assess the *local independence*. As there was no association between the criteria after controlling for the underlying latent variable, local criteria independence was assumed.

Sensitivity analyses. The aim of the study was to compare *naturally* derived samples. However, additional sensitivity analyses were conducted to investigate the robustness of the results. As the two compared groups differed in age, DIF analyses were conducted on age-matched samples. Furthermore, DIF analysis was repeated on GGP without $n = 15$ individuals favouring slot machines vs. SMG without $n = 50$ individuals preferring other games over slot machines.

Ethics

Both the ESA and the study on slot machine gamblers were approved by the Ethical Board of the German Society of Psychology (DGPs). All participants were informed about the study and provided informed consent.

RESULTS

Table 2 summarizes the 12-month prevalence of DSM-5 criteria for GD in GGP and SMG samples. The most frequently endorsed criteria in GGP were *preoccupation* (61.7%) and *chasing* (55.1%). On average, subjects met 2.7 diagnostic criteria for GD and 28.0% were classified as disordered gamblers (fulfilling at least 4 out of 9 DSM-5 criteria). Subjects of the SMG sample met on average 4.4 criteria and 59.6% were classified as disordered gamblers.

The underlying structure

Based on the TLI, CFI and RMSEA, the 2PL-IRT models were considered fitting both samples. The results of the model for GGP are presented in Table 2 and will be described in detail. For better understanding of the data we included the results for SMG in the table. The Item Characteristic Curves for GGP are graphically displayed in Figure 1. In general, the criteria of *preoccupation* and *chasing* were associated with a lower level of severity of GD and both criteria showed low discriminatory power (*a* parameter). In contrast, *jeopardizing important matters,withdrawal,* and *bail-out* criteria were associated with a more severe level

of GD. As indicated by the *a* parameter and visualized by the steepness of slope in Figure 1, all of these criteria, especially the *bail-out* criterion, showed good discriminatory power.

The graphical comparison of the total information curve (TIC) of the DSM-5 criteria for GD in the two samples is presented in Figure 2. In relation to the assessment of the discriminatory power of the DSM-5, the results show that the relative performance of all criteria was higher at more severe levels of GD. The aggregated results indicate that the set of criteria delivers substantially more information in the GGP sample than in the SMG sample.

Differential item functioning

In relation to the second aim of the study, the DIF analysis showed significant differences in both *a* and *b* parameters of the *preoccupation* and *tolerance* criteria indicating criterion biases. Exact test statistics are included in Appendix A1–A3. Criterion *preoccupation* exhibits a "*b*-DIF" ($G^2(1) = 18.2$), with a severity estimate of –0.63 for GGP (reference group) and 0.54 for SMG (focal group). Criterion *tolerance* exhibits a "*b*-DIF" ($G^2(1) = 6.7$), with a severity estimate of 0.42 for GGP (reference group) and 1.19 for SMG (focal group). Overall for both criteria the *b* parameter was higher in SMG than in GGP.

Sensitivity analyses

The sensitivity analyses confirmed the reported criterion biases related to the criteria *preoccupation* and *tolerance*. DIF was identified in analyses on both the age-matched samples and on GGP excluding individuals favouring slot machines vs. SMG excluding individuals preferring other games over slot machines. The results of the sensitivity analyses are available on request.

DISCUSSION

We aimed at (1) investigating the structure underlying the nine criteria for gambling disorder in German gamblers and (2) assessing the applicability of DSM-5 to different groups of gamblers. The results indicate a unidimensional structure with different symptoms manifested along a severity continuum. Furthermore, group-related criterion biases were found.

With regard to the first aim, the EFA indicate a one-factor model as the best fitting solution for both datasets. This is in line with previous results found in the USA (Stinchfield, 2003; Strong & Kahler, 2007; Toce-Gerstein, Gerstein & Volberg, 2003). However, psychometric analyses also reported a two factor solution (Orford et al., 2010), and Latent Class Analysis (LCA; Carragher & Williams, 2011) as well as taxometric analyses (James et al., 2014) suggested a categorical structure of GD. Differences in the number of identified factors might be due to different statistical approaches (Muthen, 2006) or cultural differences.

