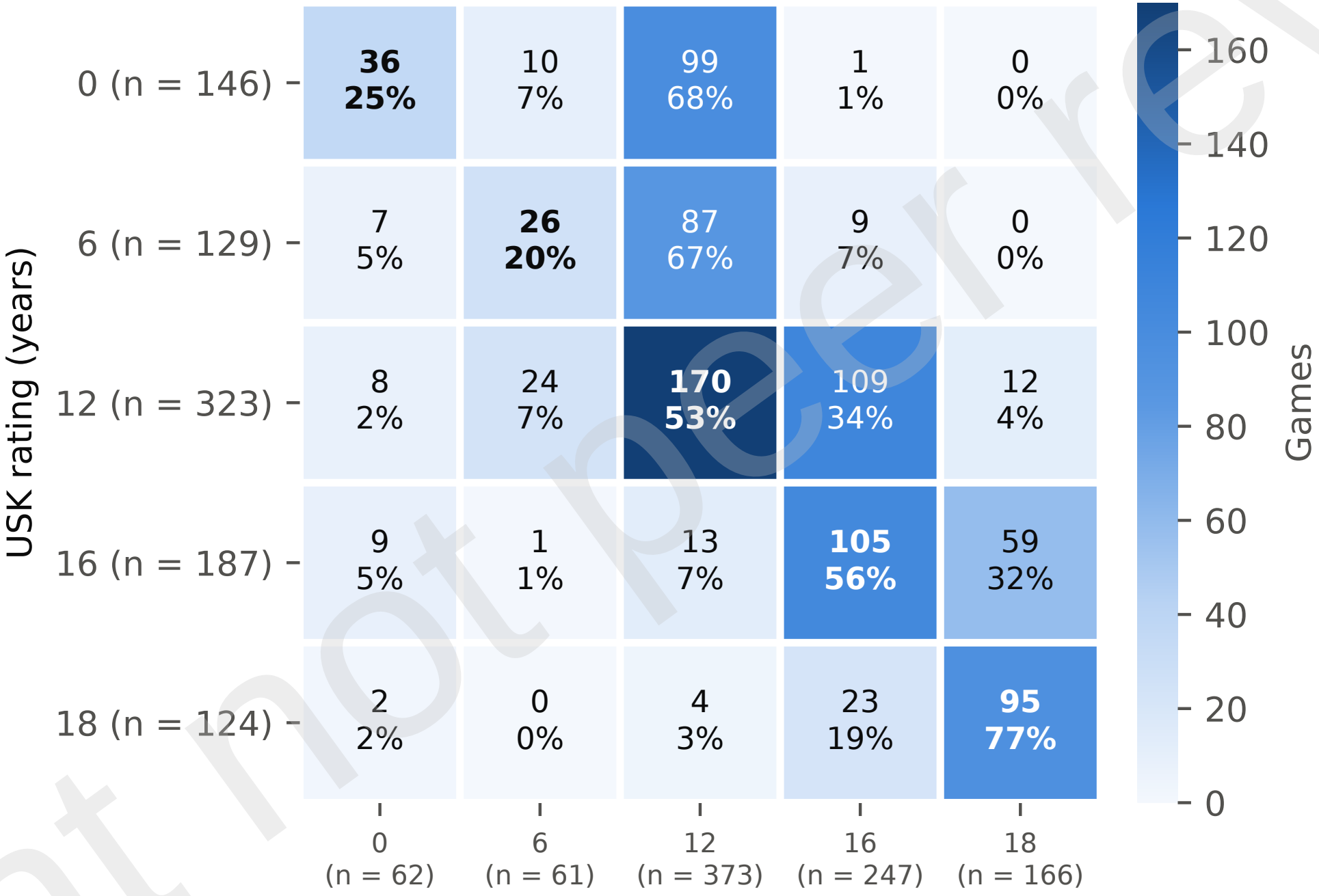
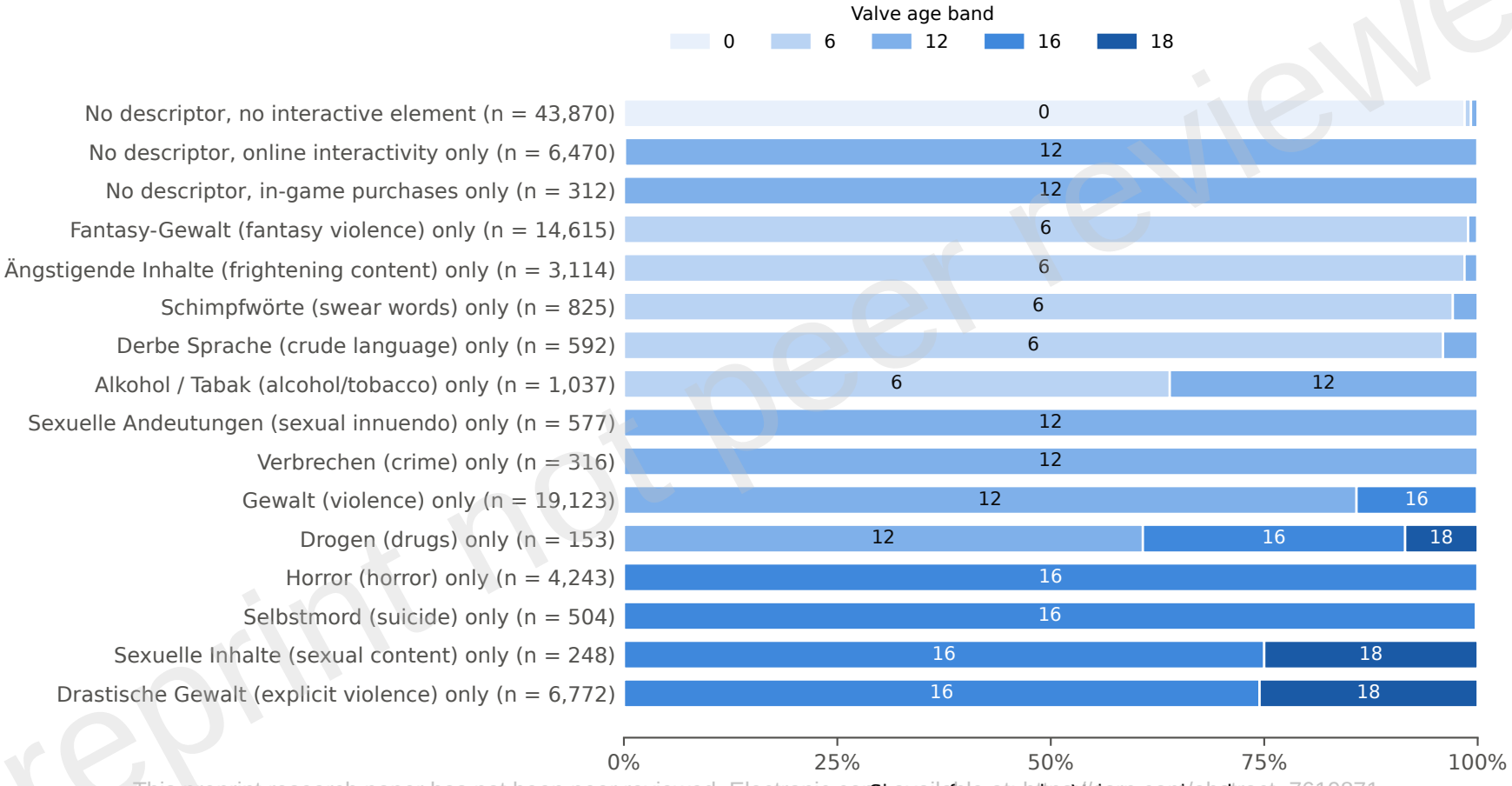
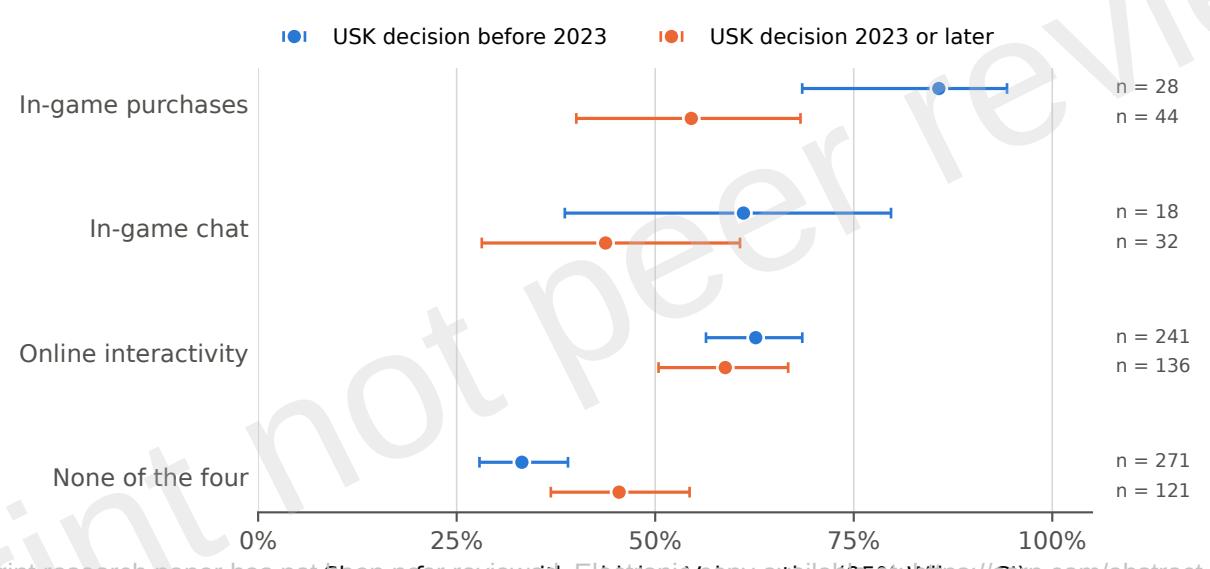


