Utrecht University

Virtual realities and categorical identity

An analysis of categorical membership and user evaluation in the video game industry

Program: Master 'Innovation Sciences'
Date: 23-06-2025
Supervisor: Koen Frenken
Second assessor: Matthijs Punt
Written by: Emre Özmuk (6592228)
e.c.ozmuk@students.uu.nl
Word count: 23,255 words

Abstract

This thesis aimed to analyse the influence of VR technology on categorical membership and user evaluation within video games. As the video game industry is steadily growing, which is partly driven by the integration of VR technology, it is of interest to understand how VR technologies have shifted genre classifications and user expectations. Categorical membership refers to an organisation or its entity to be assigned to a specific category based on the exhibited characteristics shared within this category. Therefore, the following research question was composed: ***What is the influence of VR technologies on categorical membership and user evaluation within the video game industry?***

This study adopted a mixed method research design. The research design utilised a quantitative and a qualitative analysis of which utilised the quantitative analysis tested six hypotheses. This showed that VR video games tend to be classified under more genres, make use of less conventional genre combinations than in traditional gaming and receive lower user ratings as they are disruptive of nature. To provide more nuance to these findings, a qualitative analysis was implemented, which analysed negative user sentiments and identified recurring themes that explain these negative sentiments. These themes showed and confirmed that VR video games exhibit characteristics of a disruptive innovation, which has initially lower performances and stretches out the categorical boundaries of genres.

The findings suggest that while VR technology opens up numerous opportunities for developers it also introduces challenges. Therefore, it is of importance that developers take these identified issues and the disruptive nature of VR technology into consideration, during the development of their video games. A good balance between creativity and user expectations can help innovation penalty and support the integration of VR technology in the gaming industry.

Table of contents

1. Introduction

Over the past decade, rapid technological progression has transformed numerous industries through significant investments in research, education and innovation (Khan, 2020). These investments have incentivised advancements across various domains of science and technology, which have resulted in the Fourth Industrial Revolution (Schwab, 2024). Innovations driven by the Fourth Industrial Revolution, such as artificial intelligence (AI) and 5G internet, have redefined how users interact with digital environments and are blurring the lines between the physical and digital world (Khan, 2020; Schwab, 2024). Among these, Virtual Reality (VR) and Augmented Reality (AR) have introduced new levels of immersion, which contribute to this convergence of the physical and digital world.

The introduction of these transformative technologies has resulted in significant increases in investments (Machala, 2022). This increase in investments for AR/VR is particularly evident in the video game industry, where funding rose from 93.8 million USD in 2013 to 762.4 million USD in 2021, causing an increase of 712.79% (Statista, 2024). Many stakeholders expect AR/VR to be the future of the gaming industry (Melnikava et al., 2021).

The introduction of AR/VR games potentially revolutionised the way video games are experienced through the introduction of entirely new dimensions of interaction, through new gameplay mechanics, and immersion. AR enhances the physical world by overlaying vibrant and visual elements onto the user's real-life environment, completing their experience with additional components. VR also creates a more immersive experience by engaging users in a fully simulated environment, enabling the creation of alternate worlds limited only by the player's imagination (Tulane University, n.d.). Examples of such technologies are for instance Pokémon GO, which is a prime example of an AR video game and Meta Quest, which is a VR technology that allows people to immerse themselves in a 3D space (Meta, 2024).

The impact that the introduction of AR/VR in the gaming industry has, is studied by several researchers. In the research by Tiwari et al. (2023), it is shown that AR/VR gaming can provide risk-free settings, which allow individuals to make errors without facing permanent consequences. This capability can prominently be seen in serious games within high-risk industries, such as the medical, aviation or engineering fields, where students and trainees can utilise AR/VR applications to practice specific tasks. These simulations allow students and trainees to develop their skills and knowledge, which ultimately enhances their efficiency and safety when performing these tasks in real-world settings.

Van Nguyen et al. (2022), show possibilities that AR/VR has with regards to digital tourism. According to Van Nguyen et al. (2022), it is possible to reconstruct digital heritage objects in 3D, which then can be used to create, for instance, a virtual museum. This would be an efficient way of providing cultural and historical values of heritages to communities in a more engaging manner. According to Theodoropoulos (2022), AR/VR does not only support offsite cultural experiences, but also onsite cultural experiences. The incorporation of AR/VR in museums can increase museum visiting motivation through its dynamic engagement.

A final example of this is the application of VR in educational settings, which increased student engagement. The incorporation of a VR environment led to higher levels of active participation and interest among students, due to the interactive learning experience (Shah et al, 2024). The integration of VR technology makes it possible to design complicated scenario's that otherwise would not have been possible to create or create safely. Therefore, students are able to immerse themselves in a complicated, stress-inducing situation, which will enable them to effectively and safely manage similar situations in the future.

These researches highlight the significant potential of AR/VR gaming through its immersion, however AR/VR remains an emerging and immature technology (Bian, 2023). Consequently, this research will focus exclusively on VR video games. Due to the limited maturity of AR/VR technologies, the application of AR technologies within the video game industry remains relatively scarce. This has resulted in a limited number of AR video games available on gaming platforms, making AR video games an unviable subject for analysis.

The influence of the immersion that VR brings is still not completely researched. Aforementioned researches show that VR can create video games in multiple various directions. However, the new possibilities that are introduced by VR, also give rise to certain uncertainties. The new capabilities that are achieved through VR may influence the way that video games are classified. For instance, certain combinations of video game genres may become more relevant, due to the enhanced immersion or new gameplay mechanics offered by VR, while conventional genre combinations may lose their relevance. As a result, video games classification may experience significant changes due to the introduction of these technologies.

This shift in classification may also impact the experience of users. Video game users may have developed specific preferences for certain genre combinations or may favour standalone genres. However, the change in classification from the introduction of VR technologies, could expose video game users to new genre combinations, which can influence their levels of satisfaction. While some users may find these new combinations very intriguing and engaging, others may perceive them as mismatched or dull, leading to divisions in user evaluation.

For this reason, this research could be highly valuable for game developers , that are preparing on starting or already working on VR gaming projects. According to Statista (2025), 34% of video game developers are currently working on projects for Meta Quest, while 26% are developing games for Steam VR, which showcases the impact of VR in todays market. It is essential for these developers to understand the key factors that influence user expectations and preferences, especially regarding the classification of their video games. Developers could invest heavily in a single genre or a specific combination of genres based on past success, however the introduction of VR as a new variable may have changed the dynamics of genre utilisation and classification. Failing to account for these shifts could results in tremendous losses, which could have been avoided with a better understanding of these trends. Therefore, this research aims to provide insightful results for businesses to refine their strategies, which enhances their potential of developing a successful VR game.

Consequently, this research will evaluate whether the introduction of VR technology in video games is associated with a shift in genre classification and whether these changes influence user ratings. Put differently, VR technology may make it more difficult to assign video games as members to categories and thus genres. This leads to the following research question:
What is the influence of VR technologies on categorical membership and user evaluation within the video game industry?

2. Theoretical framework

This section provides a framework consisting of four core bodies of literature to evaluate the influence of VR on video game classification and evaluation: (1) categorical membership, (2) online feedback mechanisms, (3) genres in video games and (4) technological redefinition. These core bodies will function as the foundation for the formulation of six hypotheses aimed at addressing the research question.

2.1 Categorical membership

According to Durand et al (2017), categories act as disciplinary mechanisms that bring order to organisational interactions in markets. Categories are groupings of entities that help audiences understand the world by highlighting the characteristics that are most important, which makes recognition and actions easier (Durand et al, 2017). Audiences structure their environment by organising items into relevant categories, thereby imposing order and facilitating a more comprehensible and systematic understanding of the world. These categories are defined by their respective *categorical boundaries* (Vergne, 2014). Categorical boundaries specify the criteria that determine the inclusion and exclusion of organisations or its offerings to a particular category. These boundaries are typically set through a list of category attributes, which an organisation or its offerings must posses to be considered part of the category. An organisation or its offerings can only be included in the category if it possesses all of the attributes (Vergne, 2014). According to Frenken and Punt (2023), certain items serve as representative items of a category, called *exemplars*. They function as a prototypical example that best embody the defined attributes of the respective category.

An organisation or its offerings is recognised as a *category member* by audiences if the organisation or its offerings hold all the attributes of the category boundaries and if the organisation is focused enough (Vergne, 2014). The second requirement is just as important as the first requirement, as it creates legitimacy for audiences. This can particularly be a problem in organisations that try to diversify into different categories. An example is Harley Davidson, the renowned motorcycle manufacturer, which in the nineties attempted to diversify its brand portfolio by producing perfume (Medium, 2021). Audience members were accusing Harley Davidson of producing more products just to increase sales and did not adopt the products. Even though, the product falls within the category boundaries of that industry, the firm was not focused enough and was still associated with just being a motorcycle manufacturer.

In some instances, organisations or its offerings, may be recognised only as a *partial member*. *Partial membership* occurs when the organisation or its offerings exhibit some, but not all, of the defining category attributes and when the partial member is not focused enough, in terms of its offerings, to be seen as a member of the category (Vergne 2014). Partial membership can thus be caused by solely exhibiting some of the defining attributes from a lack of focus or from a combination of both factors.

The notion of categorical membership extends beyond individual classifications, as categories are often embedded in larger categorical systems (Boghossian & David, 2021). This reflects the principle of *categorical granularity*, wherein subcategories belong to basic categories, which in turn belong to superordinate categories forming a hierarchical nested system (Durand and Cudennec, 2023). Within such nested systems, categories are vertically organised; higher level categories function as umbrella categories and meaning from these categories can flow to subordinate categories (Wry & Lounsbury, 2013; Durand & Boulongne, 2017).

The nested structure of categories plays a crucial role in giving meaning and perceived value of a category. When a category's position shifts within the system or when a new umbrella category is introduced, the interpretation and significance of a category may be, either positively or negatively, altered. This dynamic is particularly evident in the introduction of novel products by entrepreneurs. When entrepreneurs introduce novel products onto the market, audience members notice specific and unique traits of the product, which may, over time, contribute to the formation of a new categorical prototype (Boghossian & David, 2021). Products within the same category are evaluated based on their alignment with this prototype, where products with better alignment are perceived as more valuable products opposed to their counter parts.

However, the classification of an organisation or its offerings as fully, partially or not belonging to a specific category is rarely straightforward. Categorical boundaries inherently possess a degree of ambiguity, which is often referred to as *category fuzziness* (Hannan, 2010). This concept questions the clearly defined boundaries between category members and non-members (McCloskey et al, 1978). A high degree of fuzziness indicates blurred boundaries of a category, which makes it challenging to distinguish between those that belong within the category and those who do not (Motamendi, 2014). Vergne (2014), adds to this notion that a category's boundaries are fuzzier when there is a disagreement between producers and audiences about which attributes should be considered for a category's definition and if category members have multiple memberships in several categories. High fuzziness allows for audience members to be more free in understanding and using categories (Frenken and Punt, 2023).

Categorical boundaries and their fuzziness function as a stepping stone in the article by Frenken and Punt (2023), who explore the relationship between the innovation types; *radical*, *disruptive*, *breakthrough* and *incremental* innovations, and categorical membership. An innovation is considered categorical radical, when society categorises and institutionalises this innovation into a completely new category instead of in an existing category, and appreciates it for its novelty. An innovation is seen as categorical disruptive, when the innovation is categorised and institutionalised within an existing category, but by stretching the boundaries of this category often by ignoring one or more attributes, and being positively assessed as it makes an artefact or practice more accessible. An innovation is considered categorical breakthrough, when the innovation is categorised and institutionalised within an existing category and is assessed positively due to its major improvements in known attributes. On the contrary, an innovation is called categorical incremental when the innovation is categorised and institutionalised in an existing category and assessed positively due to its minor improvements in known attributes.

This view on innovation related to the theoretical notion of *categorical imperative*, which is a notion that has been extensively explored by Zuckerman (1999, 2017). According to Zuckerman (1999, 2017), a categorical imperative refers to the pressure that individuals or organisations face from their audience to conform to established expectations associated with recognised categories. Market actors encounter legitimacy constraints, when their products or services deviate from conventional products or services that are more commonly seen in said markets. Frenken and Punt (2023) also address this concept of *innovation penalty* through *categorical penalties*; when an entity deviates from the exemplar of a category, audience member could experience greater difficulty in accurately categorising the entity. This could lower the perceived valuation of the entity and potentially lead to a lack of investment or sales. This phenomenon is often seen with entities that are either radical or disruptive, as they create new categories or stretch existing categorical boundaries. This can obstruct audience members from directly comprehending their value, which often results in a lower perceived valuation.

Categorical membership theory provides a framework for understanding how users perceive and unconsciously respond to the classification of products. However, the measurement of such responses is operationalised through feedback mechanisms. A feedback mechanism that is particularly important in todays society is the utilisation of online feedback mechanisms, where user ratings and reviews are used to assess whether certain products are accepted by consumers.