The present results support the hypothesis that the latent structure underlying the DSM-5 criteria is best described as the unidimensional continuum of *gambling disorder severity*.

Table 2. Factor loadings and item response theory parameters of the DSM-5 criteria in two samples

	Gamblers in the general population (GGP)				Slot machine gamblers (SMG)			
	Prevalence (% of the total sample)	Factor loading (one-factor model)	Discrimination parameter a Estimate (SE)	Severity parameter b Estimate (SE)	Prevalence (% of the total sample)	Factor loading (one-factor model)	Discrimination parameter a Estimate (SE)	Severity parameter b Estimate (SE)
1. Preoccupation	66 (61.7%)	0.47	0.53 (0.14)	−0.64 (0.31)	247 (65.7%)	0.68	0.92 (0.15)	−0.60 (0.12)
2. Tolerance	38 (35.5%)	0.74	1.08 (0.26)	0.51 (0.17)	171 (45.5%)	0.51	0.59 (0.11)	0.22 (0.13)
3. Cessation attempts	39 (36.5%)	0.72	1.04 (0.26)	0.48 (0.18)	238 (63.3%)	0.68	0.93 (0.16)	−0.50 (0.11)
4. Withdrawal	12 (11.2%)	0.90	2.01 (1.19)	1.35 (0.25)	115 (30.6%)	0.78	1.29 (0.23)	0.64 (0.10)
5. Gambling to escape	24 (22.4%)	0.60	0.74 (0.23)	1.28 (0.34)	176 (46.8%)	0.38	0.41 (0.09)	0.21 (0.18)
6. Chasing losses	59 (55.1%)	0.46	0.51 (0.15)	−0.28 (0.28)	265 (70.5%)	0.60	0.76 (0.13)	−0.89 (0.16)
7. Concealment of own gambling	30 (28.0%)	0.73	1.05 (0.35)	0.80 (0.22)	216 (57.5%)	0.57	0.70 (0.12)	−0.33 (0.12)
8. Jeopardized or lost significant matters	14 (13.1%)	0.85	1.62 (0.69)	1.32 (0.26)	164 (43.6%)	0.72	1.03 (0.17)	0.22 (0.08)
9. Relies on others to be "bailed out"	7 (6.5%)	0.93	2.56 (2.26)	1.62 (0.30)	75 (20.0%)	0.62	0.78 (0.13)	1.36 (0.20)
Pathological Gamblers (at least 4/9 DSM-5 criteria)	30 (28.0%)				224 (59.6%)			
Total endorsed criteria (Mean)	$M = 2.7$	$(SD = 2.1)$			$M = 4.4$	$(SD = 1.2)$		
Fit statistics								
RMSEA	0.055				0.058			
CFI	0.970				0.956			
TLI	0.960				0.941			