This paper has not been peer reviewed. Electronic copy available at: https://ssrn.com



This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871



Preprint not peer reviewed
Preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=

Same game, different age: Platform ratings versus board ratings of video games on Steam in Germany

Aydın Erden[a]

[a] Department of Management Information Systems, Faculty of Business Administration, Marmara University, Recep Tayyip Erdoğan Külliyesi, 34854 Maltepe, İstanbul, Türkiye

Corresponding author: Aydın Erden, Marmara University, Faculty of Business Administration, Department of Management Information Systems, Recep Tayyip Erdoğan Külliyesi, Room I2-315, 34854 Maltepe, İstanbul, Türkiye. E-mail: aydin.erden@marmara.edu.tr. ORCID: 0000-0002-5124-8335.

Acknowledgements

None.

Declaration of competing interests

The author declares no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Aydın Erden: Conceptualization, Methodology, Software, Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review and editing.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

Same game, different age: Platform ratings versus board ratings of video games on Steam in Germany

Abstract

Since November 2024, Steam has shown German customers only games with a valid age rating, mostly rated by Valve from developers' content surveys rather than by the German self-regulatory body, the Unterhaltungssoftware Selbstkontrolle (USK). Germany thus allows a same-game comparison of platform and board ratings under the same law and age bands. We linked German rating data for the full Steam catalogue (180,201 games) to the USK's title database. For 909 games with a traceable board decision, the two ratings matched for 47.5%, whereas the USK and Pan European Game Information ratings of these games agreed far more closely (quadratically weighted kappa 0.88, against 0.61 for Valve and the USK). When the ratings differed, Valve's was higher in 80.9% of cases (69.5% among 477 games not displaying an available USK label). Divergence was largest where Valve's survey recorded a more severe violence term than the board and for online interactivity, which alone yielded a Valve rating of 12; both processes linked the same descriptor terms to different age bands. In-game purchases were associated with higher Valve ratings mainly for USK decisions before 2023, which could not be attributed to the USK's inclusion of usage risks from 2023.

Keywords: age rating; content rating; self-regulation; video games; Steam; platform governance

1. Introduction

Age ratings are one of the few pieces of information about a video game that most players and parents see before they buy it. A rating condenses a judgement about violence, fear, language, sexuality and, increasingly, purchase and communication features into a single age band and a short list of content descriptors. Traditionally, this judgement has been made by rating boards whose examiners review a game and weigh its content in context [1]. Digital distribution has changed the scale of the task, and many games are now rated through questionnaires that developers complete and that software translates into an age band [2]. On Steam's German storefront, for example, 95.9% of the visible games display a rating generated in this way (Section 4.1).

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

Germany offers an unusual setting in which both kinds of rating can be observed for the same game, under the same law and on the same scale. Since the 2021 amendment of the German Youth Protection Act (Jugendschutzgesetz, JuSchG), game platforms with at least one million users in Germany may offer a game only if it carries an age label assigned under the statutory procedure, by a recognised self-regulation body, or by a recognised automated rating system [2,3]. Steam, whose catalogue comprised 180,201 games on the collection date, responded by issuing its own German ratings, which Valve generates from the developer's content survey together with input from its content review teams and customer feedback, and since 15 November 2024 it no longer displays games without a valid rating to customers in Germany [4]. A game can instead display a rating from the Unterhaltungssoftware Selbstkontrolle (USK), the German self-regulatory body whose committees of youth-protection experts and state representatives issue the statutory labels [1]. Both ratings use the same five age bands (0, 6, 12, 16 and 18 years) and an overlapping vocabulary of content descriptors, and the Steam store data hold both ratings for many games side by side.

This study uses that overlap to ask how a platform's survey-based rating compares with a board rating of the same game. We collected the German rating information for the full Steam game catalogue (180,201 games) on 5 October 2026, linked every game that displays a USK label to the USK's public title database (88,308 records), and compared Valve's rating with the USK rating. Of the 3,719 games whose store data hold both ratings, the primary analysis uses the 909 games whose USK label could be traced to a board decision for a personal computer release. A complementary sample of 477 games that display Valve's rating although a USK decision is on record allows us to assess how the developer's choice of which label to display affects the comparison. The analysis addresses three questions. First, how often do the two ratings agree, and when they differ, is the platform's rating higher or lower? Second, which kinds of content are associated with divergence? Third, do interactive features such as in-game purchases and chat, which the USK has been able to weigh in its age decisions only since 2023 [5], account for part of the difference?

The study contributes to research on age-rating systems in three ways. Earlier work has compared the labels of different rating boards for the same games and examined whether labels generated through the questionnaire of the International Age Rating Coalition (IARC) reflect the games' features [6]. To our knowledge, this study is the first to compare a platform's own rating with

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

statutory board ratings of the same video games under a single legal standard, which holds constant the age bands and the legal test that cross-national comparisons of rating boards must take as given. It interprets the comparison through the distinction between rules and standards [7] and shows where the platform's rating departs from the board's: in the more severe content terms that the platform's survey records, in the weight it gives to online features, and in the different age bands that the two processes associate with the same descriptor terms. It also documents, at catalogue scale, where the information shown to German players about a game's suitability now comes from. For entertainment computing, where age-rating classification has been approached as a prediction problem that treats board ratings as ground truth [8], the findings show how much the label a player sees depends on which process produced it.

2. Background and hypotheses

2.1. Age ratings as consumer information

Age ratings serve parents and players as a short summary of a game's content, and their value depends on whether different raters reach similar judgements about the same game. Early validity studies compared the United States industry ratings with the judgements of other raters. Trained parent panels agreed with the industry when it marked games as unsuitable for children but judged lower ratings to be too lenient, with violence as the main source of disagreement [9]. An independent content analysis of 81 games rated Teen by the Entertainment Software Rating Board (ESRB) found content that could have warranted a descriptor without receiving one in 39 games [10]. A content analysis in this journal found that age levels do not track every content dimension: in a study of 188 popular games, neither the Pan European Game Information (PEGI) nor the ESRB rating level was correlated with the sexualisation of female characters, and games rated PEGI 12 showed on average more sexualisation than games rated PEGI 18 [11].

A second strand compares rating boards with one another. For the same best-selling games, the ratings of the USK, PEGI and ESRB were strongly correlated, most differences arose from the boards' different age categories, and the USK was somewhat more lenient for titles suitable for children under 12 and stricter for adolescents [12]. A study of the warning label for paid random items found that the ESRB and PEGI labelled the same games consistently in only 39.4% of jointly rated cases under a preregistered measure, with most disagreements arising from the ESRB not applying the label to games rated before the label existed [6]. Such comparisons hold

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

the game constant but vary the legal test, the age bands and the descriptor system, which makes it difficult to separate differences in process from differences in standards.

2.2. Questionnaire-based rating

The growth of digital storefronts led rating bodies to delegate the first step of classification to the developer. In the IARC system, the provider answers a single questionnaire, and the system generates labels for each participating rating body according to that body's criteria [2]. The same answers can therefore yield different age bands in different territories: on the Google Play Store, social casino games frequently received PEGI 18 together with ESRB Teen from the same questionnaire input [6]. Questionnaire systems also depend on what developers disclose. Among 100 popular games with paid random items on the Google Play Store, 71.0% did not display the corresponding IARC label, although all unlabelled games that could still be checked continued to contain such items [6]. Ratings declared by developers have also been found to err upwards. For 1,464 apps from the game categories of the Android store that were also available on iOS, maturity ratings selected by developers on Android were higher than the iOS ratings for 18.1% of the apps and lower for 9.7%, which the authors attributed mainly to developers' uncertainty about how to classify content [13]. In the Australian Government's pilot of the IARC tool, random audits decided by the Classification Board found that the tool assigned the Board's category for about 81% of games after weighting; above the lowest category, the tool tended to over-classify, which the evaluation attributed largely to developers over-declaring content [14].