2.2 Online feedback mechanisms

User reviews and ratings are forms of online feedback mechanisms (Dellarocas, 2003). They are important ways of distinguishing whether items are widely approved or refused. Although, the two concepts are commonly used interchangeably, they are distinct in meaning (Lackermair, 2013). A product review refers to a textual review detailing a customer's experience with a product, providing feedback. In contrast, a product rating represents the customer's evaluation expressed on a predefined scale, such as between 1 and 5 (Lackermair, 2013). Therefore, product reviews offer a qualitative measure of their opinion, while product ratings offer a quantitative measure of their opinion.

Online feedback mechanisms are widely recognised for their role in establishing trust in offerings and supporting users to make judgements and choices about offerings. However, their influence extends beyond these functions, exerting a broader impact on organisations (Dellarocas, 2003). Organisations can use the feedback generated from these mechanisms for customer acquisition, customer retention, brand building and product development (Dellarocas, 2003). Organisations can be fundamentally altered or developed based on insights from online feedback mechanisms. This feedback offers great input for decision making processes, which enables organisations to alter their strategies to align with customer preferences.

Online feedback mechanisms impact not only organisations, but also potential customers. According to Lackermair (2013), the first indicator in the decision making of a product is often the user rating of said product. As aforementioned, user ratings are expressed on a predefined scale, which often involves a Likert-scale. Potential customers can quickly and efficiently asses a product's user rating through online feedback mechanisms. For instance, if a product has an average rating of 2 out of 5 based on 500 reviews, then potential customers will be less inclined to purchase such a product as the low rating signals dissatisfaction among customers.

User reviews have the same impact on potential customers, but are seen as a less efficient mechanism (Lackermair, 2013). User reviews provide more comprehensive insights about customer dissatisfaction with certain products. By offering textual feedback, customers can express their opinions in detail about aspects they appreciate and dislike. Nevertheless, this more detailed approach results in potential consumers having to invest more time in research, which is not always preferred. Potential customers perceive longer reviews as more informative compared to shorter ones. However, the increased length can complicate comparison between reviews, which has negative influence on decision making efficiency. This ultimately may lead to certain aspects to remain unclear (Lackermair, 2013).

User ratings and reviews have also been used in empirical research about categories. In particular, empirical studies have shown that offerings that have only partial membership receive, on average, lower ratings and reviews. Hsu et al. (2009), for example, have analysed the effects of multiple categorical membership on the appeal and commercial success of films released in the United States. The films analysed in this empirical study were classified into 17 distinct genres, encompassing genres ranging from action to romance. The findings indicated that films targeting a broader audience with less specialisation in specific genres tend to perform worse in terms of appeal and box office success compared to films with a narrower, more focused genre orientation.

Kovacs and Hannan (2010) also examined the concept of partial membership, but in the context of restaurants and other food services. The study analysed restaurants based on the variety of cuisines they represent and evaluate the impact on consumer assessment. The findings indicate that restaurants offering a high contrast combination of categories, such as serving both Indian as Mexican cuisines, are more difficult for audience member to interpret. This ambiguity in categorical identity may result in lower evaluations compared to restaurants that focus on one or more low-contrast categories.

These empirical studies thus confirm the categorical imperative theory, which explains why innovations that diverge from conventional category norms often struggle to secure approval and acceptance from relevant stakeholders. Accordingly, the theories of online feedback mechanisms and categorical membership will serve as the foundational framework for examining whether categorical classification within video game industry adheres to similar dynamics of approval and refusal.

2.3 Genres in video games

Following previous research on restaurants and movies, one can evaluate user ratings and reviews to examine and analyse how users assess video games based on their categorical membership. User feedback, expressed through reviews, provides valuable insights into players' satisfaction and perceptions of the game's quality. On the other hand, sales data serve as an indicator of the game's commercial appeal and its ability to capture a broad audience. However, the influences of these two indicators are not perfectly correlated; in principle, a game that has lower reviews can still generate a lot of sales and, reversely, a game that has good reviews may still have low sales.

According to Hsu (2006), who researched the influence of genres on box office success in movies, producers that target a broad area of the market have access to greater potential revenue, despite lower appeal. The total size of the market is increased, when producers target multiple genres, thereby creating opportunities to access a greater pool of resources. This could also be the case for video games. Video games that focus on a single genre, therefore limiting the market size, may tend to get high appeal and they target a well-defined audience with niche interests. In contrast, video games that integrate multiple genres into their game expand their audience base and provide a much more diverse experience, may get lower appeal, but can, in principle, attract a broader range of potential consumers.

As aforementioned, through the production of multi-genre games a larger proportion of the audience is targeted, thus increasing sales. However, this may have negative impacts on the experience of particular audiences (Hsu, 2006). These films that target different genres create a difficult scenario as they try to appeal to multiple audiences, where they have difficulties interpreting the identities of the movies. This results in audiences disapproving the entire production, as it has a poor fit with their expectations and preferred taste (Hsu, 2006). This is also supported by Leung (2009) and Zuckerman (2017), which shows that actors that span multiple categories are devalued relative to others that focus on a single category.

The extent to which audiences may disapprove a multi-genre video game may vary, according to the actors that are involved in the production of these video games. According to Hsu et al. (2009), novel actors benefit from associating their product with a single category, because audiences interpret their niche experience as a signal of skill in this specific category. Therefore, products that target multiple categories are automatically perceived as an inferior product, with regards to category specific products. Regarding incumbent actors this may matter less, as they have already proven themselves in their respective branch of work and therefore will not be seen as unskilled if they tend to produce a product that adheres to multiple categories.

In conclusion of the previous arguments, the following hypotheses have been formulated:

Hypothesis 1: Video games that are classified under multiple genres receive lower user ratings compared to video games that are categorised under a single genre.

Hypothesis 2: Video games that are classified under multiple genres produced by novel developers receive lower user ratings compared to video games that are classified under multiple genres produced by incumbent developers.

As VR opens up possibilities for new genre combinations and classifications, due to their new capabilities, which could cause difficulties with categorisation by audience members, the third hypothesis is:

Hypothesis 3: Video games that integrate VR technology are more likely to receive lower user ratings compared to non-VR video games, due to their disruptive nature.

2.4 Technological redefinition

The gaming industry has experienced continuous improvements as more technological advances are made each year. Initially, innovative developments in gaming were focused on improvements of resolution and visual quality. For example, the first commercially successful game, Atari Pong, functioned on a display of 210x160 pixels (Weitkamp, 2019). Technological progress has enabled the development of video games to support high definition resolutions, such as 4K or 3840x2160 pixels, which has increased the visual experience for users tremendously (Yarnell, 2022).

While advancements in visual quality and resolution continue, with efforts toward the commercial implementation of 8K technology, the focus has shifted towards enhancing immersion in gaming experiences (Bove, 2018). Technological developments resulted in the creation of immersive technologies such as VR. According to Verhulst et al. (2021), VR is defined as the creation of a fully computer generated environment, resulting in a completely synthetic world.

According to Neo-Schumpeterian growth theory, new innovations emerge to successfully challenge and eventually replace existing innovations within society (Schumpeter, 1942). These radical innovations use new technology and open up new markets, which require specific organisational capabilities, such as agility and willingness to take risks, which incumbent firms often lack. Henkel et al. (2015) highlight that new entrants are typically better positioned to engage in radical innovation due to their organisational flexibility, lack of established customer relationships and the absence of existing products. These characteristics are all crucial in order to engage in radical innovation.

Building on this, Christensen (1997) introduced the concept of disruptive innovation. This concept also involves the emergence of new technologies, but instead of creating a new market, it disrupts an existing market. Disruptive innovations have unique characteristics, which are used to replace traditional technologies and completely shift an industry. At the time of introduction onto the market, these innovations typically underperform, compared to the industry standards, but offer higher value on affordability and accessibility. As the technology matures, its performance improves, which ultimately results in the innovation to surpass the dominant technology and to reshape the market. Similarly as radical innovations, disruptive innovations are often introduced by new entrants that have the flexibility, organisational capabilities and lack of established customer relationships to engage in this type of innovation.

On the contrary, incumbent firms are defined by their established customer relationships and focus on improving their existing products (Tripsas & Gavetti, 2000). These firms often engage in incremental innovation rather than radical or disruptive, because of path dependency. They have made tremendous investments in specific capabilities and processes, which keeps them locked into a set trajectory (Schumpeter, 1942). Diverging from this trajectory requires a structural reorganisation of the firm, which is costly and perilous. Therefore, incumbents are more inclined towards incremental innovation, which aligns with their organisational structure. Moreover, diversifying into different products and possibly markets could put their existing customer base at risk, therefore short term and risk free investments are more preferred. Thus, new entrants and incumbent firms fulfil different roles in driving technological advancements.

Accordingly, emerging technologies, such as VR are not likely to be pursued by incumbent firms, because of the uncertainty the technologies bring with them. The applications of VR technologies are still immature with a global user penetration rate of 1.3% that own a VR headset, compared to a penetration rate of 45% for video games in 2023 (Statista, 2023). Therefore, those who engage in the development of these technologies could very well be mostly new entrants.

Despite this low penetration rate, the implementation of VR technologies has shown its potential of increased user engagement through their immersive experiences. Examples, also mentioned in chapter 1, include serious gaming in high-risk training settings (Tiwari et al. 2023), digital tourism and heritage (Van Nguyen et al. 2022; Theodoropoulos (2022), and interactive learning in education (Shah et al., 2024).

The applications of VR technologies, however, extend beyond these specific settings, as the gaming industry also recognised its potential opportunities. The introduction of VR technologies within the video game industry has opened up numerous complex possibilities for the gaming community through innovative features. These new possibilities that VR bring into the gaming industry may disrupt how users evaluate video games in its entirety. VR video games are perceived differently due to their nature, which may lead to differences in user evaluation compared to traditional video games, and difficulties and disagreements on how to classify VR video games into genres. Critics and users are accustomed to evaluating video games to established preferences and criteria. The introduction of new characteristics, such as enhanced immersion, may change the way video games should be evaluated. This shift in evaluation is not only influenced by the possibilities that VR introduces, but also by their current limitations. As VR technology is still immature, video games that use these technologies may showcase weaker performances in graphics or bad user interfaces, which all have influence on user evaluation. Therefore, the video game industry may be in need of a reconsideration of evaluation criteria.

These advancements and limitations from VR technologies do not only have the potential to change user evaluation completely, but may be able to establish entirely new categories within the industry. The capabilities that VR introduce, may be able to combine previously seen unfitting genres to create a novel and captivating combination. Certain genres now could work well together, only because of, for instance, the enhanced immersion that VR offer. The same principle applies to the limitations of VR technologies, as certain aspects VR does not excel in could inhibit the combination of genres, which otherwise would go fine together.

To the best of my knowledge, the influence of VR technologies on categorical redefinition and user evaluation has yet to be thoroughly researched.

In conclusion of the previous arguments, the following hypotheses have been formulated:

Hypothesis 4: Video games that integrate VR technology are more likely to be classified under multiple genres compared to non-VR video games

Hypothesis 5: Video games that integrate VR technology are more likely to feature genre combinations that were previously less common in traditional video games.

As VR represent a new technology opening up new possibilities and genre combinations that newcomers are generally better at leveraging compared to incumbents focusing on incremental innovation, the last hypothesis is:

Hypothesis 6: Video games that integrate VR technology and are classified under multiple genres receive higher user ratings when developed by novel developers compared to incumbent developers.

3. Methods

This research adopted a mixed method research, which has made use of quantitative and qualitative research approaches. The methodology provides insights into the research design, data collection, operationalisation of concepts and research quality of this mixed method research.

3.1 Quantitative analysis

3.1.1 Research design

The research has utilised a cross-sectional study design for the quantitative analysis. It was a deductive analysis, in the sense that several hypotheses have been predetermined. These hypotheses were tested and have either been accepted or rejected based on the findings. The assessment of the hypotheses was done through regression analyses. The datasets that have been used for this analysis originate from 2024, as the data reflect a single point in time, the research was appropriated for a cross-sectional study. This cross-sectional approach has enabled the identification of patterns and trends regarding categorical membership within the video game industry and evaluation of said video games.

The datasets have been sourced from Steam, the largest video game distribution service, to ensure most representative outcomes for this research (Statista, 2024). With an estimated 409 million video games sold in 2022 and over 100,000 titles available, Steam is widely recognised as the leading platform in the video game industry (Statista 2024; Backlinko, 2024). Due to Steam's extensive consumer base and variability of games that are offered, Steam provides an excellent starting point for the analysis of the implementation of VR technologies in video games on categorical membership and user evaluation, which has established extensive results.

3.1.2 Data collection

This research has employed desk research for the collection of data. The data has been collected from secondary sources, such as Kaggle or GitHub, that utilise data which originates from Steam. The datasets were obtained through the use of application programming interfaces (API's) and uploaded on these sites. An initial attempt was made to carry out the API's independently, in order to secure reliability and consistency of the data, but this proved to be not viable. Therefore, the datasets were sourced from secondary sources and a manual review of 50 observations was conducted to verify the consistency and reliability of the data. This manual review cross-referenced the datasets with the data from Steam, in order to check the consistency of the datasets. This ensured validity and accuracy of the datasets that have been used in this research.