CFI: comparative fit index; RMSEA: root mean square error of approximation; SE: standard error; TLI: Tucker Lewis index.
Two samples were analysed separately: the a and b parameters are comparable within columns but not rows.
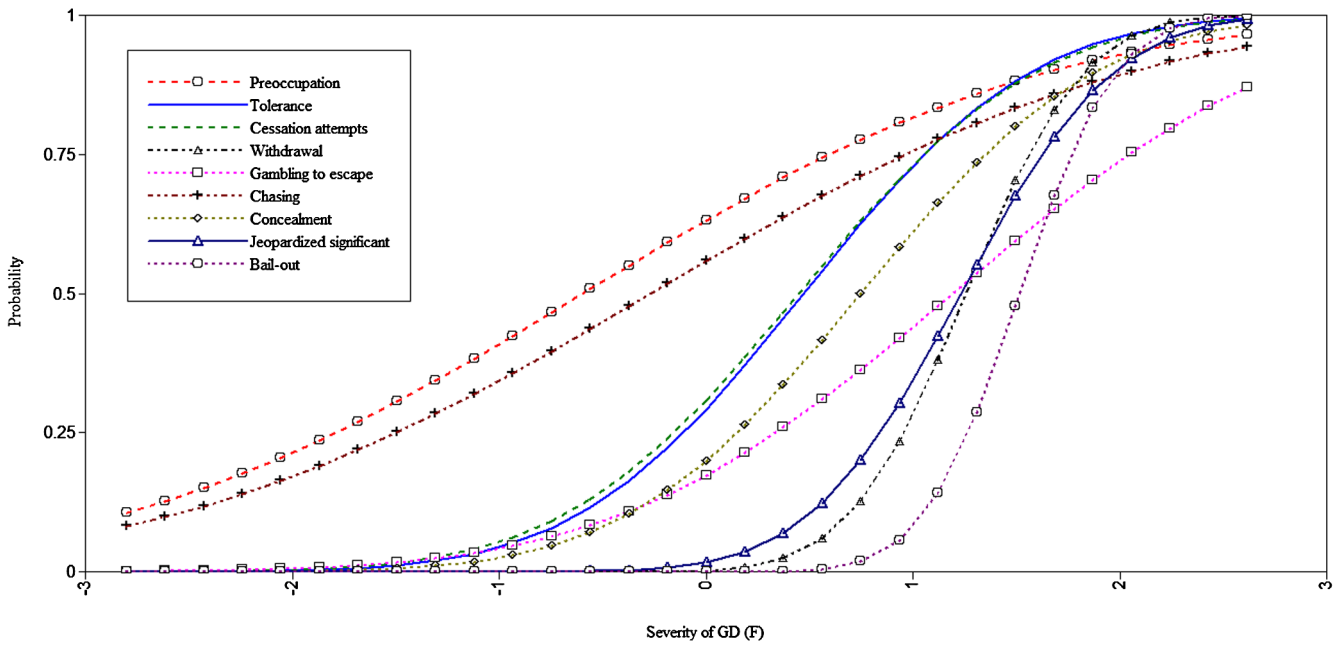
Figure 1. Criterion response curves for DSM-5 criteria for gamblers in general population (GGP)

The IRT analysis showed that the *preoccupation* and *chasing* symptoms are related to less severe levels of GD. In contrast, gamblers with a more severe gambling disorder, besides meeting these two criteria, more frequently reported to jeopardize important matters and to conceal the extent of their gambling. These differences in relations between single criteria and the level of severity of GD are in line with results of previous IRT analyses (Strong & Kahler, 2007; Sacco et al., 2011), and descriptive comparisons (Toce-Gerstein et al., 2003). They can be described as criteria endorsement pattern and suggest that while the underlying structure appears to be unidimensional, GD may be manifested differently depending on its severity.