Platform self-ratings are a further step in this development. Valve does not use the IARC system on Steam but generates its own regional ratings from the answers to its content survey [15]. Unlike IARC labels, which are issued under the authority of the participating rating bodies, Valve's German ratings are issued by the platform itself and are subject to its own review [4]. Research in entertainment computing has so far approached age ratings mainly as a classification target: models trained on cover images and descriptions of 39,212 games predicted ESRB and PEGI ratings with an accuracy of 0.678 and 0.584, and performed worst in the intermediate age categories [8]. How a platform's own rating relates to a board rating of the same game has not, to our knowledge, been examined.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

2.3. The German setting

German law requires that games sold to minors or offered on large online platforms carry one of five age labels: approved without age restriction, from 6, from 12, from 16, or no youth approval (18) [3]. Under the JuSchG, labels should also state the main reasons for the rating, and since the 2021 amendment the assessment of developmental impairment may take into account risks that arise from the use of a medium, in particular communication and contact functions, purchase functions and gambling-like mechanisms [3]. Online platforms with at least one million users in Germany may offer a game only if it carries a label assigned in the statutory procedure, by a recognised self-regulation body or certified youth-protection officer, or, where neither exists, by an automated rating system recognised by the supreme youth authorities of the federal states [3].

The USK issues labels through two routes. In the statutory procedure, a game that is submitted for the first time is presented to a committee of four youth-protection experts and a permanent representative of the federal states' youth authorities; titles that are similar in content to an already rated game, such as the same game on another platform, are handled by a smaller committee [1]. In this procedure, state representatives issue the label on the experts' recommendation [5]. Since 2013 the USK has also issued labels for online games and apps through the IARC system, in which the provider answers a questionnaire on content and usage risks and the label is generated from the IARC matrix on the connected storefronts, which include the Google Play Store, the Nintendo eShop, the Microsoft Store, the PlayStation Store and the Epic Games Store but not Steam [2]. From 1 January 2023, the statutory procedure began to consider purchase and communication features and to add notices such as "Chats" and "In-Game-Käufe" to the label; such features can raise the age band when the committee judges the risk to be substantial, taking into account protective measures such as parental controls [5].

Steam is not connected to the IARC system. Developers who distribute on Steam complete Valve's content survey, whose first section generates ratings for several regional rating systems that are displayed on the store page [15]. For Germany, a game's rating can come either from the USK or from what Valve calls its own self-rating process, which takes into account the content survey, input from Valve's content review teams and customer feedback; unless the review teams or the community provide conflicting information, Steam issues the German rating as soon as the questionnaire has been submitted [4]. Developers may enter a USK rating on Steam only if the game went through the rating process directly with the USK and the reviewed version is the

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

version distributed on Steam, and they must not enter a USK rating generated through the IARC process on another storefront [4]. When a USK rating has been entered, the German store page displays it in place of Valve's rating, but the store data continue to hold Valve's rating as well. From 15 November 2024, Steam stopped displaying games without a valid German rating to customers in Germany [4].

The German case therefore holds several things constant that cross-national comparisons cannot. The two ratings refer to the same legal test, the same five age bands and, as the data below show, a partly shared descriptor vocabulary. What differs is the process: a committee that examines the game, or a questionnaire whose answers are translated into an age band by a fixed mapping and corrected only when conflicting information arrives. In the terms of the platform-governance literature, the case combines external governance, a requirement imposed on platforms by law, with an element of self-governance, because the platform itself designs and operates the instrument that meets the requirement [16].

2.4. Rules, standards and the direction of divergence

Kaplow's distinction between rules and standards offers a compact way to describe the difference between the two processes [7]. A rule fixes the content of a norm in advance, so that applying it requires only establishing the facts; a standard leaves the content of the norm to be settled by the adjudicator when a case is decided. Assigning an age rating involves two steps, classifying the content of a game and mapping that classification to an age band, and the two processes differ in who classifies the content and in whether the band follows from the classification by a fixed mapping. In Valve's process, the developer classifies the content by answering the survey, and the age band is issued from the answers [4]; as Section 4.3 shows, the band follows the declared content profile closely. In the USK's process, the committee assesses the game as a whole: the guiding criteria require that all impairments be considered in connection with the overall effect of the game, weigh violence by its context, framing, credibility and graphic style, treat fantastic scenarios as less impairing than realistic ones, and require sexual content to be assessed in the context of the specific game [17]. The criteria, adopted in 2011 and since amended, contain no table that maps content to age bands; the descriptors state the main reasons for an age decision that has already been taken [17]. The distinction is a matter of degree [7]. The USK's decisions are structured by published criteria, and its notices on usage functions start from the applicant's

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

declarations [17], whereas Valve's ratings can be corrected by its review teams and by customer feedback [4]. Still, the platform process is closer to a rule and the board process closer to a standard.

Because a fixed content-to-age mapping takes fewer considerations into account than a board can, the two can disagree in both directions: a simpler rule is both over- and under-inclusive relative to a more complex standard [7]. The distinction therefore implies divergence rather than a particular direction. A direction follows from how the rule's boundaries are set and from how its inputs are declared, and here we offer a conjecture. A platform that issues its own ratings to meet a legal requirement bears the regulatory and reputational cost when a game turns out to be rated too low. A rating that is too high also has costs, because higher bands restrict who may buy or play a game, but these costs fall mainly on the developer and the customer rather than on the platform's compliance with the law. We therefore expect Valve's rating to be higher than the USK rating more often than it is lower (H1). If the board's advantage lies in weighing context, divergence should be largest for content whose assessment depends on framing, such as violence, fear-inducing material and coarse language, and smaller for content whose presence alone tends to determine the band. We therefore expected upward divergence to be concentrated in games whose declared content includes violence, fear-related content or crude language (H2).

Purchase and communication features have become a central concern of youth protection in games. A recent systematic review in this journal found consistent, though mostly cross-sectional, associations between random-reward monetisation and gambling-related harm among young people [18], and in an analysis of Steam reviews, 78.2% of the 417 reviews in the topic on EA FC and Ultimate Team monetisation were negative [19]. The two processes differ in the scope they give to such features. Valve's survey asks about in-game purchases, chance-based purchases, in-game chat and online interactivity for every game, and these elements can enter its rating. The USK's statutory procedure has considered usage risks only since 2023, and under its criteria such a risk raises the age band only if a concrete risk assessment finds it substantial after protective measures are taken into account; purchase functions as such are not regarded as a risk [5,17]. If interactive elements contribute to Valve's rating but were outside the scope of earlier USK decisions, they should be associated with Valve rating a game higher than the USK, and this association should be weaker for USK decisions taken from 2023 onwards (H3).

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

3. Data and methods

3.1. Data collection

All store and database data were collected from publicly accessible sources on 5 October 2026; documents cited as sources were accessed between 5 and 9 October 2026. The data contain no personal information.

Steam catalogue. We enumerated the Steam game catalogue through the store's search interface on the United States storefront, which lists games regardless of their German rating status, restricted to the product type "game" and ordered by name; this returned 180,201 distinct application identifiers. For each game we queried the store's item service with a German storefront context. The response reports whether the game is visible in Germany and which rating the German store page displays, including the issuing body (USK or Valve), the age band, the content descriptors and the interactive elements recorded in Valve's content survey, together with release date, publisher and user-review count. We decoded the four interactive-element codes by comparing the item-service output with the labels shown on 7 store pages, which agreed in all cases: in-game purchases, chance-based in-game purchases, in-game chat and online interactivity.

Paired ratings. For every game whose German store page displayed a USK label (6,407 games), we retrieved the store's application details in the German storefront. This response contains a ratings record that holds the USK rating entered for the game and, separately, Valve's German rating, which is flagged as generated by Valve. Games for which both ratings were present with a valid German age band form the pair sample (3,719 games); 26 games were excluded because Valve's record held a value outside the German bands (13 × 14, 4 × 10 and 9 × "BANNED", a value whose meaning is not documented). The same records also hold the PEGI rating for many games, which we used as an external benchmark (Section 3.5).

USK title database. We downloaded the USK's public title database (88,308 records) through the application programming interface of the USK website. Each record lists the title, publisher, examination date, platform and age band; records of decisions from April 2021 onwards also list content descriptors, and records from 2023 onwards usage notices. The database lists decisions from the USK's examination procedure.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

3.2. Linking USK labels to board decisions

We linked each game that displays a USK label on Steam to the USK database by its normalised title (case, punctuation, trademark symbols and diacritics removed; ß written as ss), using exact matches first and otherwise an order-sensitive string similarity of at least 95 out of 100, provided that both titles contained the same numbers and the same number of words and neither contained a Roman numeral from II to VI. We then assigned each label to one of four provenance classes: (A) a database record for a personal computer release with the same age band; (B) a record with the same age band for other platforms only; (C) a record with the same title but no record with the same age band; and (D) no record with a matching title. When several records qualified, we used the most recent one; using the earliest record changed the classification of USK decisions before and after 1 January 2023 for 3 games. Class A is the primary sample, because it contains labels that can be traced to a board decision for the release type distributed on Steam. Because class A requires the database band to equal the displayed label, labels that differ from the board's decision fall into class C by construction.