The required datasets had to include several important topics for them to be eligible for use. This included details on video game genres, the number of downloads a video game has, user ratings, total amount of user ratings and the developers of the video game. A thorough review of the secondary sources has been conducted for the identification of suitable datasets. The information that was acquired through these sources are all publicly available and therefore have not resulted in any legal issues with regards to Steam's privacy.

Individual datasets that were derived from the secondary sources were insufficient, because they did not include all the important information. Therefore, multiple datasets have been combined in order to create the most comprehensive and complete dataset. The different datasets have been combined with the help of R. This dataset totalled 83,008 different observations. However, the combined datasets had to be cleaned in order to remove all observations that were deemed invalid, due to unusable values in key variables. Specifically, observations with a value of 0 for estimated ownership or user ratings of -1 were excluded from the dataset. This resulted in a dataset with a total of 45,244 observations, which served as the final sample size of the study.

3.1.3 Variable construction

Based on the different hypotheses, several dependent variables had been identified for the analysis of the hypotheses. For hypothesis 1, 2, 3 and 6 the dependent variable was the user ratings. Hypothesis 4 examined the number of genres a video game is classified under, while hypothesis 5 focused on the commonness of genre combinations.

The concept of user ratings is easily presented on Steam. Steam users are given the option to evaluate a video game based on their experiences by providing either a “thumbs up” to indicate positive feedback or a ”thumbs down” to indicate negative feedback. All the individual ratings are then accumulated by Steam to produce an overall score ranging from 0% to 100% which reflects overall customer satisfaction[1]. This aggregated rating will be utilised as the primary measure for user ratings in the analysis.

The concept of the number of genres a video game has been classified under is straightforward to operationalise, as the datasets provide a clear list of genres for each game. This dataset encompasses a total of 11 distinct genres, which includes genres such as action, sports, racing and strategy. This ensures an accurate identification of genre variability for the analysis. It is important to note that Steam employs a specific genre classification system. According to Valve Corporation (2022), video games on the platform can be categorised into three levels of genres: super-genre (e.g. Action), genre (e.g. Platformer) and sub-genre (e.g. Precision Platformer). This study focussed exclusively on the super-genre level, as it captures the most noticeable differences between games. This has also been proven by Guzsvinecz & Szucs (2023), where they used super-genres or top-level genres (TLG’s) in their research. The assignment of these super-genres is carried out by the game developers themselves, who determine under which genres their games should be classified (Guzsvinecz & Szucs, 2023).

Regarding the prevalence of genre combinations required a more nuanced approach. To determine which genre combinations were regularly observed and which not, a list of all unique genre combination was composed, which resulted in a total of 1,458 different combination of genres. The frequency of combinations was calculated based on the reoccurrence of the combinations throughout the dataset. This frequency served as an indicator of whether a genre combination was usual or unusual to see in video games. Consequently, a genre combination received a commonness score, which ranged from 1 to 2,496. Higher values for the commonness score indicated that it was more common to observe the genre combination in video games.

---

[1]TotalReviews=PositiveReviews+NegativeReviews
ReviewScore=PositiveReviews/TotalReviews
Rating=ReviewScore−(ReviewScore−0.5)*2−log10(TotalReviews+1) source: SteamDB (2017)

These dependent variables did act interchangeably, as they served as a dependent variable in one hypothesis and could serve as an independent variable in others. For instance, the number of genres assigned to a video game acted as an independent variable in hypotheses 1, 2, 3 and 6, however in hypothesis 4 it acted as a dependent variable. However, the amount of genres was a categorical variable, therefore when the variable functioned as an independent variable it had to be transformed into a dummy variable, so that it could be included in the regressions.

Two additional independent variables were utilised in this study. The first independent variable was the inclusion of VR elements in video games. Video games that would have access to VR technologies had received the tag "VR", which made it possible to distinguish the video games from non-VR video games.

The second independent variable was the developer's background with regards to whether it could be classified as a novel or incumbent developer. In this study, the distinction between novel and incumbent developers was based on the types of games they produced. Developers characterised as incumbent were those producing AAA games, which are video games that have significant backing and funding from major publishers (Prolific Studio, 2024). Contrarily, developers of AA or Indie games, which operate on a significantly smaller budget and smaller team, were characterised as novel developers. The identification of the type of developer has been done manually, as datasets would only include information on whether the video games were Indie games. Therefore, developers from video games that were not Indie were analysed to determine whether they develop AAA video games.

Several control variables have been incorporated in order to prevent omitted variable bias, which occurs when a third variable is absent in an analysis, while this variable causes the independent and dependent variable (Wilms et al. 2021). In the study by Philp & Nepomuceno (2024), several control variables had been considered for their research on the influence of reviews on product usage. One such variable was whether a video game included a multiplayer mode. The presence of a multiplayer functionality may influence the genre classification of video games, the type of developer behind the game, and ultimately user ratings. Another control variable that was included in the analysis is the release date of a video game. Video games released prior to the introduction of VR technologies did not experience the potential classification shift and therefore their evaluations may differ from those games released after the introduction of VR technologies.

The estimated number of owners has also been taken into account, as this has influence on the representation of user evaluation, where games with a lot of owners tend to have more user ratings. This does not necessarily have negative or positive consequences for user ratings, but it does give a more nuanced view of user evaluation and therefore had to be included in the analysis. The amount of supported languages in a video game was included as a control variable. A greater number of supported languages increases the accessibility of the video game to a more diverse audience. People from all different kind of countries are able to engage with the game. This broader accessibility enables people of different backgrounds to form an opinion on the video games and share them through user ratings. Therefore, it was deemed important to include this variable in the analysis.

The final control variable that had been included was whether a video game had an official age rating of 16+. Video games with age ratings could have different genre combinations, as certain immersive genres are more effectively expressed without age restrictions. These differences in game expression could influence user evaluations as well. The inclusion of these variable as a control variable ensured a more nuanced analysis of the relationships between the variables.

3.1.4 Regression analyses

These variables have all been used in linear regression models, also known as ordinary least squares (OLS) regression, for either the acceptance or rejection of the hypotheses 1, 2, 3, 5 and 6. The dependent variables regarding these hypotheses were all continuous and therefore are suitable for linear regression. This method was used to minimise the sum of square differences between the observed and predicted values (XLstat, n.d.). This assumption of a linear relationship has been made due to the fact that several similar studies, such as the studies by Philp & Nepomuceno (2024) and Brunt et al. (2020) showed that this was the case.

With regard to hypothesis 4, the dependent variable represents count data, which makes linear regression deemed as inappropriate. To determine the most suitable regression model, an analysis of the dispersion was conducted. This analysis indicated that there was a presence of under dispersion. Consequently, a quasi-Poisson regression model was selected, as quasi-Poisson regression models have great power to determine significance of covariates in states of under dispersion (Harris et al., 2012). For the interpretation of the estimates derived from the quasi-Poisson regression models, it was required to take the exponential estimated values. This allows for an accurate interpretation of the actual effect of the predictors on the dependent variable.

For the regression analyses a confidence interval had to be determined. This was important as this range represents the probability that the estimate will fall between the range of this interval for a certain percentage of times (Scribbr, 2023). It was decided that, for this research, the confidence interval of 90% would be used for all the regression analyses. This confidence interval was chosen, because the sample size for the analyses was very large. Large sample sizes give narrower intervals, which enables the use of larger confidence levels. This allows for more efficient detection of effects between variables (Statsig, 2025). Furthermore, to assess the proportion of variance in the dependent variable explained by the independent variables of the regression analyses, the adjusted R^2 was taken for the OLS models and for the quasi-Poisson model the pseudo R^2. It is common to observe low R^2 and pseudo R^2 values in social sciences, which would indicate poor model fit. However, this is not necessarily bad as the primary goal is not to predict behaviour, but rather to analyse whether predictors have a significant effect on the dependent variable (Ozili, 2023). Therefore, even if these values turn out to be low, they can still be accepted.

To assess the strengths of the regression results, robustness checks were conducted for all models through the alteration of the regression specifications. This was done by performing the regressions using a subset of the data, which only included video games that had at least 100 user ratings. This approach allowed to observe whether the estimated coefficients were consistent under different conditions. This threshold of user ratings was taken to preserve a large proportion of the VR video games, as their representation would diminish quickly with a higher amount of minimum user ratings. This would not have given a valid illustration on how VR technology is influencing categorical membership and user evaluation.

To provide more contextual information prior to the regression analyses and, consequently, the robustness checks, descriptive statistics are shown for the complete dataset, the dataset for the robustness checks and also for the proportion of VR video games within these datasets. In addition, correlation tests have been conducted for both datasets to assess the presence of correlations between independent variables that were not expected.

Following the robustness checks, the models were assessed for multicollinearity through the use of the Variance Inflation Factor, also known as the VIF test. VIF values between 1 and 5 are generally considered acceptable and indicate low to moderate multicollinearity and suggest that the independent variables are not correlated. VIF values exceeding 5 indicate multicollinearity, which make statistical references less reliable. Therefore, it is important to maintain VIF values between 1 and 5, closer to 1 is preferred, to ensure validity of the results.

3.1.5 Hypotheses equations

The relationships between the different variables have been calculated through the following equations:

Hypothesis 1: $\beta_1 < 0$, User rating = β_0 + β_1(Genre Dummy) + β_2(Multiplayer) + β_3(Release Date) + β_4(Age Rating) + β_5(Supported Languages) + β_6(Estimated Owners) + error

Hypothesis 2: $\beta_1 < 0$, User rating = β_0 + β_1(Genre Dummy)*β_2(AAA Developers) + β_3(Multiplayer) + β_4(Release Date) + β_5(Age Rating) + β_6(Supported Languages) + β_7(Estimated Owners) + error

Hypothesis 3: $\beta_1 < 0$, User rating = β_0 + β_1(VR) + β_2(Genre Dummy) + β_3(Multiplayer) + β_4(Release Date) + β_5(Age Rating) + β_6(Supported Languages) + β_7(Estimated Owners) + error

Hypothesis 4: $\beta_1 > 0$, Genres = β_0 + β_1(VR) + β_2(Multiplayer) + β_3(Release Date) + β_4(Age Rating) + β_5(Supported Languages) + β_6(Estimated Owners) + error

Hypothesis 5: $\beta_1 < 0$, Genre combination commonness = β_0 + β_1(VR) + β_2(Multiplayer) + β_3(Release Date) + β_4(Age Rating) + β_5(Supported Languages) + β_6(Estimated Owners) + error

Hypothesis 6: $\beta_1 < 0$, User ratings = β_0 + β_1(VR)*β_2(Genre Dummy)*β_3(AAA Developers) + β_4(Multiplayer) + β_5(Release Date) + β_6(Age Rating) + β_7(Supported Languages) + β_8(Estimated Owners) + error

3.2 Qualitative analysis

The quantitative approach shows the relationship between categorical membership, user evaluation and VR technology, however the underlying motives from users are not highlighted in these figures. The reasoning behind, for instance, lower ratings are not directly reflected in the results of the quantitative analysis. However, understanding the reasoning is of great importance for comprehending what factors have influenced the outcomes. Therefore, a qualitative approach has been added, in order to gain deeper insights into the results from the quantitative analysis. Developers may be able to reflect those insights with their current or future VR video games to understand the sources of innovation penalty.

3.2.1 Research design

The qualitative analysis has utilised an inductive netnographic research design. According to Kozinets (2010), netnography is a form of ethnographic research that uses the participant-observational approach in online interactions. This approach was chosen to examine the factors contributing to innovation penalties associated with VR video games. Therefore, a content analysis was conducted on user reviews of the two lowest rated genres, which are affected the most by this phenomenon.

This provides insights into the specific bottlenecks associated with lower rated genres in VR video games. These insights can be used by VR video game developers on how to better align their video games with user expectations or by identifying what types of genres are less suitable for VR compatibilities. Themes regarding reasoning behind innovation penalty were not predetermined, but were rather observed by reading the reviews. Consequently, this research was of inductive nature.

3.2.2 Data collection

The data for this analysis was collected using similar methods to those used in the quantitative analysis. Specifically, the data was collected from a secondary source, Kaggle, which has sourced the data directly from Steam through the use of API's. The original dataset that has been taken from this source contained approximately 94 million user reviews. This has been systematically reduced through several filtering methods.

First, only user reviews belonging to video games with "VR" in their title were allowed in the dataset. This criterion was applied to differentiate video games that had access to VR functionalities from video games that did not, as the dataset did not explicitly contain any information that included this. Manually verifying whether video games were compatible with VR functionalities was not feasible due to time constraints, therefore this method was used as an alternative. Second, the remaining reviews were filtered based on the language of the review. Only reviews written in English were preserved in the dataset, all other reviews written in a different language were omitted from the dataset, as these reviews could not be interpreted.