While our results support a unidimensional structure, they do not support the additive severity rating in DSM-5. The nine criteria vary in discriminatory power and therefore have an unequal share in the IRT-based assessment of GD severity. Each criterion has a different influence at a different level of GD severity. As illustrated in Figure 1, in line with previous findings (Strong & Kahler, 2007), the DSM-5 set of criteria discriminates best at the higher levels of GD severity and performs poorer at lower levels. We conclude that introducing new GD criteria that discriminate well (high *a* parameter) in less severe stages (low *b* parameter) might improve the performance of further editions of the DSM and possibly allow reliable measurement of *diagnostic orphans,*
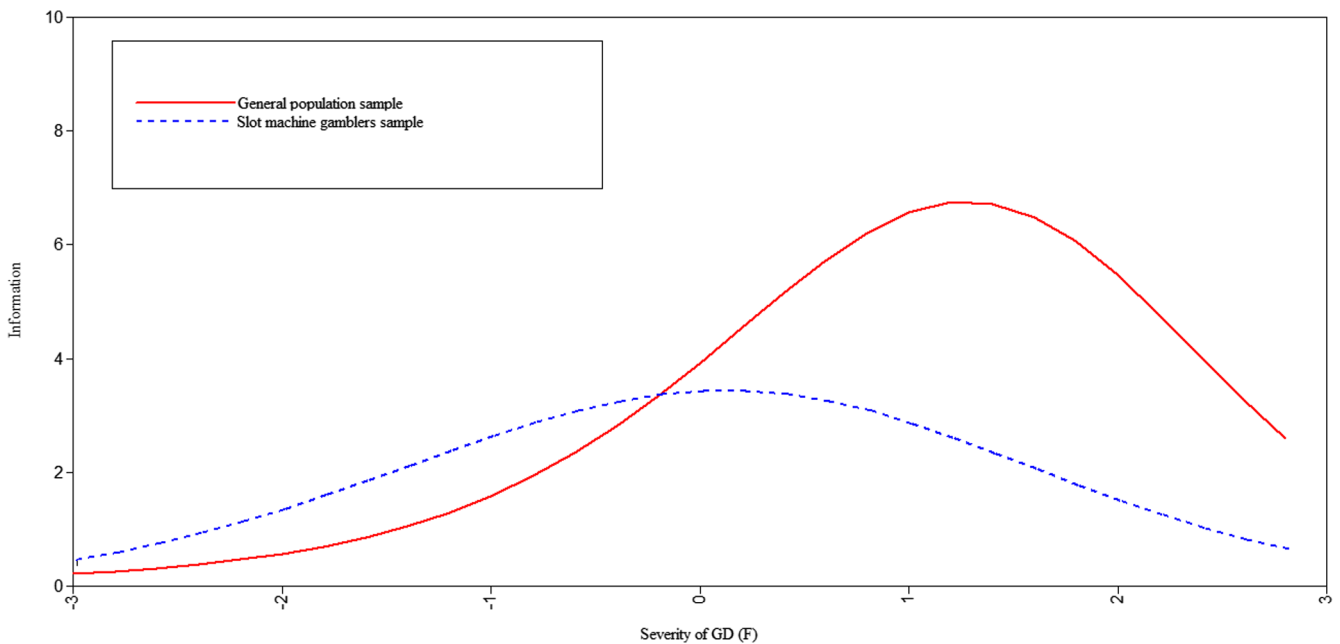

Figure 2. Aggregated test information of the DSM-5 criteria for gambling disorder in gamblers in general population and slot machine gamblers

i.e. individuals fulfilling less than 4 criteria. Altogether a valid diagnostic instrument should address not only the question *how many,* but also *which* criteria are met. For example, assigning higher weights to more severe criteria would produce more accurate results.

With regard to our second aim, the results of the DIF analyses indicate that, compared to gamblers in the general population, slot machine gamblers reported *preoccupation* and *tolerance* symptoms less frequently regardless of GD severity. This leads to the conclusion that, independently of the severity of the disorder, slot machine gamblers display partially different symptoms of GD than gamblers in the general population. Possible explanations of these differences might be that: (1) SMG have certain characteristics (other than age and severity of disorder) that are related to the choice of game and the development of specific symptoms; (2) slot machine gambling facilitates the occurrence of certain symptoms of GD, or (3) SMG respond differently to DSM-5 criteria due to either their personal characteristics or the context of slot machine gambling. Independently of cause, criterion biases are indicators of limited construct validity of DSM-5 criteria for GD in this group. As can be seen in Figure 2, the set of criteria generally differentiates better in GGP, which is the more heterogeneous sample. Thus, it seems that DSM-5 provides more reliable data when applied to gamblers in general rather than to a specific group of gamblers.