We checked the linkage in three manual audits. Of 50 randomly drawn class-A links, 50 referred to the same game; of the 9 class-A links obtained by fuzzy matching, 8 did, and the incorrect link was removed. The removed link concerned a game without a Valve rating in the store data and therefore did not change the pair sample. Of the 55 class-C pairs, Valve's band equalled the board's band in 16 and the displayed label in 20. For a random sample of 50 class-D games, a manual search of candidate records found 3 records that the procedure had missed and 3 games for which only an earlier edition had been rated; for the remaining 44, no corresponding record existed.

3.3. A complementary sample of games that display Valve's rating

Developers decide whether to enter a USK label on Steam, and they may be more inclined to do so when the USK band is lower than Valve's. We therefore also linked games that display Valve's rating to USK database records for personal computer releases. We required an exact match of the normalised title, a title that occurs only once in the visible catalogue and is written in Latin script, a single age band across the matching USK records, and a USK decision within two years of the game's Steam release. In a manual audit of 40 randomly drawn links, 39 referred to the same game, judged by title, publisher and release timing; for the remaining link, identity could

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

not be confirmed. This sample (sample V, 477 games) is the complement of class A under the same display decision: it contains games whose developers did not enter an available USK label, either by choice or because the USK decision did not concern the version distributed on Steam [4]. If developers display the USK label mainly when it is lower than Valve's rating, class A overstates and sample V understates the share of games that Valve rates higher, so that the two samples bracket the share that would be observed without the display decision.

3.4. Measures

Age bands were coded on an ordinal scale (0 = 0 years, 1 = 6, 2 = 12, 3 = 16, 4 = 18). For each pair, the difference Δ is Valve's band minus the USK band, so that a positive Δ means that Valve's rating is higher. Content was measured with Valve's German content descriptors for the game. Because these descriptors draw largely on the USK's descriptor vocabulary, we grouped them into categories defined in advance on the basis of the USK vocabulary: explicit violence ("Drastische Gewalt"), violence ("Gewalt"), fantasy, comic or implied violence, fear-related content, crude language, sexual content, substances (alcohol, tobacco, drugs), gambling themes, and pressure to act or to keep playing. Descriptors that occur in Valve's vocabulary but not in the USK's were assigned to the closest category: frightening content ("Ängstigende Inhalte") and suicide ("Selbstmord") to fear-related content, swear words ("Schimpfwörte" [sic]) to crude language, and sexualised violence to sexual content; crime ("Verbrechen") was not assigned. No game in the pair sample carried a gambling or pressure descriptor. Following Section 2.4, violence, fear-related content and crude language were classified as context-sensitive, and sexual content and substances as presence-based; in the regression models, horror, frightening content and suicide were entered separately. Interactive elements were coded as four binary indicators. The examination date of the matched USK record was used to distinguish decisions taken before and from 1 January 2023.

For 350 class-A games, the USK record lists the board's own descriptors or usage notices (334 list descriptors and 16 only notices; 278 have decisions from 2023 onwards, when notices were introduced). The USK lists the main reasons for its decision, on average 1.41 descriptors per game for the records that list descriptors, against 2.01 Valve descriptors for the same games. For these games we coded the highest violence descriptor of each process on a common scale (implied, comic or fantasy violence; violence; explicit violence); the USK term for war themes

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

("Kriegsthematik", 24 records) was not assigned to a level. We also recorded whether the USK record carries a notice for in-game purchases or chats; comparisons involving notices are restricted to decisions from 2023 onwards.

3.5. Analysis

Agreement between the two ratings was summarised by the share of identical age bands, Cohen's weighted kappa with linear and quadratic weights [20] and Krippendorff's alpha for ordinal data [21], with 95% confidence intervals (CIs) from 2,000 bootstrap resamples of games. As a benchmark, we computed the same quadratically weighted kappa between the USK and PEGI ratings, and between the Valve and PEGI ratings, for the class-A games that also hold a PEGI rating, mapping PEGI 3, 7, 12, 16 and 18 to the ordinal positions of the German bands.

H1 was tested with an exact binomial test of the share of discordant pairs in which Valve's band is higher, complemented by Bowker's test of symmetry, the Stuart–Maxwell test of marginal homogeneity and a Wilcoxon signed-rank test of Δ. Because games rated 0 by the USK cannot be rated lower by Valve and games rated 18 cannot be rated higher, we repeated the test without both extreme bands. Shares are reported with Wilson CIs. H2 and H3 were tested with logistic regressions of an indicator for Valve's band being higher, estimated for class-A games with a USK rating below 18. The H2 model includes the content categories and the interactive elements; the H3 model includes the interactive elements, an indicator for USK decisions from 2023 onwards, its interactions with in-game purchases and online interactivity, and an indicator for context-sensitive content. Both models control for the USK band and the year of the USK decision, so that the 2023 indicator measures a change net of the trend. Because several categories are almost perfectly separated, the models were estimated by Firth's penalised likelihood, with profile penalised-likelihood CIs and penalised likelihood-ratio tests, and results are reported as odds ratios (ORs). The coefficients of the content categories and of the interactive elements and their interactions were adjusted for multiple testing within each model with the Benjamini–Hochberg procedure. For H3 we also report the underlying cell counts with Fisher's exact test, average marginal effects of in-game purchases (the mean change in the predicted probability when the purchase indicator is set to 1 instead of 0 for the games of each period) from an ordinary logistic regression with bootstrap CIs, and an ordinary logistic regression with standard errors clustered by publisher.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

Three further analyses are exploratory. The first describes how Valve's age band depends on the declared content profile, defined as the combination of content descriptors and interactive elements, across all games that display a valid Valve rating. It also compares how predictable each process's band is from each process's content profile among the 334 games whose USK record lists descriptors, measured as the leave-one-out share of games whose band equals the most frequent band of the other games in the same profile, with CIs from 2,000 bootstrap resamples of profiles. Because the USK records its descriptors as reasons for a decision already taken, this comparison describes the association between descriptors and bands in each process and does not show how either process reaches its decisions. The second compares the bands that each process assigns to the same descriptor terms and cross-tabulates the highest violence descriptor of each process for the games for which both list a violence descriptor. The third compares classes A and D in a penalised logistic regression of a higher Valve rating on the USK band and an indicator for class D. Finally, to examine the mechanism behind H3, we cross-tabulated games with purchases by USK band and period, compared their mean USK and Valve bands across periods, and estimated an ordinary least-squares model of the USK band on Valve's band, the purchase indicator, the period indicator and their interaction.

The hypotheses, sample definitions and main models were fixed in a written analysis plan before the full-catalogue data were analysed; the plan was preceded by a pilot comparison of 30 games. The following deviations from the plan were made: the planned ordinary logistic regressions were replaced by Firth regressions because of separation; the planned proportional-odds models (as an alternative specification and for the comparison of classes A and D) were replaced by penalised binary models, because the cumulative logits showed separation at the lowest thresholds; the content categories were extended to Valve-specific descriptors and fear-related content was split; and sample V, the PEGI benchmark, the descriptor comparisons, the average marginal effects and the analysis of the mechanism behind H3 were added. The analysis was carried out in Python with pandas, SciPy and statsmodels; the Firth estimator was implemented by the author and checked against closed-form results.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

4. Results

4.1. Where the German ratings on Steam come from

Of the 180,201 games in the catalogue, 159,457 were visible on the German storefront on the collection date and 20,744 (11.5%) were not; the item service does not report why a game is not visible. Among the visible games, 152,990 (95.9%) displayed Valve's own German rating and 6,407 (4.0%) displayed a USK label; 60 visible games displayed no rating. Valve's rating is therefore the age information that German Steam users see for the great majority of games. A small number of Valve's German ratings carried values that are not German age bands (558 games rated 14, 43 rated 10 and 2 rated "l"); these were excluded from all analyses of Valve's ratings, leaving 152,387 games.

Table 1 traces the construction of the analysis samples. Of the 6,407 games with a USK label, 3,745 (58.5%) also held a Valve rating in the store data. The 2,662 USK-labelled games without a Valve rating were older than those with one (median Steam release year 2017 versus 2022) and had a similar share of class-A labels (22.1% versus 24.4%). The primary sample consists of the 909 class-A games with both ratings; the USK decisions in this sample date from 1995 to 2026 (median 2021), and 308 were taken in 2023 or later. The class-A games are older and far more widely played than the typical game in the catalogue (median Steam release year 2022 versus 2024; median number of user reviews 2,220 versus 3). Sample V contains older games with fewer reviews than class A (median release year 2016, median 484 reviews) and more games rated 0 by the USK (32.3% versus 16.1%).

Table 1. Construction of the analysis samples.