Third, a minimum threshold of ten words per review was applied. This was done to enhance the consistency and reliability of the data, because this filtering process helped with the removal of reviews that would not contribute to the content analysis, as they would not provide any meaningful information. Fourth, only negative reviews of the two lowest rated genres were included in the datasets; sports and adventure. These genres were determined through their proportions of negative reviews. A percentage was calculated of negative user reviews for all 11 genres that were portrayed in the datasets.

Lastly, the provisional dataset was cross referenced with the quantitative dataset to ensure that all the video games, that were included in the qualitative data, were also present in the quantitative data. Only video games that were present in the quantitative dataset were retained in the final sample for the qualitative dataset. This resulted in the facilitation of the most comprehensive conclusions, which are in line with the results from the quantitative analysis.

This filtering process had resulted in a data sample with a total of almost 350,000 reviews. From this 350,000 reviews, a subset of reviews have been taken through purposive sampling (Researcher, 2024). In purposive sampling, samples are chosen based on their characteristics, which fit the study's objective. The reviews of the genres were chosen based on their proportion of negative sentiments, which resulted in two definite datasets with approximately 25,000 reviews for both; one with reviews from VR sports video games and the other with VR adventure video games.

The data in this dataset is also publicly accessible and therefore did not pose any legal issues concerning Steam's privacy policies. However, there were some ethical concerns regarding the use of netnography in research (Anne-Marie et al, 2017). Therefore, several precautions were implemented to ensure ethical integrity in the analysis of the user reviews. Firstly, all reviews were used anonymously so users can not be identified or held accountable for their online statements. However, anonymity can not be fully guaranteed through omission of usernames alone, as direct quotes may still be traceable via search engines, which could lead to the identification of their username. To address this, all reviews were paraphrased, while preserving the original meaning and intent of the users. This ensured the user's anonymity, while keeping the core of their review intact so that it maintains its informational value for the analysis.

3.2.3 Operationalisation

The aim of the analysis of the reviews was to identify recurring themes that would give contextual meaning of innovation penalty within VR video games. Therefore, the reviews were meticulously read and interpreted, through an inductive thematic analysis, in order to identify themes that would explain this phenomenon. As aforementioned, the datasets were extensive, hence not all reviews could be analysed. Accordingly, random samples of the reviews were taken of which no more than five reviews per video game were allowed. This method was applied, because there was overrepresentation from certain video games, which would have a high proportion of reviews. This avoided that all analysed reviews would stem from these video games, which could have resulted in an inaccurate depiction of recurring themes.

The random samples contained 50 reviews per sample, which have been analysed till thematic saturation was realised. According to Lowe et al. (2018), thematic saturation is achieved when further observations do not provide any new themes. This research determined that thematic saturation was realised when the following 50 reviews did not contain any new themes. This was assessed after 200 reviews for sports and 150 for adventure. The identified themes were then elaborated with quotes to provide more context, which resulted in recommendations for developers.

3.3 Research quality

Certain methodological choices have been made in order to ensure credibility and rigor of the research.

Reliability is supported through various methods, one of which are the robustness checks. As aforementioned, for the quantitative analyses robustness checks have been performed to assess, whether the results are consistent across different samples. This ensures that the results are reproducible and therefore increases reliability of the research. Furthermore, reliability is supported by the use of publicly available data from Steam. The data is obtained through secondary sources, but they have directly sourced the data from Steam. This data provides consistent and high-quality data, which guarantees reproducibility.

This is also ensured within the qualitative analysis, where the reviews are sourced from a secondary source, which collected it directly from Steam. This data is also reliable and will ensure reproducibility. Besides, the results that stemmed from this data have been categorised carefully with corresponding quotes which can all be traced back, and therefore contributing to an increase in reliability.

To ensure validity, control variables have been added into the quantitative analyses. The inclusion of these variables makes sure that the influence of confounding variables is limited, thereby helping to establish a causal relationship between the dependent and independent variables (Scribbr, 2023). Furthermore, the research is representative due to the extensive nature of the datasets. The generalisability of the findings is enhanced due to the inclusion of a large and diverse sample of video games.

With regard to the qualitative analysis, the filtering method used for the final sample of reviews made sure that validity of the research is also intact. Due to this method, overrepresentation and irrelevant reviews were omitted, which could have given a distorted view, as sentiments would only be applicable to the respective video games and not the genres in general. Therefore, this ensured that the qualitative analysis really measured what it was intended to measure. In addition, the results from the qualitative analysis were shown after thematic saturation was determined. Thus, it was made sure that no additional themes were omitted by accident, hence it increases validity of the research.

However, this research does face certain limitations that are difficult to mitigate. One such limitation involves the user ratings in the datasets, which are vulnerable to manipulation. For instance, stakeholders that are involved in the development may influence the ratings through an inflow of positive ratings to simulate a more successful video game. This could result in a biased representation of user evaluations. This was addressed through the inclusion of a qualitative analysis. Additionally, the coding of the qualitative data has been done by a single researcher. Therefore, others might have interpreted the content of the reviews differently, which would have resulted in different themes. While efforts were made to remain objective and consistent, the reliability of the themes could not be checked with other researchers. So, the thematic determinations could possess a small sample of subjectivity.

4. Quantitative results

The results are presented in two distinct sections. The first section discusses the results based on the complete dataset, which constitutes of 45,244 observations. The second section presents a robustness check including only video games that had at least 100 different user ratings to assess the validity and generalisability of the initial results, which constitutes of 18,077 observations.

4.1 Descriptive statistics

Several variables have been utilised in the analysis of the complete dataset and their descriptive statistics are presented in table 1.

User ratings range from 0 to 100% with an average rating of 76.86% and a median of 81%. Regarding genre classification, video games are, on average, classified between two and three genres, as indicated by the mean of 2.87 and a median of 3. This variable was also transformed into a dummy variable, when it functioned as an independent variable in the regressions. This shows that approximately 85% of the video games are classified under more than one genre, with a mean of 0.85. The commonness of genre combinations range from 1 to 2,496, with a mean of 853.35 and a median of 551, which indicates that the average genre combination is present in 853 video games. Therefore, higher values of commonness of genre combinations indicate that it is more common to observe such a combination across different video games.

Whether video games have access to VR functionalities has been observed in 7% of the video games, as the mean is 0.07. This is a very small portion of the video games, which indicates that VR is still an emerging technology within the industry and has not reached maturity yet. Comparably, only around 2% of the video games were developed by incumbent developers, which indicates that novel developers release much more video games, controlling the industry. A similar proportion of video games have an age rating of 16+ as also approximately 2% of video games have this rating.

Multiplayer video games also seem to be less prevalent when compared to single player video games, as around 21% of the video games have access to multiplayer functionalities. The estimated number of people that own a specific video game varies considerably between 10,000 and 350,000,000 owners, with the average video game having a total of 191,376 different owners and a median of 10,000. The variable is utilised by the million which makes interpretation of its effect easier, given the broad range of the estimated owners. The video games have been released between 1997 and 2024, which gives a range of 27 years of which the most video games have been published during 2019. Ultimately, video games utilise different languages into their video games, ranging from 1 to 103 languages, with an average of 5.17 different languages supported by a video game.

These descriptive statistics summarise and organise the characteristics of the dataset and provide important context for the interpretation of the results in the following sections.

Variable	Min	Max	Mean	Median	Std_Dev
Age Rating (16+)	0	1	0.02	0	0.15
AAA Developers	0	1	0.02	0	0.13
Multiplayer	0	1	0.21	0	0.41
Estimated Owners	10,000	350,000,000	191,376.87	10,000	2,350,293.85
VR	0	1	0.07	0	0.25
User Rating	0	100	76.86	81	17.74
Supported Languages	1	103	5.17	2	10.41
Genre Commonness	1	2,496	853.35	551	842.26
Genres	1	11	2.87	3	1.35
Genre Dummy	0	1	0.85	1	0.35
Release Date	1997	2024	2019	2020	3.26

Table 1: Descriptive statistics of the complete dataset

4.2 Descriptive statistics VR video games

For comparison purposes, table 2 represents the descriptive statistics for only the VR video games that were present in the complete dataset.

Variable	Min	Max	Mean	Median	Std_Dev
Age Rating (16+)	0	1	0.01	0	0.11
AAA Developers	0	1	0.01	0	0.10
Multiplayer	0	1	0.23	0	0.42
Estimated Owners	10,000	350,000,000	111,246.25	10,000	850,503.00
VR	1	1	1	1	0
User Rating	0	100	73.58	77	17.76
Supported Languages	1	103	3.36	1	5.39
Genre Commonness	1	2,496	634.81	198.00	797.77
Genres	1	11	2.98	3	1.39
Genre Dummy	0	1	0.87	1	0.34
Release Date	2008	2024	2018	2018	2.35

Table 2: Descriptive statistics of the VR video games in the complete dataset

4.3 Correlation test

Prior to constructing a correlation matrix, several assumptions regarding data distribution had to be tested in order to determine, which correlation test was suited the best to perform. The distribution of the data was assessed through the use of the Kolmogorov-Smirnov test as well as through the analysis of histograms and Q-Q plots. The results of these analyses indicated that the variables were non-Gaussian distributed, as $p < 0.05$. Therefore, the Spearman rank correlation method was observed as fitting the best, as it is appropriate for non-Gaussian data (Spearman, 1961). The correlation matrix that followed from this test is presented in table 3.

The coefficients of Spearman's test are interpreted based on their proximity to zero; values between ±0.00 and ±0.39 are considered very weak to weak correlation, which indicate little to no negative or positive monotonic correlation (Laerd, n.d; Statstutor, n.d.). The correlation matrix shows that the vast majority of the variable coefficients fall well within this range, which indicates that there is little to no correlation between the variables.

However, there are two relationships that approach the upper boundary of this range, which are the relationships between age rating and incumbent developers and estimated owners and release date with coefficients of 0.3 and -0.33 respectively. Even though, they are approaching moderate correlation, the values still remain within the range of weak correlation and therefore are not considered problematic.

On the other hand, there are two relationships that do exceed this range and even are considered strongly correlation relationships. These relationships are that of genres and genre dummy and genre commonness and genres. The relationship between genres and genre dummy has a coefficient of 0.63 which suggests a strong positive relationship, while the relationship between genres and genre commonness has a coefficient of -0.55 which suggests a strong negative relationship. Nonetheless, these results are theoretically expected, as the variable genre dummy is derived from the variable genres, as it is a dummy version of it. This is also the case for genre commonness, as the value for commonness is derived from the genres that are used. Therefore, the observed strong correlation effects do not pose concerns for the regression analyses that follow.

	AAA_Developers	Multiplayer	Estimated_Owners	VR	User_Rating	Supported_Languages	Genre_Commonness	Genres	Genre_Dummy	Release_Date
Genre_Dummy										0.06
Genres									0.63	0.1
Genre_Commonness								-0.55	-0.18	-0.05
Supported_Languages							0.04	-0.02	0	0.1
User_Rating						0.02	0.15	-0.1	-0.03	0.21
VR					-0.06	-0.08	-0.09	0.02	0.01	-0.08
Estimated_Owners				-0.05	0.02	0.2	-0.06	-0.02	-0.03	-0.33
Multiplayer			0.19	0.02	-0.08	0.09	-0.18	0.07	0	-0.09
AAA_Developers		0.11	0.19	-0.01	0.02	0.13	0.01	-0.13	-0.13	-0.09
Age_Rating	0.3	0.04	0.15	-0.02	-0.01	0.08	0.03	-0.09	-0.09	-0.09

Table 3: Correlation matrix of the complete dataset

4.4 User rating

The regression results in table 4 illustrate how user ratings are influenced by the independent variables. These results show the influence across models 1, 2, 3 and 6, which correspond to hypotheses 1, 2, 3 and 6 respectively. Approximately 0.032-0.036 of the variation in the dependent variable is explained across all of the models.

Firstly, it can be stated that almost all independent and control variables are statistically significant across all models with a CI of 90%. Among the most notable variables are the genre classification and multiplayer functionalities. Regarding genre classification, which specifies whether a video game is classified under more than one genre, consistently shows a strong and statistically significant negative effect on user ratings. This is most evident in model 1, where the coefficient is -1.915, which indicates that if a video game is classified under more than one genre, the user rating is, on average, 1.915 points lower than that of single genre video games.

When the models get more complex, this strong and statistically significant negative effect remains, but declines in effect size. The effect sizes now are -1.597, -1.887 and -1.628 respectively for models 2, 3 and 6. This reduction in effect size in model 2 suggests that some of the penalty that multi-genre video games experience are explained by whether the video games was developed by either incumbent or novel developers. However, in models 3 and 6 the addition of VR functionalities slightly increases the effect size again, which suggests that VR functionalities may have positively influenced the decline in user ratings.

Regarding the control variables, multiplayer functionalities display a strong and statistically significant negative effect on user ratings. Model 1 shows a coefficient of -2.265, which indicates that video games that have access to multiplayer functionalities, on average, are rated 2.265 points lower than video games that have only access to single player modes. This strong effect is maintained across the different models and is even slightly amplified in the subsequent models. The coefficient for models 2, 3 and 6 are -2.409, -2.245 and -2.392, respectively, which are both marginally lower than in model 1.