The conducted analyses have some limitations: (1) due to low prevalence of GD, the sample size is rather small. It has been suggested, however, that the better the data fulfil the assumptions, the smaller samples are needed for applying 2PL-IRT models (Reeve & Fayers, 2005). Interestingly, the aggregated test information was higher in the smaller sample; (2) methodological differences pose some limitations for DIF analysis. The SMG study involved contact to interviewers and it cannot be excluded that criterion biases are to some extent caused by the presence of an interviewer. Furthermore, in the SMG study, long term (two years) gambling involvement was an inclusion criterion and GGP included all gamblers who gambled in the last 12 months. Thus, it cannot be ruled out that the differences are caused by different stages (not severity) in the development of GD. These issues need further investigation; (3) one might argue that the inclusion of slot machine gamblers in GGP and vice versa, gamblers preferring other games in SMG might confound the results. The goal of the study was to examine the applicability of the DSM-5 as it is currently used; thus, the choice of the samples was deliberate. Nevertheless, additional sensitivity analysis on samples excluding these cases confirmed the two identified criterion biases.

CONCLUSIONS

The present study challenges the current application of the DSM-5 criteria for GD. The assumed equal impact of each criterion on the diagnosis lacks support in the results of the IRT analysis. The presence of DIF implies that the criteria *preoccupation* and *tolerance* function unequally in different groups of gamblers. Inequalities in the performance of the criteria question the validity of the DSM-5 in specific subgroups. Based on the presented findings, we suggest a discussion on the use of severity rating of GD and further research on the applicability of DSM-5 diagnosis to different groups of gamblers and on the structure underlying the criteria.

REFERENCES

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders (Fourth edition)*. Washington, DC: American Psychiatric Press.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders: Fifth edition*. Washington, DC: American Psychiatric Association.

Bühringer, G., Kraus, L., Höhne, B., Küfner, H. & Künzel, J. (2010). *Untersuchung zur Evaluierung der Fünften Novelle der Spielverordnung vom 17.12.2005*. Berlin: Bundesministerium für Wirtschaft und Technologie.

Carragher, N. & McWilliams, L. A. (2011). A latent class analysis of DSM-IV criteria for pathological gambling: Results from the National Epidemiologic Survey of Alcohol and Related Conditions. *Psychiatry Research,187*, 185–192.

Costello, A. B. & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation: A peer-reviewed electronical journal,10*, 1–9.

Currie, S. R., Hodgins, D. C., Wang, J., el-Guebaly, N., Wynne, H. & Chen, S. (2006). Risk of harm among gamblers in the general population as a function of level of participation in gambling activities. *Addiction,101*(4), 570–580.

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling,6*(1), 1–55.

James, R., O'Malley, C. & Tunney, R. (2014). On the latent structure of problem gambling: A taxometric analysis. *Addicition,109*, 1–10.

Kraus, L. & Pabst, A. (2010). Studiendesign und Methodik des Epidemiologischen Suchtsurveys 2009. *Sucht,56*(5), 315–326.

Kupfer, D. J., First, M. B. & Regier, D. E. (2002). Introduction. In D. J. Kupfer, M. B. First & D. E. Regier (Eds.), *A Research agenda for Dsm V*. Virginia: American Psychiatric Association.

Ludwig, M., Kräplin, A., Braun, B. & Kraus, L. (2013). Gambling experiences, problems, research and policy: Gambling in Germany. *Addiction,108*(9), 1554–1561.

MacCallum, R. C., Zhang, S., Preacher, K. J. & Rucker, D. D. (2002). On the practice of dichotomization of quantitatives variables. *Psychological Methods,7*, 19–40.

Muthen, B. (2006). Should substance use disorders be considered as categorical or dimensional?*Addiction,101*(Suppl 1), 6–16.

Muthén, L. K. & Muthén, B. O. (2007). *Mplus: User`s guide*. Fifth edition. Los Angeles: CA: Muthén & Muthén.