Step	Games	Note
Games in the Steam catalogue (type game)	180,201	
Visible on the German storefront	159,457	
Displaying Valve's German rating	152,990	95.9% of visible
Displaying a USK label	6,407	4.0% of visible
USK label without Valve rating in the store data	2,662	

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

USK label and Valve rating in the store data	3,745	58.5% of USK-labelled
Excluded: Valve value outside the German bands	26	
Pair sample (valid bands in both ratings)	3,719	
Class A: board record, PC release, same band	1,501 (909)	primary sample
Class B: board record, other platforms only	486 (308)	
Class C: title found, different band	91 (55)	
Class D: not located in the USK database	4,329 (2,447)	
Sample V: Valve rating displayed, USK PC record	477	outside the pair sample; USK label not displayed
Class-A sample (n = 909)	**Value**	
USK band 0 / 6 / 12 / 16 / 18	146 / 129 / 323 / 187 / 124	
Valve band 0 / 6 / 12 / 16 / 18	62 / 61 / 373 / 247 / 166	
Games with at least one Valve content descriptor	80.0%	
Most frequent Valve descriptors	Drastische Gewalt (300); Gewalt (241); Alkohol / Tabak (205); Sexuelle Andeutungen (139); Derbe Sprache (133); Schimpfwörte (128)	
Interactive elements: purchases / chance-based purchases / chat / online interactivity	10.0% / 1.5% / 7.0% / 48.7%	
Median Steam release year; median year of USK decision	2022; 2021	
USK decisions from 2023 onwards	33.9%	

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

Median number of user reviews	2,220

Note. Upper panel: for the provenance classes, the first number refers to all USK-labelled games and the number in parentheses to games in the pair sample. PC = personal computer. Lower panel: characteristics of the primary sample.

4.2. Agreement and direction (H1)

In the primary sample, the two ratings placed 47.5% of the games in the same age band (95% CI [44.3%, 50.9%]). The quadratically weighted kappa was 0.62 [0.58, 0.66], the linearly weighted kappa 0.47 [0.43, 0.51] and Krippendorff's ordinal alpha 0.65 [0.61, 0.69] (Table 2). The 729 class-A games that also hold a PEGI rating provide a benchmark. On these games, the USK and PEGI ratings agreed considerably more closely (quadratically weighted kappa 0.88 [0.86, 0.90]) than Valve's rating did with the USK (0.61 [0.57, 0.65]) or with PEGI (0.62 [0.58, 0.66]). Valve's rating was higher than the PEGI rating in 287 games and lower in 82, whereas the PEGI rating was higher than the USK rating in 153 games and lower in 101. The PEGI rating on Steam is entered by the developer, PEGI's bands 3 and 7 were treated as equivalent to the German bands 0 and 6, and the two boards often rate the same submitted material, so that the benchmark is favourable to agreement between the boards.

Where Valve's rating and the USK rating differed, Valve's rating was higher for 386 games and lower for 91, so that 80.9% of the discordant pairs (95% CI [77.1%, 84.4%]) lay above the diagonal ($p < 0.001$, exact binomial test; Wilcoxon signed-rank test $p < 0.001$). The marginal distributions differed ($\chi^2(4) = 155.8$, $p < 0.001$, Stuart–Maxwell test), and the mean difference was 0.42 ordinal bands. Excluding games rated 0 or 18 by the USK, for which a difference is possible in one direction only, Valve's rating was higher in 81.7% of the 276 + 62 discordant pairs. The results were similar for the 802 games whose USK decision fell within two years of their Steam release, for which the release on Steam is most likely to correspond to the version examined by the USK (45.5% identical bands; quadratically weighted kappa 0.60 [0.56, 0.64]; Valve higher in 82.2% of the discordant pairs, $p < 0.001$). Fig. 1 shows the cross-tabulation with row percentages; the largest off-diagonal cells are USK 0 to Valve 12 (99 of 146 games rated 0 by the USK), USK 6 to Valve 12 (87 of 129), USK 12 to Valve 16 (109 of 323) and USK 16 to Valve 18 (59 of 187).

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

In sample V, in which developers did not display an available USK label, 52.6% of the 477 games received the same band (95% CI [48.1%, 57.1%]), and Valve's rating was higher in 69.5% of the discordant pairs (95% CI [63.0%, 75.4%]; $p < 0.001$). The upward share is smaller than in class A restricted to the same two-year window (82.2%; see above), and the USK 0 to Valve 12 cell is less frequent (38.3% of games rated 0 by the USK, against 67.8% in class A). The difference remained when the USK band, online interactivity and release year were held constant (penalised regression of a higher Valve rating on an indicator for sample V among 1,143 games with a USK rating below 18 from class A within the two-year window and sample V, OR = 0.41, 95% CI [0.30, 0.55], $p < 0.001$; without controls OR = 0.50, 95% CI [0.39, 0.64], $p < 0.001$), which is consistent with developers displaying the USK label more often when it is lower than Valve's rating. The display decision affects the two samples in opposite directions, and both lie well above one half; pooling sample V with the class-A games in the same two-year window, Valve's rating was higher in 77.8% of 663 discordant pairs (95% CI [74.5%, 80.8%]). The direction also held when the class-A sample was restricted to exact title matches with a publisher similarity of at least 80 (85.5%, n = 516). H1 is supported.

Fig. 1. Cross-tabulation of USK and Valve age bands for the 909 games whose USK label could be traced to a USK decision for a personal computer release (class A). Each cell shows the number of games and the percentage of its row; column totals are given below the axis labels. Cells on the diagonal (bold) are identical ratings; cells to the right of the diagonal are games that Valve rated higher. Games rated 0 by the USK can only be rated equal or higher by Valve, and games rated 18 only equal or lower.

Table 2. Agreement between Valve's rating and the USK rating.

	Class A	**Sample V**	**Class B**	**Class D**	**All pairs**
Games	909	477	308	2,447	3,719
Identical band [95% CI]	47.5% [44.3%, 50.9%]	52.6% [48.1%, 57.1%]	51.3% [45.8%, 56.5%]	48.3% [46.3%, 50.3%]	48.2% [46.7%, 49.9%]
Valve higher	386	157	116	823	1,345
Valve lower	91	69	34	441	581
Share higher among	80.9% [77.1%,	69.5% [63.0%,	77.3% [69.8%,	65.1% [62.4%,	69.8% [67.7%,

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

discordant pairs [95% CI]	84.4%]	75.4%]	83.8%]	67.7%]	71.9%]
Kappa, quadratic weights [95% CI]	0.62 [0.58, 0.66]	0.67 [0.61, 0.72]	0.69 [0.64, 0.75]	0.64 [0.62, 0.67]	0.65 [0.63, 0.67]
Mean difference (ordinal bands)	0.42	0.29	0.36	0.22	0.28

Benchmark on the 729 class-A games with a PEGI rating (quadratically weighted kappa [95% CI]): USK and PEGI 0.88 [0.86, 0.90]; Valve and PEGI 0.62 [0.58, 0.66]; Valve and USK 0.61 [0.57, 0.65].

Note. Class C (n = 55) is included in all pairs only. Exact binomial tests of the share of discordant pairs in which Valve is higher: $p < 0.001$ in all columns. PEGI ratings are developer-entered on Steam; PEGI 3 and 7 were treated as equivalent to German 0 and 6.

4.3. How each process relates content descriptors to age bands

Valve's ratings follow the declared content closely. Across the 152,387 games with a valid Valve rating, the combination of content descriptors and interactive elements formed 893 distinct profiles, and 90.3% of the games received the most frequent age band of their profile. Fig. 2 shows the age bands for the profiles without content descriptors and for all single-descriptor profiles that occur in at least 150 games. Games without descriptors or interactive elements were rated 0 in 98.5% of cases; online interactivity alone led to 12 (99.8% of 6,470 games) and in-game purchases alone to 12 (100.0% of 312). In-game purchases combined with chance-based purchases led to 16 (97.7% of 44 games), and all four interactive elements to 16 (99.1% of 115). Fantasy violence alone led to 6 (98.9%), violence alone to 12 (85.8%), horror alone to 16 (100.0%) and explicit violence alone to 16 (74.5%) and 18 (25.5%); shares may not sum to 100% because of rounding. Within class A, 79.9% of Valve's ratings equalled the most frequent catalogue band of the same profile, close to the 82.2% expected from the catalogue for the same mix of profiles; the remaining ratings lay above the catalogue mode in 156 games and below it in 24.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

The two processes use partly the same descriptor terms, but they associate them with different bands (Table 3). Among the 334 class-A games whose USK record lists descriptors, games whose highest violence descriptor at the USK was "Gewalt" (violence) were rated mostly 16 by the USK, whereas games whose highest violence descriptor at Valve was "Gewalt" were rated mostly 12 by Valve; the USK used implied, comic or fantasy violence mostly with 12, Valve mostly with 6 or 12. This statement rests mainly on the violence terms and horror, because the USK lists the other shared terms for few games. Each process's band could be predicted similarly well from its own descriptors. The leave-one-out share of games receiving the most frequent band of their profile was 78.4% [71.0%, 84.8%] for Valve's band within Valve's profiles (294 games in 66 profiles) and 78.5% [71.1%, 84.8%] for the USK's band within the USK's profiles (291 games in 44 profiles). Across processes, it was 49.0% [39.3%, 58.6%] for the USK's band within Valve's profiles and 57.6% [47.6%, 68.3%] for Valve's band within the USK's profiles. With profiles of at least five games, the four shares were 81.1% [71.5%, 89.3%], 80.7% [72.4%, 87.8%], 48.5% [34.0%, 61.8%] and 56.4% [44.7%, 69.0%]. Because the USK's descriptors state the reasons for a decision already taken, this similarity does not show that the USK applies a fixed mapping.