This increase in effect size after the introduction of the type of developer, suggests that some of the penalty that multiplayer video games experience are explained by the type of developer that developed the video game. The minimal change in effect size between models 2 and 6, with the introduction of VR functionalities, suggest that VR functionalities have a small positive influence on the decline in user rating, as the effect size became smaller.

While these two variables showed that they had negative effect on user ratings, there are also variables that are perceived as positive on user rating. Release date is one of those variables, as it shows a statistically significant positive effect on user ratings with a coefficient of 0.888 in model 1. In models 2 and 6, this is effect is intensified again, with coefficients of 0.910 and 0.902 respectively, while model 3 shows a similar effect size of 0.878. These results suggest that more recently released video games, on average, receive slightly higher user ratings.

Supported languages also show a statistically significant positive effect on user ratings across all models. This indicates that video games that support a broader range of languages are rated averagely higher. The coefficient of model 1 is 0.031 and the effect sizes appear to diminish with the introduction of the additional variables in models 2, 3 and 6, with coefficient of 0.029, 0.029 and 0.026 respectively. Even though, the effect sizes decrease they are not of a considerable magnitude and therefore have not a big significant effect on user ratings.

The estimated owners of a video game has also a consistent positive and statistically significant effect on user ratings across all models, however similarly to supported languages, the effect size of this variable appears small. The effect sizes are, respectively, 0.304, 0.269, 0.301 and 0.266 for models 1, 2, 3 and 6. The results are all statistically significant and positive, thus it indicates that games that are owned by more users tend to receive, on average, higher user ratings even if it is marginally.

On the other hand, age rating does not show any statistically significant effect on user ratings across all models. This suggest that there is not any concluding evidence on whether a game is rated 16+ has any influence on how users would evaluate the video games.

As aforementioned, in models 2 and 6 another variable was added, which is whether a video game is developed by an incumbent or novel developer, i.e. AAA developers. This variable shows a considerable and statistically significant positive effect on user ratings for both model 2 and 6. With coefficient of 7.090 and 7.366 respectively for models 2 and 6, it suggests that video games that are developed by incumbent developers, averagely receive a user rating that is 7.090 or 7.366 points higher than video games developed by novel developers. These effect sizes are very large and therefore this variable appears to have the most influence on user evaluation in video games.

In models 3 and 6 another additional variable is included which is the variable of VR functionalities. This variable appears to have a strong and statistically significant negative effect on user ratings, with a coefficient of -2.809 and -3.521 for models 3 and 6. This suggests that video games that have access to VR functionalities are on average lower rated than video games that do not have access to these functionalities.

In model 2, an interaction term between genre classification and type of developers behind the video game has been included to test hypothesis 2. The coefficient for this interaction term is -3.078. In order to interpret the actual influence of this interaction, the coefficients of the individual variables; genre classification (-1.597), type of developer (7.090) and their interaction have to be taken together. The resulting coefficient is 2.415, which means that the negative effect of multi genre classification diminishes and the positive effect of being developed by an incumbent is partially nullified when both variables are found in video games. However, the effect remains positive, so video games that are both developed by an incumbent developer and are classified under multiple genres are expected to receive a user ratings that is averagely 2.415 points higher than those developed by novel developers and are classified under a single genre.

Model 6 extends on this interaction term by incorporating three additional interaction terms: (1) VR functionalities and developer type; (2) VR functionalities and genre classification; (3) VR functionalities, genre classification and developer type. Table 4 shows that the same interaction term as in model 2 has a similar effect in model 6, as it has a strong and statistically significant negative coefficient of -3.409. When the coefficients of the individual variables and the interaction term are combined, the actual predicted effect is 2.329. This suggests that even though the negative effect of multi genre classification and the positive effect of incumbent developers are both reduced, the overall effect remains positive with an estimated user rating increase of, on average, 2.329 points.

The interaction between VR functionalities and developer type shows an even stronger effect, as the coefficient for this interaction is -8.533. The combined individual effects give an actual effect of-4.688, which has a strong and statistically significant negative effect on user ratings. This implies that while video games developed by incumbents typically receive higher user ratings (7.366), this effect is lowered when the video games includes VR functionalities (-3.521). Therefore, video games that are both developed by incumbents and have access to VR functionalities, it is expected that these video games averagely receive 4.688 points lower than video games that do not.

The remaining two interaction terms are that of VR functionalities and genre classification and VR functionalities, genre classification and developer type. These interactions have a coefficient of 0.910 and 8.972 respectively. The actual effects for these interactions are -4.239 for multi genre video games with VR functionalities and -0.843 for multi genre video games with VR functionalities developed by incumbents. However, neither of these coefficients is statistically significant, which indicates that there is no concluding evidence that these interaction effects have meaningful influence on user ratings. As such, interpretation of these interactions should be made carefully and no definitive claims about their impact should be drawn.

In conclusion, the results presented in table 4 show that the hypotheses 1, 2 and 3 should be accepted. In the case of hypothesis 6, the results are also consistent with the hypothesis, however the lack of statistical significance results in the rejection of the hypothesis.

Ordinary Least Square	**User Rating**			
	(1)	**(2)**	**(3)**	**(6)**
	Coef.	**Coef.**	**Coef.**	**Coef.**
Variables	**(se.)**	**(se.)**	**(se.)**	**(se.)**
Genre Dummy	-1.915***	-1.597***	-1.887***	-1.628***
	(0.234)	(0.239)	(0.234)	(0.247)
Multiplayer	-2.265***	-2.409***	-2.245***	-2.392***
	(0.203)	(0.204)	(0.203)	(0.203)
Release Date	0.888***	0.910***	0.878***	0.902***
	(0.026)	(0.026)	(0.026)	(0.026)
Age Rating	0.920	-0.411	0.828	-0.479
	(0.568)	(0.591)	(0.568)	(0.590)
Supported Languages	0.031***	0.029***	0.029***	0.026***
	(0.008)	(0.008)	(0.008)	(0.008)
Estimated Owners	0.304***	0.269***	0.301***	0.266 ***
	(0.035)	(0.036)	(0.035)	(0.036)
AAA Developers	-	7.090***	-	7.366***
	-	(0.916)	-	(0.934)
VR	-	-	-2.809***	-3.521***
	-	-	(0.330)	(0.925)
Genre Dummy:AAA Developers	-	-3.078**	-	-3.409***
	-	(1.239)	-	(1.267)
VR:AAA Developers	-	-	-	-8.533*
	-	-	-	(4.415)
VR:Genre Dummy	-	-	-	0.910
	-	-	-	(0.990)
VR:Genre Dummy:AAA Developers	-	-	-	8.972
	-	-	-	(6.030)
Constant	-1,713.659***	-1,759.040***	-1,693.567***	-1,741.804***
	(51.883)	(52.120)	(51.896)	(52.153)
Observations	45,244	45,244	45,244	45,244
R^2	0.032	0.034	0.034	0.036
Note:	*p<0.1; **p<0.05; ***p<0.01			

Table 4: Regression estimates for user ratings

4.5 Genres

Table 5 presents the regression results for the fourth model, which analyses the influence of the independent variables on genre classification. The coefficient presented in this table were derived from a quasi-Poisson regression model, this means that the coefficients do not represent the actual effects. Therefore, for the interpretation of their actual effects, the exponential of all coefficients have been calculated. The interpretation of these values will be discussed for each variable individually.

All the independent variables in the model are statistically significant, except for the variable of estimated owners. It is also interesting to note that all independent variables have almost the same effect size and are all positive. Approximately 0.034 of the variation in the dependent variable is explained within this model.

The variable of VR functionalities shows a positive and statistically significant coefficient of 1.05. Therefore, video games with VR functionalities are, on average, classified under 1.05 more genres than video games without VR functionalities.

Similarly, multiplayer functionalities also has a positive and statistically significant effect on genre classification. Video games that have access to multiplayer functions are averagely classified under 1.11 more genres than video games that only have access to singleplayer modes.

The release date of a video game is also positively and statistically significantly associated with its genre classification. More recently released video games classify a higher amount of genres than those that were released earlier. For every additional year, that a video game is released about 1.02 more genres are classified under that video game.

Two additional variables, supported languages and estimated owners, also show positive. Specifically, each additional language that is supported within a video game results in a statistically significant increase in genre classifications of averagely one. Highly diffused video games are also approximately classified under one more genre, however the results are statistically insignificant.

Age rating is the variable that has the lowest effect size in this model with a coefficient of 0.762, which indicates that video games with an age restriction of 16+ are associated with more genre classifications, as they are, on average, classified under 0.762 more genres than those that do not have this age restriction.

In conclusion, the results presented in table 5 indicate that hypothesis 4 should be accepted.

Quasi-Poisson	Genres
	(4)
	Coef.
Variables	(se.)
VR	0.049***
	(0.009)
Multiplayer	0.106***
	(0.005)
Release Date	0.020***
	(0.001)
Age Rating	-0.272***
	(0.017)
Supported Languages	0.001***
	(0.0002)
Estimated Owners	$9.14*10^{e-10***}$
	$(8.57*10e^{-10})$
Constant	-39.016***
	(1.401)
Observations	45,244
Pseudo R^2	0.034

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5: Regression estimates for genres

4.6 Genre commonness

Table 6 presents the regression results for model 5, which analyses the influence of the independent variables on the commonness of genre combinations in video games. This model explains approximately 0.03 of the variation in the dependent variable. It is important to note, that the genre commonness variable ranges from 1 to 2,496. This indicates that a value of 1 corresponds to a genre combination that appears in only one video game in the dataset, therefore is highly uncommon, while a value of 2,496 indicates that the combination is found in 2,496 video games and therefore is highly common.

The accessibility to VR functionalities has a statistically significant negative effect on genre commonness, with a coefficient of -233.446. This indicates that, on average, video games that have access to VR functionalities are associated with 233 fewer video games, than non-VR video games that share the same genre combinations.

Similarly, multiplayer functionalities exhibit a statistically significant negative effect on genre commonness, with a coefficient of-295.477. This effect size suggests that video games that incorporate multiplayer elements into their game, averagely are shared with 295 fewer games, opposed to single player video games that share the same genre combinations.

The release date of video games also shows a statistically significant negative effect on genre commonness, but its effect size happens to be of smaller magnitude. The coefficient regarding release date is -17.139, therefore this indicates that for each additional year the genre combination will see 17 less video games.

The diffusion of video games also has a statistically significant negative effect on genre commonness. Although significant, the effect size is extremely small, which indicates that its effect is almost negligible. Video games that have a higher amounts of owners, averagely are associated with 6.57 less video games than those with lower amounts of owners.

Conversely, supported languages has a statistically significant positive effect on genre commonness, with a coefficient of 1.926. This indicates that every additional language that is supported in a video game results in approximately two extra games associated with the same genre combination.

Finally, the age rating variable shows to have a positive coefficient of 40.098, which would indicate that video games that have an age restriction of 16+ have about 40 more video games associated with the same genre combinations, than those without age restriction. However, the variable is statistically insignificant and therefore suggests that no reliable conclusion can be drawn, regarding it impact on the commonness of genre combinations that are found in these video games.

In conclusion, the results presented in table 6 indicate that hypothesis 5 should be accepted.

Ordinary Least Square	**Genre commonness**
	(5)
	Coef.
Variables	**(se.)**
VR	-233.446***
	(15.730)
Multiplayer	-295.477***
	(9.657)
Release Date	-17.139***
	(1.218)
Age Rating	40.098
	(26.967)
Supported Languages	1.926***
	(0.377)
Estimated Owners	-6.57***
	(1.68)
Constant	35,528,390***
	(2,459,308)
Observations	45,244
R^2	0.03

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 6: Regression estimates for genre commonness

4.7 Multicollinearity

The regression models are all assessed for multicollinearity through the use of the VIF test. The results presented below concern all the hypotheses and to increase clarity and readability the VIF results have been put in a range across all models. The results show that all VIF values were below the threshold of 5 and are all even close to 1, which indicates that there is no multicollinearity throughout the different models.

Variable	VIF range	Outcome
Genre Dummy	1.01-1.03	No multicollinearity
AAA Developers	1.01-1.03	No multicollinearity
Multiplayer	1.01-1.02	No multicollinearity
Release Date	1.03-1.04	No multicollinearity
Age Rating	1.02-1.06	No multicollinearity
Supported Languages	1.00-1.01	No multicollinearity
Estimated Owners	1.01-1.02	No multicollinearity
VR	1.00-1.01	No multicollinearity

Table 7: VIF results complete dataset

5. Robustness check

A robustness check was conducted to enhance the validity and generalisability of the regression analysis. This check assesses the sensitivity of the results to changes in the dataset, which can increase the confidence in the findings. Therefore, a subset of the original dataset was constructed, which included only video game with at least 100 user ratings under the assumption that a higher number of ratings ensures more reliable outcomes.

5.1 Descriptive statistics

The descriptive statistics of the filtered dataset are presented in table 8.