National Research Council (1999). *Pathological gambling: A critical review*. Washington, DC: National Academy Press.

Orford, J., Wardle, H., Griffith, M., Sproston, K. & Erens, B. (2010). PGSI and DSM-IV in the 2007 British Gambling Prevalence Survey. *International Gambling Studies,10*(1), 31–44.

Petry, N. M. (2010). Pathological gambling and the DSM-V. *International Gambling Studies,10*(2), 113–115.

Reeve, B. B. & Fayers, P. (2005). Applying item response theory modelling evaluating questionnaire item and scale properties. In P. Fayers & R. Hays (Eds.), *Assessing quality of life in clinical trials: Methods and practice* (pp. 55–74). New York, u.a.: Oxfrod University Press.

Reilly, C. & Smith, N. (2013). *White paper on the evolving definition of pathological gambling in the DSM-5*. Washington, Beverly: National Center for responsible gambling.

Rounsaville, B. J., Alarcón, R. D., Andrews, G., Jackson, J. S., Kendell, R. D. & Kendler, K. (2002). Basic Nomenclature Issues of DSM-IV. In D. J. Kupfer, M. B. First & D. A. Regier (Eds.), *A research agenda for Dsm V* (pp. 1–29). Virginia: American Psychiatric Association.

Sacco, P., Torres, L. R., Cunningham-Williams, R. M., Woods, C. & Unick, G. J. (2011). Differential item functioning of pathological gambling criteria: An examination of gender, race/ethnicity, and age. *Journal of Gambling Studies,27*(2), 317–330.

Sassen, M., Kraus, L. & Bühringer, G. (2011). Differences in pathological gambling prevalence estimates: Facts or artefacts? *International Journal of Methods in Psychiatric Research,20*(4), 83–99.

Sassen, M., Kraus, L., Bühringer, G., Pabst, A., Piontek, D. & Taqi, Z. (2011). Gambling among adults in Germany: Prevalence, disorder and risk factors. *Sucht,57*(4), 249–257.

Solcum-Gori, S. L. & Zumbo, B. D. (2010). Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Social Indicators Research,102*(3), 443–461.

Steppan, M., Piontek, D. & Kraus, L. (2014). The effect of sample selection on the distinction between alcohol abuse and dependence. *International Journal of Alcohol and Drug Research,3*(2), 159–168.

Stinchfield, R. (2003). Reliability, validity, and classification accuracy of a measure of DSM-IV diagnostic criteria for pathological gambling. *American Journal of Psychiatry,160*(1), 180–182.

Strong, D. R. & Kahler, C. W. (2007). Evaluation of the continuum of gambling problems using the DSM-IV. *Addiction,102*(5), 713–721.

Thissen, D. (2001). *IRTLRDIFv.2ob: Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning (computer program)*.

Toce-Gerstein, M., Gerstein, D. R. & Volberg, R. A. (2003). A hierarchy of gambling disorders in the community. *Addiction,98*(12), 1661–1672.

APPENDIX

Appendix A1. Test statistics of the Differential Item Functioning analysis: Step 1