Fig. 2. Valve's German age band by declared content profile: the three profiles without content descriptors and all profiles with a single content descriptor and no interactive element that occur in at least 150 of the 152,387 games with a valid Valve rating. Rows are ordered by the most frequent band. Descriptors are shown in Valve's German wording with an English gloss.

Table 3. Age bands associated with the same descriptor terms in the two processes, and highest violence descriptors for the same games (class-A games whose USK record lists descriptors or usage notices, n = 350).

Descriptor term (highest violence term for violence rows)	**Valve: games by Valve band**	**USK: games by USK band**
Fantasy, comic or implied violence	6: 15; 12: 21; 16: 1 (n = 37)	6: 26; 12: 99; 16: 9 (n = 134)
Gewalt (violence)	12: 73; 16: 22; 18: 2 (n = 97)	12: 12; 16: 59; 18: 9 (n = 80)

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

Drastische Gewalt (explicit violence)	16: 60; 18: 81 (n = 141)	16: 5; 18: 37 (n = 42)
Horror	16: 10; 18: 2 (n = 12)	16: 5; 18: 14 (n = 19)
Sexuelle Andeutungen (sexual innuendo)	12: 20; 16: 29; 18: 22 (n = 71)	12: 1; 16: 1; 18: 1 (n = 3)
Sexuelle Inhalte (sexual content)	16: 6; 18: 14 (n = 20)	18: 4 (n = 4)
Derbe Sprache (crude language)	6: 1; 12: 14; 16: 21; 18: 26 (n = 62)	16: 2 (n = 2)
Drogen (drugs)	12: 5; 16: 11; 18: 22 (n = 38)	12: 1; 16: 2; 18: 2 (n = 5)

Highest violence term, Valve compared with USK	**Games**	**Valve higher / lower (all)**	**Valve higher / lower (no online interactivity)**
Valve more severe	139	68 / 3	44 / 1
Same term	79	12 / 9	2 / 6
Valve less severe	1	0 / 0	0 / 0
Only the USK lists a violence term	37	18 / 7	13 / 5
Only Valve lists a violence term	56	25 / 9	9 / 5

Note. Upper panel: the 334 games whose USK record lists descriptors; the two columns refer to different sets of games, namely those for which each process lists the term; a game appears in a violence row of a process if that term is the highest violence descriptor the process lists for it. The USK lists the main reasons for its decision (1.41 descriptors per game, Valve 2.01), so its non-violence rows rest on few games. Lower panel: games by the highest violence descriptor of each process; the first three rows contain the 219 games for which both processes list a violence descriptor; 38 games list no violence term in either process. The USK term for war themes was not assigned a level; treating it as violence gives 148, 80 and 1 games in the first three rows (n = 229), and treating it as explicit violence 130, 96 and 3 (n = 229).

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

4.4. Where the divergence arises

The USK lists fewer descriptors than Valve, so the absence of a USK violence descriptor does not mean that the USK found no violence. Among the 219 games for which both processes list a violence descriptor, Valve's highest violence descriptor was more severe than the USK's in 139 games, the same in 79 and less severe in 1 (Table 3). Where Valve's descriptor was more severe, Valve's rating was higher in 68 and lower in 3 of these games (95.8% of the discordant pairs, 95% CI [88.3%, 98.6%]). Where the descriptors agreed, Valve's rating was higher in 12 games and lower in 9; among those without online interactivity, it was higher in 2 and lower in 6. Where only the USK listed a violence descriptor (37 games), Valve's rating was higher in 18 and lower in 7, and where only Valve listed one (56 games), higher in 25 and lower in 9. Upward divergence thus occurs across these groups and is most pronounced where Valve's survey records a more severe violence term than the board lists as a reason. In the small group of games with the same term and no online interactivity (42 games, 25 of them rated 18 by the USK, where Valve cannot be higher), Valve's rating was not more often higher. Because the same term corresponds to different bands in the two processes, these data do not separate differences in how content is classified from differences in how the terms are used.

The second place where upward divergence is pronounced is online interactivity. Of the 99 class-A games rated 0 by the USK and 12 by Valve, 92 had online interactivity and 83 had no content descriptor. Games with online interactivity and no descriptor accounted for 103 of the 386 upward differences in class A. Setting these games aside, Valve's rating remained higher in 76.1% of the discordant pairs, and setting aside all games with online interactivity, in 70.1%. Among the 278 games with a USK decision from 2023 onwards whose record lists descriptors or notices, both processes recorded in-game purchases for 44 games, Valve alone for 5 and the USK alone for 51; where both recorded purchases, Valve's rating was higher in 21 of 23 discordant pairs (95% CI [73.2%, 97.6%]), and where only the USK recorded them, Valve was higher in 25 and lower in 3 games. The USK attached a chat notice to 70 records; Valve recorded in-game chat for 36 and online interactivity for 68 of these games. The two processes thus record interactive features differently, and the features recorded by Valve may be defined more narrowly than the USK's notices.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

4.5. Content categories (H2)

H2 expected upward divergence to concentrate in context-sensitive content. The planned contrast with presence-based content could not be estimated, because only 2 class-A games with a USK rating below 18 declared presence-based content alone, and the unadjusted shares do not follow the expected pattern. Valve's rating was higher for 65.8% of the games with explicit violence (n = 187), 75.9% with sexual content (n = 29) and 61.6% with sexual innuendo (n = 112), but only for 32.3% with violence (n = 235), 39.6% with fantasy or comic violence (n = 111) and 40.0% with crude language (n = 210); games without any descriptor were rated higher by Valve in 60.1% of cases (n = 178). The low unadjusted share for violence reflects that most of these games were rated 12 by the USK (67.2%), the band to which Valve's mapping assigns violence alone. In the penalised regression that holds the USK band, the decision year and the interactive elements constant (Table 4), the descriptors associated with a higher Valve rating were explicit violence, violence, horror, suicide, sexual content and sexual innuendo, which Valve's mapping places at 12 or above; fantasy or comic violence, crude language and substances (which combines alcohol and tobacco, mostly rated 6, with drugs, rated 12 or above) were not statistically significant (for substances, after Benjamini–Hochberg adjustment), and the estimate for frightening content, recorded for few games, was uninformative. Several of these estimates exceed 100 because Valve rated almost every game with the descriptor above the USK band; they indicate the direction of the association rather than its size. H2 is not supported in the form in which it was planned.

4.6. Interactive elements and the 2023 change (H3)

In the penalised regression, in-game purchases were associated with a higher Valve rating for USK decisions taken before 2023 (OR = 8.51, 95% CI [2.85, 30.80], $p < 0.001$), and this association was weaker for decisions from 2023 onwards (OR = 0.20, 95% CI [0.05, 0.71], $p = 0.012$; $q = 0.020$; Table 4, Fig. 3). On the probability scale, games with in-game purchases were 0.37 [0.21, 0.50] more likely to be rated higher by Valve when the USK decision predated 2023 and 0.11 [-0.05, 0.26] when it did not (average marginal effects with 95% CIs). In the underlying counts, Valve rated 24 of 28 versus 24 of 44 games with in-game purchases higher in the two periods (Fisher's exact test $p = 0.010$), and 18 of 21 versus 23 of 42 when the sample was restricted to USK decisions within two years of the Steam release ($p = 0.024$). The interaction was of similar size with standard errors clustered by publisher (OR = 0.18, $p = 0.005$, 263 publishers) and when the earliest instead of the latest matching USK record was used (OR = 0.20,

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

95% CI [0.05, 0.70], p = 0.012); among exact title matches with high publisher similarity it pointed in the same direction but was not statistically significant (OR = 0.30, 95% CI [0.06, 1.26], p = 0.100). Online interactivity was associated with a higher Valve rating (OR = 3.10, 95% CI [2.06, 4.72], p < 0.001); its interaction with the decision period pointed in the same direction but was not statistically significant (OR = 0.52, 95% CI [0.26, 1.03], p = 0.061). In-game chat showed no association.