The user ratings in this dataset show an average rating of 80.06% and a median of 83%, which closely aligns with the full dataset. Similarly, video games are on average classified between two and three genres. This is indicated by the mean of 2.84 and median of 3. The dummy version for genre classification, which represents whether a game is classified under more than one genre, has a mean of 0.84, indicating that 84% of the video games is classified under more than one genre. The commonness of genre combinations still ranges from 1 to 2,496 in this dataset and has a mean of 777.06 and median of 478, which indicates that the average genre combination can be found in approximately 777 video games.

There are some shifts that have been noticed in the filtered dataset:

- VR functionalities are present in 5% of the video games, which was 7%.
- The incumbent developers now account for 4% of the video games, which was 2%.
- Video games that have an age rating of 16+ has increased from 2% to 5%
- Multiplayer functionalities can now be found in 28% of the video games, which was 21%.

This suggests that video games that have more than 100 user ratings are more likely to be developed by incumbent developers, target mature audiences and support multiplayer. On the other hand, they are less likely to incorporate VR functionalities into their video games.

The extent of estimated number of owners remains between 10,000 and 350,000,000 owners, but the mean regarding this variable has shifted considerably. The average video game now has a total of 447,539 different owners and a median of 75,000.

The ranges for the supported languages and release date are also maintained as in the full dataset. However, the average number of languages that are supported in video games have risen to 6.18 with a median of 4, which is a one language increase. And for the release date the year that the most games have been released has shifted from 2019 to 2018.

As mentioned, there are differences in the descriptives for the dataset of the robustness check with regards to the full dataset, however the changes are relatively minimal. The descriptives of the robustness check dataset is closely similar to that of the full dataset. This confirms it suitability for further analysis and supports the validity of the results.

Variable	Min	Max	Mean	Median	Std_Dev
Age Rating (16+)	0	1	0.05	0	0.21
AAA Developers	0	1	0.04	0	0.20
Multiplayer	0	1	0.28	0	0.45
Estimated Owners	10,000	350,000,000	447,539.97	75,000	3,703,021.10
VR	0	1	0.05	0	0.21
User Rating	7	100	80.06	83	14.07
Supported Languages	1	103	6.18	4	8.82
Genre Commonness	1	2,496	777.06	478	807.29
Genres	1	11	2.84	3	1.35
Genre Dummy	0	1	0.84	1	0.36
Release Date	1997	2024	2018	2019	3.72

Table 8: Descriptive statistics robustness check dataset

5.2 Descriptive statistics of VR video games in the robustness checks

For comparison purposes, table 9 present the descriptive statistics for VR video games in the robustness check dataset. The descriptive statistics are closely in line with those in the complete dataset.

Variable	Min	Max	Mean	Median	Std_Dev
Age Rating (16+)	0	1	0.03	0	0.16
AAA Developers	0	1	0.03	0	0.18
Multiplayer	0	1	0.38	0	0.49
Estimated Owners	10,000	350,000,000	340,394.89	75,000	1,561,973.77
VR	1	1	1	1	0
User Rating	25	99	79.89	82	12.65
Supported Languages	1	28	4.77	2	4.99
Genre Commonness	1	2496	673.20	238	821.19
Genres	1	11	2.92	3	1.35
Genre Dummy	0	1	0.87	1	0.33
Release Date	2008	2024	2018	2018	2.60

Table 9: Descriptive statistics of the VR video games in the robustness check dataset

5.3 Correlation test

For the correlation test of the dataset for the robustness check the same assumptions were tested and the same outcomes were observed. Therefore, the Spearman rank correlation method was used, which is presented in table 10.

The correlation matrix shows that the vast majority of the variable coefficients fall within the range of very weak to weak, which indicates that there is little to no correlation between the variables.

The same two relationships as in the full dataset approach the upper boundary of this range, which are the relationships between age rating and incumbent developers and estimated owners and release date with coefficients of 0.33 and -0.36 respectively. the same is considered for this dataset, as the coefficient are similar; the values remain within the range of weak correlation and therefore are not considered problematic.

The relationships that exceed the range and are considered strongly correlated are also the same; genres and genre dummy and genre commonness and genres. Their coefficients are 0.65 and -0.55 respectively. However, these results were again expected so the observed strong correlating effects do not pose concerns for the regression analyses.

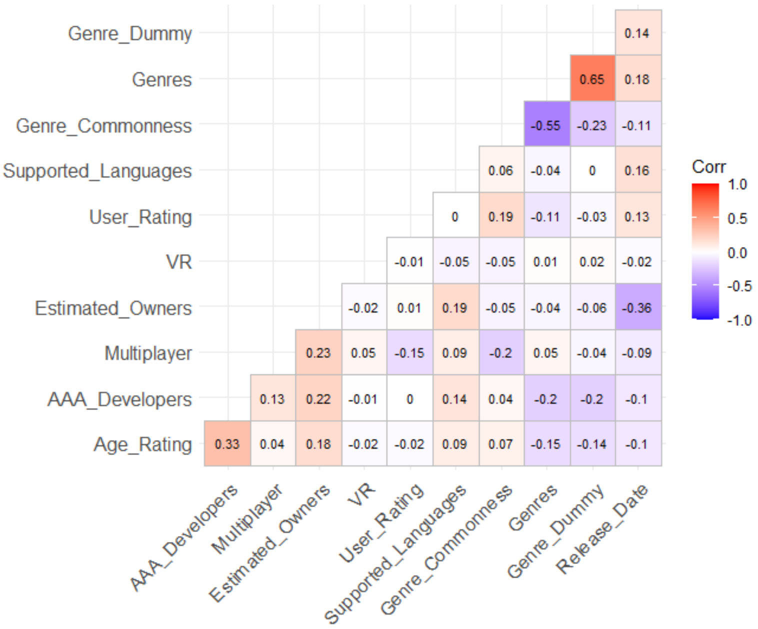

Table 10: Correlation matrix robustness checks

5.4 User rating

The regression results in table 11 provide insights into how user ratings are influenced by the independent variables under the robustness check conditions. This robustness check includes only video games that have a minimum of 100 user ratings.

The effect of genre classification on user ratings remains consistent with the results from the full dataset. The variation that is explained of the dependent variable remains around 0.036. Additionally, it displays a strong and statistically significant negative influence on user ratings. In model 1, the coefficient for genre classification is -1.735. which indicates that video games classified under multiple genres with more user ratings are rated, on average, 1.735 points less. As the models become more complex, with the addition of variables, this negative effect slightly diminishes with coefficients of -1.519 in model 2 and -1.473 in model 6, while it increases slightly in model 3 with a coefficient of -1.739. This increase is minimal, so the results are in line in the full dataset.

The effect of multiplayer functionalities remains the same as in the full dataset, where it maintains a strong and statistically significant negative effect on user ratings. The coefficient is -4.285, which significantly larger than the -2.265, which was observed in the full dataset. This suggests that video games with more user ratings experience stronger penalty for including multiplayer functionalities. This negative effect further intensifies in the other models, where coefficients increased from -2.409, -2.245 and -2.392 for models 2, 3 and 6 respectively to -4.346, -4.294 and -4.362.

Contrary to the full dataset, where age rating was statistically insignificant across all models, it is now statistically significant in the robustness check. The effect of age rating is consistently negative across all models, with a coefficient of -0.979 in model 1. This effect intensifies in the following models, with coefficients of -1.393 for model 2 and
-1.350 for model 6, while it decreases slightly with -0.973 for model 3. This indicates that video games with an age restriction of 16+ are associated with lower user ratings, especially in video games that have more user ratings.

Release date maintains a statistically significant positive effect on user ratings across all models, which is still consistent with the full dataset. Although, the effect sizes have changed a little bit, as these have become slightly lower in the robustness check. In model 1, the variable of release date exhibits a coefficient of 0.394. This effect is intensified or remains the same in the subsequent models with a coefficient of 0.403 in model 2, 0.394 in model 3 and 0.404 in model 6. This indicates the same effect as in the full dataset, where more recently released video games tend to receive higher user ratings, even though the effect size of this effect is slightly reduced.

The effect of supported languages also remains statistically significant and positive on user ratings across all models. The coefficient of 0.066 in model 1 is nearly doubled with that of the full data set. This suggests that video games with more user ratings are more positively impacted by the amount of languages that are supported in the video game. This effect remains stable in the other models, as these coefficients are 0.064, 0.066 and 0.065 for models 2, 3 and 5 respectively.

The estimated owners of a video game exhibit a positive and statistically significant effect on user ratings across all models, which is in line with the results of the full dataset, with coefficients of 0.179, 0.170, 0.179 and 0.170 for the models 1, 2, 3 and 6 respectively. It is a small decrease with regard to the full dataset, but it still indicates that an increase in the number of owners that a video game has, results in a slightly higher user rating. However, the effect size of this variable suggests that the effect is not that influential.

The influence of the type of developer behind a video game remains positive and statistically significant in both models 2 and 6. However, the effect sizes of the coefficients has decreased significantly in the robustness check for both models. The coefficients for the type of developer were 7.090 and 7.366 for models 2 and 6, considering that the coefficients now are 2.148 and 2.549, respectively. This implies that the benefit that video games experience by being developed by incumbent developers is less pronounced in video games with more user ratings.

VR functionality was statistically significant in the full dataset, but now is statistically insignificant in the robustness check for both models 3 and 6. This change is likely because of the reduction in share of VR video games in the data, which decreased from 5,700 to 871. This decreased sample size could have resulted in an inability to detect an effect on user ratings.

Model 6 had an additional three interaction terms incorporated in the model. The interaction term's coefficient between VR functionalities and developer type has remained statistically significant and negative, however it has increased from -8.533 to -11.222 with regards to the full model. The actual predicted effect that was perceived was -4.688, which had a negative effect. This actual effect has stayed negative, with now a coefficient of -6.904. This suggests that video games with more user ratings, that have access to VR functionalities and developed by incumbents are penalised more, by approximately two points.

The interaction term between genre classification and type of developers had a negative coefficient in the full dataset, which can also be seen in the robustness check. However, the coefficient is statistically insignificant, which was not the case in the full dataset. This could indicate that the interaction term is mostly dependent on less widely-rated video games, where differentiation in genre classification may play a more important role in user ratings.

The remaining two interaction terms, between VR functionalities and genre classification, and VR functionalities, genre classification and developer type, presented similar results in both the full dataset as in the robustness check. The actual predicted effects for both interaction terms were negative and they did not show any statistical significance, which was also captured in the results of the full dataset.

Overall, the robustness check generally confirms the results from the full dataset, but with some difference in the magnitudes of effect sizes and statistical significance. The most prominent changes were those observed for the variable of multiplayer functionalities, which has a much stronger negative effect on user ratings and the variable of age rating, which becomes statistically significant in this check. The variable of VR functionalities showed a statistical insignificance, which is most likely due to the reduced number of VR games in the dataset.

The results are generally consistent with the hypotheses. Model 1 shows a consistent statistically significant result and supports the hypothesis. Similarly, model 6 shows a consistent statistically insignificant result and rejects the hypothesis. However, in the case of model 2 the result is consistent with the hypothesis, but lacks statistical significance. Since the hypothesis was statistically significant in the results from the complete dataset, the hypothesis will still be accepted, however it is only weakly supported due to the statistical insignificance in the robustness check. The same applies to model 3, for which the robustness check results indicate statistical insignificance, therefore the hypothesis is still accepted based on the results from the complete dataset, but is only weakly supported.

Ordinary Least Square	User Rating			
	(1)	(2)	(3)	(6)
	Coef.	Coef.	Coef.	Coef.
Variables	**(se.)**	**(se.)**	**(se.)**	**(se.)**
Genre Dummy	-1.735***	-1.519***	-1.739***	-1.473***
	(0.290)	(0.306)	(0.290)	(0.312)
Multiplayer	-4.285***	-4.346***	-4.294***	-4.362***
	(0.230)	(0.231)	(0.231)	(0.232)
Release Date	0.394***	0.403***	0.394***	0.404***
	(0.029)	(0.029)	(0.029)	(0.029)
Age Rating	-0.979*	-1.393***	-0.973*	-1.350**
	(0.502)	(0.524)	(0.502)	(0.525)
Supported Languages	0.066***	0.064***	0.066***	0.065***
	(0.012)	(0.012)	(0.012)	(0.012)
Estimated Owners	0.179***	0.170***	0.179***	0.170***
	(0.028)	(0.028)	(0.028)	(0.028)
AAA Developers	-	2.148***	-	2.549***
	-	(0.774)	-	(0.787)
VR	-	-	0.376	1.769
	-	-	(0.484)	(1.454)
Genre Dummy:AAA Developers	-	-1.069	-	-1.364
	-	(1.040)	-	(1.060)
VR:AAA Developers	-	-	-	-11.222***
	-	-	-	(4.035)
VR:Genre Dummy	-	-	-	-1.321
	-	-	-	(1.545)
VR:Genre Dummy:AAA Developers	-	-	-	8.565
	-	-	-	(5.386)
Constant	-713.449***	-730.613***	-713.167***	-732.871***
	(57.830)	(58.103)	(57.832)	(58.127)
Observations	18,077	18,077	18,077	18,077
R^2	0.036	0.036	0.036	0.036
Note:	*p<0.1; **p<0.05; ***p<0.01			

Table 11: Regression estimates for user rating in robustness checks

5.5 Genres

Table 12 presents the regression results of the robustness check for the fourth model, which analyses the influence of the independent variables on genre classification. The exponential of all coefficient from the table are again calculated for the interpretation of the actual effects. The findings from this robustness check largely align with those from the full dataset. However, there are two notable differences in the variables of VR functionalities and estimated owners and the explained variation is seen to be doubled to 0.067.