Criterion	Hypothesis Test		$G2^1$	d.f.	Gamblers in the general population		Slot machine gamblers		μ	δ
					a	b	a	b		
1 Preoccupation	All	equal	19.9	2	0.74	−0.75	1.73	0.49	1.02	0.87
	a	equal	4	1	1.43	−0.52	1.43	0.4	1.01	0.92
	b	equal	15.9	1	1.01	0	1.01	0	0.96	0.97
2 Tolerance	All	equal	6.7	2	1.72	0.45	0.93	1.2	0.97	0.97
	a	equal	2.8	1	1.06	0.64	1.06	1.18	0.98	0.96
	b	equal	3.9	1	0.93	1.1	0.93	1.1	0.96	0.97
3 Cessation attempts	All	equal	0.1	2	1.43	0.5	1.43	0.44	0.96	0.97
4 Withdrawal	All	equal	1.4	2	3.22	1.43	1.95	1.65	0.96	0.98
5 Gambling to escape	All	equal	4.7	2	0.97	1.52	0.65	1.17	0.95	0.99
	a	equal	0.9	1	0.71	1.93	0.71	1.15	0.95	0.98
	b	equal	3.9	1	0.81	1.24	0.81	1.24	0.96	0.97
6 Chasing	All	equal	3.6	2	0.66	−0.37	1.3	0.11	0.98	0.93
7 Concealment	All	equal	3.4	2	1.7	0.81	1.12	0.61	0.94	1.01
8 Jeopardized matters	All	equal	5.3	2	2.42	1.46	1.69	1.17	0.94	0.99
	a	equal	1	1	1.8	1.63	1.8	1.17	0.94	0.97
	b	equal	4.3	1	1.92	1.23	1.92	1.23	0.96	0.97
9 Bail out	All	equal	4.1	2	3.73	1.77	1.33	2.33	0.96	0.99
	a	equal	3.9	1	1.53	2.36	1.53	2.22	0.96	0.97
	b	equal	0.2	1	1.55	2.22	1.55	2.22	0.96	0.97

[1]Values exceeding 3.84 (the α = 0.05 critical value of the $χ^2$ distribution, when *df* = 1) are considered as significant difference (Thissen, 2001).

Appendix A2. Test statistics of the Differential Item Functioning analysis: Step 2

Criterion	Hypothesis Test		$G2^1$	d.f.	Gamblers in the general population		Slot machine gamblers sample		μ	δ
					a	b	a	b		
1 Preoccupation	All	equal	10.7	2	0.82	−0.68	1.44	0.31	0.92	0.93
	a	equal	1.6	1	1.25	−0.54	1.25	0.23	0.9	0.99
	b	equal	9.2	1	0.96	−0.12	0.96	−0.12	0.84	1.04
2 Tolerance	All	equal	5.3	2	1.83	0.46	1.17	1.04	0.86	0.82
	a	equal	1.3	1	1.32	0.56	1.32	1.02	0.86	0.79
	b	equal	3.9	1	1.14	0.94	1.14	0.94	0.83	0.82
5 Gambling to escape	All	equal	3.4	2	0.83	1.71	0.79	1.1	0.92	0.79
8 Jeopardized matters	All	equal	3.2	2	2.2	1.48	1.84	1.19	0.99	0.88
9 Bail out	All	equal	2.7	2	2.42	1.9	1.26	2.25	0.91	0.84

[1]Values exceeding 3.84 (the α = 0.05 critical value of the $χ^2$ distribution, when *df* = 1) are considered as significant difference (Thissen, 2001).

Appendix A3. Test statistics of the Differential Item Functioning analysis: Step 3

Criterion	Hypothesis Test		$G2^1$	d.f.	Gamblers in the general population		Slot machine gamblers sample		μ	δ
					a	b	a	b		
1 Preoccupation	All	equal	20.1	2	0.94	−0.63	1.7	0.54	1.09	0.88
	a	equal	2	1	1.49	−0.5	1.49	0.49	1.09	0.93
	b	equal	18.2	1	1	0.03	1	0.03	1.02	0.99
2 Tolerance	All	equal	9.2	2	2.03	0.42	1.13	1.19	0.99	0.8
	a	equal	2.5	1	1.31	0.56	1.31	1.18	1.01	0.78
	b	equal	6.7	1	1.07	1.08	1.07	1.08	0.98	0.8

[1]Values exceeding 3.84 (the α = 0.05 critical value of the $χ^2$ distribution, when *df* = 1) are considered as significant difference (Thissen, 2001).

Appendix B1. Items (in German) measuring the criteria for GD

Es folgen nun Fragen zum Glücksspielen. Dieser Teil dauert ca. 5 Minuten. Bitte lesen Sie sich die folgenden Fragen sorgfältig durch und lassen Sie sich ausreichend Zeit zur Bearbeitung.