The pattern therefore matches the direction that H3 predicted, but not its mechanism. Valve's mapping assigns in-game purchases alone to 12 (Section 4.3), and among games rated 0 or 6 by the USK, Valve rated all 20 games with purchases higher in both periods. The period difference arose entirely among games rated 12 or 16 by the USK (14 of 18 versus 14 of 34; Fisher's exact test p = 0.019), where purchases alone cannot raise Valve's rating; in this group, Valve's ratings of 16 or 18 (16 of 18 games before 2023, 25 of 34 afterwards) reflect content descriptors and, for 5 and 8 games respectively, chance-based purchases. The evidence on the USK side is mixed. Among the games with purchases in the model sample, the mean USK band rose slightly from 1.79 to 1.95 (ordinal positions; n = 28 and 44) while the mean Valve band fell from 2.82 to 2.66, so that both sides contribute to the smaller difference; 30 of the 34 post-2023 games with purchases at USK 12 or 16 carry the USK's purchase notice; and holding Valve's band constant, the USK band of games with purchases did not rise significantly more after 2023 than that of other games (interaction 0.22 [-0.17, 0.60], p = 0.268 in an ordinary least-squares model of the USK band, treated as an interval scale, on Valve's band, the purchase indicator, the period indicator and their interaction, with heteroskedasticity-robust standard errors; n = 785). Because the analysis by USK band conditions on the quantity that the mechanism would change, it cannot exclude that the USK raised some bands for games with purchases. The period difference therefore cannot be attributed to the USK's inclusion of usage risks, nor can this explanation be ruled out. The test of H3 thus does not support its proposed mechanism. Sample V cannot provide a check, because it contains only seven games with purchases below USK 18, none with a decision from 2023 onwards.

The models also show a trend over the decision year: holding the other variables constant, Valve's rating was more often higher for more recent USK decisions (OR = 1.07, 95% CI [1.04, 1.11], p < 0.001 per year). Among games with none of the four interactive elements, Valve's rating was higher for 33.2% of the games with a USK decision before 2023 (n = 271) and 45.5%

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

of those with a later decision (n = 121); the increase was concentrated among games rated 16 by the USK (23.5%, n = 68, versus 53.1%, n = 32), that is, at the boundary between 16 and 18. Whether Valve's survey or descriptor list changed during this period cannot be determined from the data.

Fig. 3. Share of class-A games with a USK rating below 18 that Valve rated higher than the USK, by interactive element recorded in Valve's content survey and by the date of the USK decision (unadjusted shares with 95% Wilson CIs). Games can have more than one element; chance-based purchases are omitted because they were recorded for few games; "None of the four" contains games without any of the four elements. The sample contains 519 games with a USK decision before 2023 and 266 with a later decision.

Table 4. Penalised (Firth) logistic regressions of a higher Valve rating (class A, USK rating below 18).

Predictor	**H2 content model: OR [95% CI]; p**	**H3 model: OR [95% CI]; p**
Explicit violence	> 100 [> 100, –]; p < 0.001; q < 0.001	
Violence	9.11 [3.59, 26.55]; p < 0.001; q < 0.001	
Fantasy, comic or implied violence	1.32 [0.61, 2.89]; p = 0.480; q = 0.567	
Horror	> 100 [> 100, –]; p < 0.001; q < 0.001	
Frightening content	0.65 [0.04, 11.05]; p = 0.762; q = 0.794	
Suicide	> 100 [> 100, –]; p < 0.001; q < 0.001	
Crude language	1.54 [0.88, 2.72]; p = 0.135; q = 0.175	
Sexual innuendo	2.25 [1.09, 4.64]; p = 0.028; q = 0.045	

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

Sexual content	22.24 [6.05, > 100]; p < 0.001; q < 0.001	
Substances	1.88 [1.04, 3.45]; p = 0.037; q = 0.053	
Context-sensitive content (any)		3.38 [1.85, 6.40]; p < 0.001
In-game purchases	3.34 [1.41, 8.30]; p = 0.006; q = 0.011	8.51 [2.85, 30.80]; p < 0.001; q < 0.001
In-game chat	0.87 [0.32, 2.41]; p = 0.794; q = 0.794	0.53 [0.24, 1.17]; p = 0.118; q = 0.118
Online interactivity	7.53 [4.67, 12.45]; p < 0.001; q < 0.001	3.10 [2.06, 4.72]; p < 0.001; q < 0.001
USK decision 2023 or later		1.24 [0.73, 2.10]; p = 0.426
In-game purchases × 2023 or later		0.20 [0.05, 0.71]; p = 0.012; q = 0.020
Online interactivity × 2023 or later		0.52 [0.26, 1.03]; p = 0.061; q = 0.076
Year of USK decision (per year)	1.04 [1.01, 1.08]; p = 0.020	1.07 [1.04, 1.11]; p < 0.001
Games (games with a higher Valve rating)	785 (386)	785 (386)

Note. Both models include indicators for the USK band (reference: 0). CIs are profile penalised-likelihood intervals; p-values are from penalised likelihood-ratio tests; q-values are Benjamini–Hochberg-adjusted within each model across the content categories and interactive elements. ORs above 100 indicate near-separation: Valve rated almost every game with the descriptor above the USK band; they show the direction, not the size, of the association, and their upper confidence limits are extremely large and uninformative. Numbers of games with each descriptor: explicit violence 187, horror 20, suicide 19, sexual content 29, frightening content 10. In the H3 model, the main effects of in-game purchases and online interactivity refer to USK decisions before 2023.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

4.7. Labels not located in the USK database

The 2,447 class-D pairs showed the same overall pattern with a weaker upward tendency. Agreement was similar (48.3% identical bands; quadratically weighted kappa 0.64), but Valve's rating was higher in 65.1% of the discordant pairs (95% CI [62.4%, 67.7%]) compared with 80.9% in class A. Holding the USK band constant, Valve's rating was less often higher in class D than in class A (OR = 0.45, 95% CI [0.38, 0.54], $p < 0.001$ for class D, penalised regression of games with a USK rating below 18).

5. Discussion

The German Steam storefront makes it possible to compare a platform's rating with a board's rating of the same game under one law and one set of age bands. For the games whose USK label could be traced to a board decision, the two ratings agreed on the exact band for about half of the games, considerably less often than the USK and PEGI ratings of the same games agreed with each other, and when they differed, Valve's rating was mostly higher. The comparison of class A with sample V shows that the developer's choice of which label to display inflates this share, but the share remains well above one half in both samples.

5.1. Where the platform's rating departs from the board's

The documentary evidence places the two processes at different points between a rule and a standard [7]. The USK's guiding criteria require an assessment of the overall effect of a game and give no table that maps content to age bands [17], whereas Valve's rating is issued from the developer's survey answers unless conflicting information arrives [4], and the data show that it follows the declared content profile closely. The comparison of the two ratings for the same games identifies where the outcomes differ, though not how either process reaches its decisions.

The upward divergence is most pronounced in two places. Among games with violent content, it is strongest where Valve's survey records a more severe violence term than the board lists as a reason for its decision, although it also occurs where only one of the two processes lists a violence term. This pattern resembles findings from other settings in which ratings derived from developers' own declarations were more often higher than lower than a reference rating [13,14]; in the Australian evaluation of the IARC tool, this over-classification occurred above the lowest category and was attributed largely to developers over-declaring content [14]; the present data cannot distinguish such over-declaration from cautious boundaries in the platform's mapping.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

Because the board uses the violence terms and horror with higher age bands than Valve does (Table 3), the shared vocabulary does not measure content on a common scale, and the data cannot separate whether the survey classifies the same content as more severe or whether the two processes simply use the terms differently. The second is online interactivity, which Valve's mapping places at 12 even when no other content is declared. Under the USK's criteria, such a feature raises the age band only if a concrete risk assessment finds a substantial risk [17]. This difference in weighting accounts for the largest single cell of disagreement, and because it concerns how the two processes treat the feature rather than when they rated the game, it may persist.

The results for in-game purchases show the limits of reading such differences as effects of a rule change. The association between purchases and a higher Valve rating was concentrated in USK decisions taken before the USK began to consider usage risks in 2023 [5], as H3 expected, but the closer analysis could not attribute it to the proposed mechanism: Valve's mapping lifts purchases alone only to 12, the period difference arose among games that the USK had rated 12 or 16, and the change in the difference came partly from the Valve side. It may reflect differences between the games behind older and newer decisions, for example in their content or in features added after the board's decision, although the difference largely remained when the comparison was restricted to decisions close to the Steam release. The trend towards more upward divergence for recent USK decisions, which among games without interactive elements was concentrated at the boundary between 16 and 18, remains unexplained; it may reflect changes in Valve's survey, in the USK's decisions or in the games that obtain a USK label.

5.2. Implications

For players and parents in Germany, the label shown on Steam is in most cases the platform's rating. For the games studied here, which are older and more widely played than the typical game in the catalogue, this label tended to be higher than the board's rating of the same game. A higher label is not necessarily more protective: in an experiment with Dutch children and adolescents, restrictive age labels made games more attractive to young players [22]. Nor is it necessarily more informative: a rating that bundles online features, purchase functions and content into the same age band conveys less about why a game received it. The descriptors and interactive-

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

element notices shown with the label carry this information, which underlines the value of displaying them prominently, as the JuSchG requires for statutory labels [3].