Even though, the coefficient and positive effect for VR functionalities are quite similar, which was 1.05 and now is 1.01, the effect is not statistically significant in this robustness check. As aforementioned, this statistical insignificance was also seen in the models for user ratings and this was most likely due to the reduced amount of VR video games in the dataset, which is probably the same case for this instance.

Multiplayer functionalities maintain a consistent and statistically significant positive effect on genre classifications, with a coefficient of 1.10 This is nearly identical to the coefficient that was observed in the full dataset, which was 1.11.

The variable regarding the release date also shows consistent results between the robustness check and the full dataset. It holds onto a statistically significant and positive effect with a coefficient of 1.03, which is a slight increase compared to the 1.02 in the full dataset.

Two variables showed similar positive effect sizes in the full dataset, which were supported languages and estimated owners. This remains true in the robustness check with regards to the magnitude of the effect sizes, as they are both identical, to be more precise, they are both 1.00. However, the effect of estimated owners has shifted, with regards to statistical significance, as the effect size now has become statistically significant.

Age rating remains the variable with the least pronounced positive effect on genre classification, as the statistically significant and positive coefficient is 0.725. which is almost identical to the results observed in the full dataset.

In summary, the results that have been observed in the robustness check regarding genre classifications are highly consistent with the results of the full dataset. The biggest changes are the shifts in statistical significance for VR functionalities and estimated owners in genre classifications. However, the sample size for VR video games decreased substantially, which may be the cause for the observed shift in statistical significance. Regarding estimated owners, the robustness check excludes niche video games with low player counts, which may have been the main cause for this shift in statistical significance.

Therefore, due to statistical insignificance hypothesis 4 is still accepted, but weakly supported.

Quasi-Poisson	Genres
	(4)
	Coef.
Variables	(se.)
VR	0.010
	(0.016)
Multiplayer	0.097***
	(0.008)
Release Date	0.027***
	(0.001)
Age Rating	-0.322***
	(0.019)
Supported Languages	-0.001***
	(0.0004)
Estimated Owners	0.166***
	($8.07*10e^{-4}$)
Constant	-52.491***
	(1.959)
Pseudo R^2	0.067
Observations	18,077

Note: *p<0.1; **p<0.05; ***p<0.01

Table 12: Regression estimates for genres in robustness checks

5.6 Genre commonness

Table 13 presents the regression results for the fifth model, which analyses the effect of the independent variables on the commonness of genre combinations in video games. The findings from the robustness check generally align with those of the full dataset, with some differences in effect sizes and a notable difference in the variable age rating. The explained variation in the dependent variable has increased a little bit to 0.043

VR functionalities remain having a statistically significant negative effect on genre commonness, with a coefficient of -66.552. This coefficient indicates that, on average, video games that have access to VR functionalities are associated with 66 fewer video games, than non-VR video games that share the same genre combinations. This effect is slightly lower than that observed in the full dataset. VR functionalities proved to be statistically insignificant in the other models, however for this model it is statistically significant. This suggests that VR functionality has a more direct impact on how genres are combined, instead of on the amount of genres or user evaluation for video games with more than 100 user ratings.

Multiplayer functionalities continue to exhibit a strong and statistically significant negative effect on genre commonness, with a coefficient of-299.575. This effect size suggests that video games with multiplayer functions in their game, on average, are shared with 299 fewer games, compared to single player video games that share the same genre combinations. The magnitude of the effect is nearly identical to that observed in the full dataset.

The release date of video games also maintains a statistically significant negative effect on genre commonness, with a coefficient of -25.158. This indicates that for each additional year the genre combination will see 25 less video games. The effect is slightly increased when compared to the full dataset, but the difference is minor so it is not something to be wary of.

The diffusion of video games also continues to have a statistically significant negative effect on genre commonness. Therefore, video games that have a higher amounts of owners, averagely are associated with 5.72 less video games than those with lower amounts of owners. However, the magnitude of the effect remains negligible, which indicates that the amount of owners has no significant impact on the commonness of genre combinations.

Supported languages remains having a statistically significant positive effect on genre commonness, with a coefficient of 6.622. This indicates that every additional language that is supported in a video game results in approximately two extra games associated with the same genre combination. The effect size of supported languages has slightly increased, but the direction and significance remain consistent.

Finally, the age rating variable displays a positive coefficient of 94.425, which indicates that video games that have an age restriction of 16+ have about 94 more video games associated with the same genre combinations, than those without age restriction. This is a substantial increase in effect size as it has more than doubled compared to the full dataset, where the coefficient was 40. Furthermore, the variable was statistically insignificant in the full dataset, but now does fall in the category of statistical significance. The higher effect size and statistical significance in the robustness check may be the results of a higher proportion of age restricted video games in the dataset.

Overall, the results from the robustness check for genre commonness are consistent with the findings from the full dataset. The directions of the effects remain the same, but there are some differences in effect sizes for certain variables. The most notable variable, age rating, has become statistically significant with a much higher effect size, which suggests that age restricted video games have more common genre combinations. In addition, the effect of VR functionalities on genre commonness is also statistically significant, which suggests that video games with VR functionalities are less likely to have commonly used genre combinations.

Therefore, hypothesis 5 is supported through the robustness check as well.

Ordinary Least Square	Genre commonness
	(5)
	Coef.
Variables	**(se.)**
VR	-66.552**
	(27.643)
Multiplayer	-299.575***
	(13.184)
Release Date	-25.158***
	(1.615)
Age Rating	94.425***
	(28.486)
Supported Languages	6.622***
	(0.673)
Estimated Owners	-5.72***
	(1.61)
Constant	51,601.870***
	(3,259.080)
Observations	18,077
R^2	0.043

Note: *p<0.1; **p<0.05; ***p<0.01

Table 13: Regression estimates for genre commonness in robustness checks

5.7 Multicollinearity

As the number of observations changed due to restricting the dataset to only contain video games with more than 100 user ratings, this could have influence on the multicollinearity of the independent variables. Therefore, VIF tests have also been conducted for the robustness checks, in order to ensure that there is no correlations between independent variables. The results below confirm that also in the robustness checks the VIF tests are all well below the threshold of 5 and indicate that there is no multicollinearity between variables in the regressions.

Variable	VIF range	Outcome
Genres	1.01-1.05	No multicollinearity
AAA Developers	1.01-1.03	No multicollinearity
Multiplayer	1.01-1.02	No multicollinearity
Release Date	1.04-1.07	No multicollinearity
Age Rating	1.02-1.07	No multicollinearity
Supported Languages	1.01-1.02	No multicollinearity
Estimated Owners	1.01-1.02	No multicollinearity
VR	1.00-1.01	No multicollinearity

Table 14: VIF results robustness checks

6. Qualitative results

This section provides more nuance to the results of the quantitative analyses, through the use of content analysis of user reviews to identify which genres of VR video games are perceived as less enjoyable and to explore the underlying reasons for the innovation penalty.

6.1 User reviews

The dataset used for this content analysis comprised approximately 350,000 user reviews. Each review is associated to a single or multiple genres, which have been used to calculate the proportion of negative reviews per genre. This analysis revealed that, of the 11 genres represented in the data, the sports and adventure genres received the highest proportion of negative user reviews. In particular, 30.3% of user reviews for VR video games that have sports elements and 29.7% of user reviews for VR video games with adventure elements were negatively rated. To understand the reasons behind these lower ratings, the contents of user reviews of these genres will be further examined.

6.2 Overview of themes

During the thematic analysis of user reviews for both VR sports and VR adventure video games, several recurring themes were identified. Notably, the main issues that were identified through this content analysis were largely consistent across both genres. As a result, the analysis discusses these themes collectively, instead of separately for each genre. There were some themes that were not consistent across both genres, which have been discussed separately. The identified themes that were present in both genres have been combined into three main issues; technological issues, immersive issues and developers issues. These will be discussed firstly, then to provide further clarification on any themes specific to a genre these will be discussed following the discussion of the three core issues. This has been illustrated in figure 1.

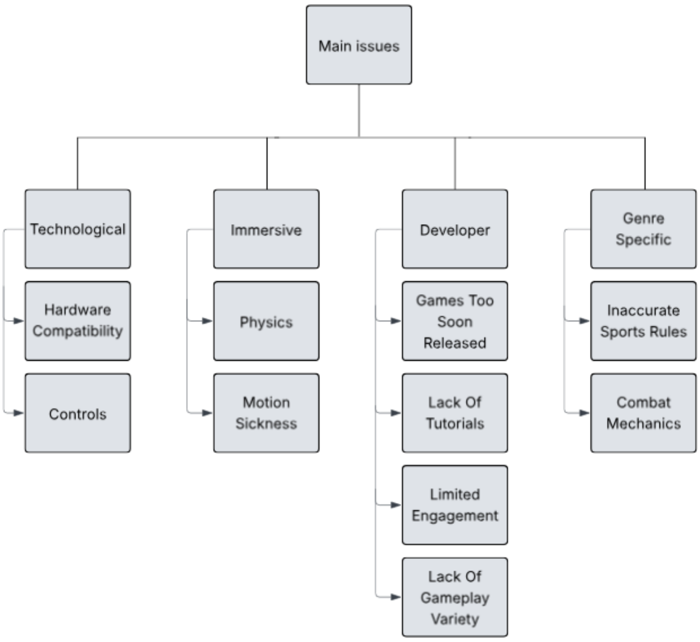

Figure 1: Thematic diagram of qualitative analysis

6.2.1 Technological issues

Hardware compatibility

One of the prominent themes observed from the user reviews was the compatibility of VR video games with certain VR hardware. In particular, a VR sports review stated, *"I am not able to say anything positive about this game. I tried playing it, but the only headset that works is this seven year old one. Message me when they support newer headsets"* (review 37). Users can not enjoy recently published video games, as they only support older VR devices. This resulted in a lot of frustration with users, which is also illustrated in a VR adventure review, *"It is impossible to progress past the main menu. It says I have to use either this VR or this VR headset, unbelievable"* (review 103).

The VR market has been established only recently, but it is steadily growing and therefore diversification regarding VR hardware is seen more. It is important for developers to understand these changes in the market and realise that technical accessibility is important for users to engage with their video game. Otherwise, it is only natural that they receive negative feedback.

Controls

Building on the previously discussed technological issue, another theme regarding technological issues occurred; the controls of the VR video games. Users frequently report difficulties with the controls, which is especially important in VR video games, because these video games are more physical and therefore pose more difficulties with interactions.

Many users mentioned that the controls of the video games were very "clunky". One VR adventure review stated, *"I don't like the controls, I try to hold my gun but because of the controls I keep pointing it to myself. The way items are hold should be different."* (review 95). Other users stated that there was a lack of precision with regards to their physical actions. One user in specific noted: "*I get what they tried with the game, but the hit detection is completely off. I swing with accuracy but the trajectory is nowhere near in the right direction.*" (review 68). This shows that this resulted in outcomes which were not anticipated by the user.

These issues create frustration among users. Users often seek to experience physical actions in VR that they may not be able to perform in real life. However, the risks on physical injury makes it difficult for developers to calibrate controls system that are both realistic and safe. This results in developers simplifying their controls systems, which could make it counterintuitive or more difficult to replicate real life experiences. This trade off between realism and safety is a big challenge for VR video games and therefore these video games are very risky for developers to develop.

6.2.2 Immersive issues

Physics

In addition to the technical issues regarding these VR video games, users express concerns regarding the lack of realistic physics, which ultimately negatively impacts their immersion. Several users reported frustration with the inability to perform certain tasks that would be perfectly fine to perform in real life. One user writhing a VR sports stated, "*It is a good game to learn this sport, but I didn't like that I couldn't pull myself up onto things like I can outside of the game*" (review 78). On the other hand, some noted that they could perform certain actions that defied logic, because of inconsistencies in the physics. For instance, "*It is not a good game. I can lightly touch someone, but the game thinks that I used my full power*" (review 115). This resulted in an unpleasant and unexciting experience.

While VR video games are known for their immersion, which is not always depended on whether the video games adhere to real world physics, however this is valued by users in the context of VR sports and adventure video games. In these genres, users highly value physics as a means to understand and engage with the gameplay. If the video games do not adhere to basic physics principles, users may experience difficulties with how to approach a video game, which in turn can contribute to negative user experience. This suggests that for VR sports and adventure video games, compliance to real world physics is crucial in order for them to maintain immersion. Therefore, it is important for developers of VR sports and adventure video games to prioritise real world physics, so user can be fully engaged and to support user expectations.