Geben Sie für alle Aussagen an, ob diese in den letzten 12 Monaten auf Sie zutreffen.

Unter dem Sammelbegriff Glücksspielen sind alle Wetten und Spiele mit Geldeinsatz zu verstehen.

Gab es Phasen, in denen Sie sehr viel Zeit damit verbrachten, über Ihr vergangenes oder zukünftiges Glücksspielen nachzudenken/zu grübeln?

Haben Sie oft über Möglichkeiten nachgedacht, wie Sie den Geldeinsatz für Glücksspiele beschaffen könnten (z.B. durch Kredit, Leihgabe durch Freunde oder Verwandte, Diebstahl)?

Gab es Zeiten, in denen Sie häufiger als vorher spielen mussten, um denselben Reiz beim Glücksspiel zu erleben (z.B. durch längeres und/ oder häufigeres Spielen)?

Gab es Phasen, in denen Sie mit größeren Geldbeträgen oder höheren Einsätzen als vorher spielen mussten, um denselben Reiz beim Glücksspiel zu verspüren (z.B. indem Sie immer mehr Geld für das Glücksspiel ausgegeben haben)?

Haben Sie mehrmals versucht, Ihr Glücksspielen zu reduzieren oder zu kontrollieren und das als schwierig empfunden?

Haben Sie mehrmals ohne Erfolg versucht, mit dem Glücksspielen aufzuhören?

Haben Sie sich unruhig oder reizbar gefühlt, nachdem Sie versucht hatten, Ihr Glücksspielen zu reduzieren oder ganz damit aufzuhören?

Hatten Sie das Gefühl, dass Sie gespielt haben, um vor persönlichen Problemen zu fliehen?

Hatten Sie den Eindruck, dass Ihnen Ihr Glücksspielen geholfen hat, unangenehme Gefühle wie Angst oder Depression zu mildern?

Ist es öfter vorgekommen, dass Sie Geld verloren haben und innerhalb weniger Tage erneut gespielt haben, um das verlorene Geld wieder zu gewinnen?

Ist es vorgekommen, dass Sie große Spielschulden hatten und Sie dann immer häufiger gespielt haben in der Hoffnung, Ihre Verluste wieder zurück zu gewinnen?

Haben Sie Familienmitglieder, Freunde, Mitarbeiter oder Lehrer oft angelogen, wenn es um das Ausmaß Ihres Glücksspielens oder um die Höhe Ihrer Spielschulden ging?

Haben Sie Ihr Glücksspielen gegenüber anderen (z.B. Familienmitgliedern) oft verheimlicht oder versucht, es zu verheimlichen?

Es kommt vor, dass Personen einen Scheck fälschen oder etwas stehlen, um ihr Glücksspiel zu finanzieren. Ist dies bei Ihnen in den letzten 12 Monaten vorgekommen?

Es kommt vor, dass Personen etwas Illegales tun, etwa Veruntreuung oder Betrug, um Geld für das Glücksspielen zu haben. Ist dies bei Ihnen in den letzten 12 Monaten vorgekommen?

Gab es Phasen, in denen Ihr Glücksspielen zu Problemen in der Beziehung zu Ihrer Familie, Ihren Freunden, Mitarbeitern oder Lehrern geführt hat?

Haben Sie wegen Ihres Glücksspielens Arbeits- oder Schultage, soziale Aktivitäten oder Familienaktivitäten versäumt?

Haben Sie andere Personen wegen Ihrer finanziellen Probleme durch das Glücksspielen gebeten, Ihnen Geld zu leihen?

Haben Sie andere Ihre Glücksspielschulden bezahlen lassen (d.h. sich aus der Klemme helfen lassen), wenn Sie wegen Ihrer finanziellen Lage verzweifelt waren?