For regulators, the German case shows that a platform can meet a mandatory rating requirement at catalogue scale through its own questionnaire, and that the resulting ratings can be compared with board decisions using public data. A routine comparison of this kind could serve as a calibration check for automated rating systems, similar in purpose to the quality checks that rating bodies carry out on questionnaire-based labels [2]. The finding that divergence is most pronounced where the survey records more severe content terms and online features suggests where such a check would be most informative: in the questions and categories through which developers describe their games and in the weight given to interactive features.

For research on age ratings in entertainment computing, the results caution against treating any single label as ground truth. Classification models trained on board ratings [8] learn the decisions of a process that assesses content in context; a platform's labels follow a survey and a fixed mapping, and the two differ systematically even for the same game. Even the shared descriptor terms carry different meanings in the two processes. Studies that relate ratings to content, as in the analysis of character sexualisation across rating levels [11], may benefit from recording which process produced each label.

5.3. Limitations

The study has several limitations. First, the pair sample depends on developers' decisions to display a USK label; class A and sample V are affected by this decision in opposite directions, and the pooled estimate rests on title matching within a two-year window. Second, the date on which Valve's rating was generated is not observable, while the USK decisions in the primary sample span three decades, so that some pairs may refer to different versions of a game, for example a German version examined by the board and an international version distributed on Steam, or a version to which purchase functions were added later. The direction of the results was the same for USK decisions taken before and after the Steam release (79.0% and 81.8% of the discordant pairs), but the time trend and the analysis of purchases in Section 4.6 show that the timing of decisions matters. Third, Valve's process combines the content survey with review-team input and customer feedback [4], and for games that display a USK label, Valve's rating is not shown to users; these ratings largely followed the catalogue mapping, with deviations mostly

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

towards higher bands. Fourth, content was measured with the descriptors recorded by Valve, which reflect the developer's declarations, and for the USK only with the main reasons it lists, available from 2021 onwards; the two processes also record interactive features differently. Fifth, the planned classification of sexual content and substances as presence-based differs from the USK's guiding criteria, which assess sexual content in the context of the specific game [17], and very few games declared such content alone, so H2 could not be tested as planned. Sixth, the analysis of in-game purchases rests on few games, and its period difference could not be attributed to the mechanism that H3 proposed. Seventh, the origin of the labels that could not be located in the USK database (class D) remains open. Finally, the study covers one platform in one jurisdiction; whether other platforms' ratings show the same pattern is a question for further research.

6. Conclusion

When the same game is rated under the same law by a rating board and by the platform that sells it, the two ratings agree for about half of the games studied, and the platform's rating is higher in most of the remaining cases. On Steam's German storefront, where the platform's rating is the label most players see, this divergence is most pronounced in games for which the platform's survey records more severe content terms and in games with online features, and the two processes attach different age bands to the same descriptor terms. Comparing platform ratings with board ratings for the same games offers a practical way to make such differences visible and to calibrate automated rating systems.

Data availability

The compiled dataset is not publicly shared, because it consists of records taken from two third-party sources, the Steam store and the USK title database, and permission to redistribute these records has not been obtained from their providers. Both sources remain publicly accessible but change over time; Section 3 describes the collection, linkage and coding procedures in enough detail for a comparable dataset to be compiled from the same sources. The analysis plan, the analysis code and the audit files are available from the corresponding author on reasonable request. The data contain no personal information.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

References

[1] Unterhaltungssoftware Selbstkontrolle, Arten der Alterskennzeichnung [Types of age rating]. https://usk.de/fuer-unternehmen/spiele-und-apps-pruefen-lassen/arten-der-alterskennzeichnung/ (accessed 5 October 2026).

[2] Unterhaltungssoftware Selbstkontrolle, Spiele und Apps im IARC-System [Games and apps in the IARC system]. https://usk.de/fuer-unternehmen/spiele-und-apps-pruefen-lassen/spiele-und-apps-im-iarc-system/ (accessed 5 October 2026).

[3] Jugendschutzgesetz (JuSchG) [German Youth Protection Act] of 23 July 2002 (BGBl. I p. 2730), last amended by Art. 12 of the Act of 6 May 2024 (BGBl. 2024 I No. 149), §§ 10b, 12, 14, 14a. https://www.gesetze-im-internet.de/juschg/ (accessed 5 October 2026).

[4] Valve Corporation, Age ratings mandatory in Germany, Steamworks Documentation. https://partner.steamgames.com/doc/gettingstarted/contentsurvey/germany (accessed 5 October 2026).

[5] Unterhaltungssoftware Selbstkontrolle, In-Game-Käufe, Chats und Lootboxen: USK erweitert Prüfkriterien [In-game purchases, chats and loot boxes: USK extends its examination criteria], press release, 14 December 2022. https://usk.de/usk-pressemitteilung-umsetzung-neues-jugendschutzgesetz/ (accessed 5 October 2026).

[6] L.Y. Xiao, Beneath the label: unsatisfactory compliance with ESRB, PEGI and IARC industry self-regulation requiring loot box presence warning labels by video game companies, R. Soc. Open Sci. 10 (2023) 230270. https://doi.org/10.1098/rsos.230270.

[7] L. Kaplow, Rules versus standards: An economic analysis, Duke Law J. 42 (1992) 557–629. https://doi.org/10.2307/1372840.

[8] C. Balim, Deep multimodal fusion for video game age rating classification, Entertain. Comput. 55 (2025) 100980. https://doi.org/10.1016/j.entcom.2025.100980.

[9] D.A. Walsh, D.A. Gentile, A validity test of movie, television, and video-game ratings, Pediatrics 107 (2001) 1302–1308. https://doi.org/10.1542/peds.107.6.1302.

[10] K. Haninger, K.M. Thompson, Content and ratings of teen-rated video games, JAMA 291 (2004) 856–865. https://doi.org/10.1001/jama.291.7.856.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

[11] W. Pan, Y. Zhou, Changing trends in female character portrayals: representation and stereotypes across popular video games (2011–2022), Entertain. Comput. 57 (2026) 101107. https://doi.org/10.1016/j.entcom.2026.101107.

[12] L. Dogruel, S. Joeckel, Video game rating systems in the US and Europe: Comparing their outcomes, Int. Commun. Gaz. 75 (2013) 672–692. https://doi.org/10.1177/1748048513482539.

[13] Y. Chen, H. Xu, Y. Zhou, S. Zhu, Is this app safe for children? A comparison study of maturity ratings on Android and iOS applications, in: Proceedings of the 22nd International Conference on World Wide Web, ACM, 2013, pp. 201–212. https://doi.org/10.1145/2488388.2488407.

[14] Australian Government, Department of Communications and the Arts, Evaluation of the IARC pilot, August 2016, released under freedom of information (FOI 25-1920). https://www.infrastructure.gov.au/sites/default/files/documents/foi-25-1920-pilot-international-age-rating-coalition-classification-tool.pdf (accessed 9 October 2026).

[15] Valve Corporation, Content survey, Steamworks Documentation. https://partner.steamgames.com/doc/gettingstarted/contentsurvey (accessed 5 October 2026).

[16] R. Gorwa, What is platform governance?, Inf. Commun. Soc. 22 (2019) 854–871. https://doi.org/10.1080/1369118X.2019.1573914.

[17] Unterhaltungssoftware Selbstkontrolle, Leitkriterien der USK für die jugendschutzrechtliche Bewertung von digitalen Spielen [Guiding criteria of the USK for the youth-protection assessment of digital games], last amended December 2024. https://usk.de/die-usk/grundlagen-und-struktur/grundlagen/ (accessed 6 October 2026).

[18] J. Arnedo-Moreno, J.J.P. López, P. Rodríguez-Rivera, A. Baena, D. Aranda, The impact of video game use of dark patterns and random reward mechanisms on the youth: A systematic literature review, Entertain. Comput. 57 (2026) 101122. https://doi.org/10.1016/j.entcom.2026.101122.

[19] B. Bora, Digital gamblification in video games: Player backlash, topic structure, and regulatory event effects in Steam reviews, Entertain. Comput. 58 (2026) 101223. https://doi.org/10.1016/j.entcom.2026.101223.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871

764 [20] J. Cohen, Weighted kappa: Nominal scale agreement provision for scaled disagreement or
765 partial credit, Psychol. Bull. 70 (1968) 213–220. https://doi.org/10.1037/h0026256.

766 [21] A.F. Hayes, K. Krippendorff, Answering the call for a standard reliability measure for
767 coding data, Commun. Methods Meas. 1 (2007) 77–89.
768 https://doi.org/10.1080/19312450709336664.

769 [22] M. Nije Bijvank, E.A. Konijn, B.J. Bushman, P.H.M.P. Roelofsma, Age and violent-content
770 labels make video games forbidden fruits for youth, Pediatrics 123 (2009) 870–876.
771 https://doi.org/10.1542/peds.2008-0601.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7610871