Motion sickness

Another factor regarding users' immersion is the reoccurrence of motion sickness in these VR video games. Multiple users mentioned that specific elements from the video games cause nausea, which had forced them to stop playing the video game. This is exemplified in the following VR sports review: *"I got motion sickness after 5 minutes playing this video game. The video game itself seemed great, but I did not have the opportunity to play it again"* (review 43). Even though, the video game seemed to be of high quality, the discomfort brought by the motion sickness resulted in less engagement with the video game. Similarly, a VR adventure review declared, *"I need to constantly move and there is a lot of rotation going on, so the game is unplayable for me because I get motion sickness"* (review 21). This suggests that repetitive or advanced movements, which involve rotations or other forms of movement that changes the users' perspective, may negatively influence user experience.

While VR video games may be perfectly developed, where functionalities and mechanics are optimised, the realism may result into unintended inconveniences. Repeated actions such as jumping, rotating or flipping could be used too many times in a short time window without considering the possible consequences of this. Specifically in VR sports and adventure video games this is pertinent, as such dynamic movements are generally frequently seen. Therefore, it is important for developers to create a good balance between simple and advanced movements and limit movements that are likely to induce motion sickness. These changes have the potential to benefit user experience of VR video games.

6.2.3 Developer issues

Video games are released too soon

Users frequently experienced issues with the state of the VR video games that were released. According to multiple users, the video games appeared as unfinished and therefore it was perceived that the video games were released prematurely. So did the user behind one review, which stated: "*The game is buggy, it doesn't listen to you and it lacks polish. I can't recommend it to anyone, I don't know why it was ever released*" (review 12).

Several users stated that full released video games resembled early access games when compared to their functionalities and refinement. This is accurately described in one review, which states: "*The game is not fun at all, even for a VR game. I read some reviews that said that it resembles a trial and not a video game, they were right*" (review 55). In some cases, video games that were categorised as early access games were blamed for being functionally and visually terrible and were unjustified to be even released.

This suggests that developers should put more emphasis on sufficient testing and refinement of their video games prior to release. The visual and functional deficiencies that would have been noticed if the video games were published before testing can now be avoided through several testing methods, such as appointing a group of testers beforehand. This would help to confirm that there are little to no unintended consequences of in-game actions, which in turn would benefit user experience.

Lack of tutorials or insufficient tutorials

Another interesting issue that was identified is the absence or insufficiency of tutorials in these VR video games. Users frequently report that the tutorials were either missing or were insufficient with regards to detail in providing information on the understanding of the video game, which led to them stop playing before they could experience the gameplay. One user mentioned, "*If you want to keep your sanity, don't ever play this game. It is so frustrating to understand how to play this game. The tutorial is so bad that it only contains pictures that don't say anything useful*" (review 85). Another user stated that the video game did not have any form of tutorial, which ultimately forced them to quit the video game: "*I only got a small idea of what I had to do through internet searches, because there just wasn't a tutorial. I got through the first couple of levels, but after that I just didn't know what to do and quit*" (review 36).

Video games should possess a detailed and working tutorial, which in particular is important for VR video games, as the controls for these video games are often more complex and less obvious for the average user. In the absence of such a tutorial, the threshold of backing out of a game becomes much lower, when it requires more effort to understand the video game rather than playing it. This may result in negative user feedback, even in cases when the video game might be well designed, however the user could not experience this well designed video game, as it was not accessible for them. Therefore, it is important for developers to ensure that their VR sports and adventure video games include a tutorial, but also make sure that the tutorial is detailed, understandable and accessible for the users. This could improve user experience and consequently increase the proportion of positive user feedback.

Limited engagement from developers

One of the most prominent themes identified in the user reviews is the engagement from developers with their community. Significant users have indicated that developers are not engaged with their community and even had the feeling that they did not care about them. One review of a VR adventure video game stated, "*The game could be fun, but is doesn't work. It feels like the developers put the game on Steam and then just vanished*" (review 33). Recurring issues with gameplay and request for quality of life changes were mentioned on online forums by the community, but without any response from the developers. This is portrayed by an user stating, "*Why do you release a game and not respond on forums? Good luck getting feedback when you don't recognise us*" (review 42). There were several users that even had direct contact with developers, where they offered suggestions or criticism, but which were disregarded.

It should be essential for developers to engage with their community, as their user base is the most important source of information regarding their video game as they, ultimately, are responsible for whether the video game becomes successful or not. If developers lack in communication, many users might feel undervalued, which could have detrimental consequences for user retention. Users might choose to abandon the video game and search for better alternatives.

Therefore, it is crucial that developers carefully consider community feedback and where possible implement the feedback into the video game. In the end, there will be cases where feedback can not be implemented in the video, however transparency on the rationale behind the decision making process is very important. This could benefit compliance and trust from the community even when the expectations are not completely fulfilled, which increases user loyalty.

Lack of gameplay variety

The final theme related to developer issues concerns the lack of variety in gameplay. Multiple users stated that the VR sports and adventure video games they played quickly became boring, due to repetitive gameplay or they were not replayable. One user in particular said, "*Didn't play the game for long. Its just the same things over and over again and becomes boring. Not worth it*" (review 88). Others acknowledged the potential of the video games, but due to the lack of playable levels it became too simple and monotonous. One reviewer stated, "*It is a really exciting game to play, but there isn't enough content. If new things would be released it could be a good game, but for now I can't recommend it*" (review 93).

These statements from the users indicate that even though a video game can inherently be well designed, but fail to retain user engagement, because the gameplay is linear and lacks variety. This issue may especially occur in the context of VR sports video games, where users could have the opportunity to play in different environments and conditions to replicate it for in the real world. For that reason, a lack of variety in gameplay could decrease user engagement and therefore user experience.

Consequently, it is advised for developers to carefully consider, whether the VR sports and adventure video games they want to publish offer sufficient longevity. This is mostly related to whether the video games provide adequate and diverse content to satisfy the users' needs. This would keep users engaged for a longer period of time, which may positively influence their experience.

6.2.4 Genre specific themes

As aforementioned, certain themes were identified unique to either VR sports video games or VR adventure video games. These genre specific themes are discussed in the following section to provide more clarification.

Inaccurate sports rules

An issue regarding the immersion of the users unique to VR sports video games, is the implementation of the rules of the sports that they are representing. According to several users, there are certain VR sports video games that did not adhere to the standard rules of the respective sport. One user in particular said, "*The game stops way too soon, the rules say it should not be this soon. Serving also is just not correct: the way I have to serve in the game is not how you should do it in real life*" (review 68). These inaccuracies made playing the game feel counterintuitive and created an unpleasure experience.

Another user and dedicated fan of the sport stated, "*I think I played all the games of this sport, but this one is just not it. I don't think the developers have ever seen or played a game of this sport once in their life. This game has the worst gameplay, if you want to improve it learn the basics of the sports at least*" (review 111). This has resulted in frustration and less immersion.

When the rules within the VR sports video game diverge from the rules that the users are familiar with can lead to negative experiences as immersion is disrupted. This is especially the case if users have set certain expectations on how the video game will work and then it is completely not in line with those expectations. Therefore, developers should conduct research into nature of the sports that they are incorporating in their video game. This would help ensure that the rules are accurately portrayed in their video game. This may result in less resistance and more positive feedback from users.

Combat mechanics

A recurring theme that has been identified in user reviews of VR adventure video games is the dissatisfaction of combat mechanics in these video games. Multiple users shared how they are dissatisfied with how combat works in the video games. For example, one review stated: "*There are only a couple moments of combat and the mechanics are so bad maybe the worst I've seen. This feels more like a simulation as you are almost only walking around*" (review 12). Some noted that the combat was too easy and simplistic, which reduced their enjoyment from the video game. Hence, one review stated: "*When there was combat it wasn't fun. It was like I was just poking the enemy and they would just vanish*" (review 65). Others thought that the mechanics were not appealing, which made it difficult to immerse themselves in the video game. One user stated, "*The game is boring, fighting is just not fun. I would not waste your money on this*" (review 89).

VR adventure video games revolve a lot around combat, therefore it is crucial that this is implemented correctly and engagingly. However, incorporating complex combat mechanics, which would make them more profound and engaging is challenging to implement in VR video games. More complex combat mechanics could require more physical restrain on the user, which may raise concerns about user safety. These constraints intrinsic to the VR technology of today may limit developers to implement combat mechanics that are engaging, but also safe to use. Future developments in VR technology may be necessary, so that more complex combat mechanics can be safely incorporated in such video games.

7. Discussion

This section reflects on the findings from the quantitative and qualitative analyses, while examining their broader theoretical implications. In addition, limitations regarding this research will be discussed and practical implications will be made for developers in accordance to the results.

7.1 Theoretical implications

The quantitative analysis aimed to assess how categorical membership and user evaluation are influenced by VR technology within the video game industry. This has been put into practice by formulating six hypotheses to test their association.

Hypotheses 1 and 2 were formulated to assess whether video games that are classified under multiple genres receive lower ratings than single-genre video games and whether this is affected by the nature of the developers. The results are in accordance with these two hypotheses, as they both showed statistically significant negative results. That is, video games that are classified under multiple genres received lower user ratings compared to video games that are categorised under a single genre. And, multi-genre video games produced by novel developers received lower user ratings compared to multi-genre video games produced by incumbent developers. Hypothesis 3 predicted that VR video games tend to receive lower user ratings than non-VR video games, which was also validated. Contrary to these hypotheses, however, hypothesis 6 did not receive empirical support. That hypothesis predicted that multi-genre VR video games developed by novel developers receive higher ratings than. multi-genre VR video games developed by incumbents. Given the lack of empirical support, this hypothesis consequently was rejected.

Hypotheses 4 and 5 addressed categorical membership with regards to VR video games. Hypothesis 4 expected that VR video games are more frequently classified under multiple genres than non-VR video games. This was supported by the results given the statistically significant, positive effect. Additionally, hypothesis 5 gave more insights into the definitions of categorical membership by VR video games. It suggested that VR video games use less conventional combinations of genres, than observed in non-VR video games. This hypothesis also received empirical support.

The results from this analysis are closely related to the theory of disruptive innovation. The theory of disruptive innovation states that these types of innovations initially show lower performance, when they are introduced onto the market. This lower performance is observed in the lower user ratings that VR video games receive, which is used as a proxy of their performance. This is further confirmed through the findings from the qualitative analysis. Users that experience difficulties with the controls not working, the video games not being compatible with their hardware, or constant experience of motion sickness. These are all instances, where the technology does not perform up to standards with regard to traditional video games. However, VR video games do provide new value through other characteristics that are absent in traditional video games, such as its immersion. Therefore, due to its initial lower performance and new added value, VR video games can be seen as disruptive innovations.

The findings also support categorical theory on disruptive innovation by Frenken and Punt (2023). According to Frenken and Punt (2023), disruptive innovations are innovations that eventually are categorised in an existing category by stretching its boundaries so that it can fit in the category. VR did not form a distinct genre, but rather was seen to be added as a tag, which suggests that it is still categorised within the same category as traditional video games. Therefore, it can be concluded that the technology does not function as a radical innovation, as it would have opened up a completely new category, but rather functions as a disruptive innovation that stretches the categorical boundaries of video games.

Additionally, the findings indicate that VR video games tend to adopt more distinct genres and use less conventional combinations of genres. This suggests that the introduction of VR technology has disrupted how genre classification functions in the gaming industry. The value that VR offers creates a shift in genre classification, as more genres are now applicable simultaneously, which has caused this increase in genre applicability. This has opened up more variations in genre usage as less conventional genre combinations can be used. This immersion that VR offers creates numerous possibilities for developers to explore new elements, which have become viable due to VR functionalities. As a results, categorical boundaries have become more fuzzy, which allows for more creativity from developers to innovate.

This is also supported by the results of the qualitative analysis. In particular, it is mentioned that certain elements of the VR video games did not work as expected or intended by the developers, which could be genre defining. For instance, this is shown in the issue regarding combat mechanics for VR adventure video games. If characteristics that define category boundaries are missing in a video game, it could contribute to fuzziness of categories and therefore stretch these boundaries. Sports video games where realistic physics do not apply could also be seen as a fantasy video game, where the laws of nature can be defied or an adventure video game without proper combat mechanics may be perceived as a simulation where they only can wander the digitally constructed world. These findings suggest that VR technology does not adhere to traditional categorical boundaries and therefore stretch these boundaries through fuzziness.

However, even though the fuzziness of boundaries may suggest that developers have more freedom to experiment with more innovative and unconventional video game designs, this research also highlights the relevance of the theory described by Zuckerman (1999), where he introduces the notion of innovation penalty. This is prominently evident in the context of VR video games, as it is observed that VR video games deviate from the exemplars in traditional gaming, making it difficult for users to accurately categorise the video game. This ambiguity has resulted in notably less positive user feedback than in traditional video games, which suggests that deviation from the categories established in traditional gaming has consequences. This is supported throughout the findings, such as the lower user ratings for VR video games, which shows that defection from the established norms has contributed to negative feedback. In addition, user sentiments illustrate how new features in VR video games have lead to frustration. The lack of tutorials or controls not working properly have all been the result of deviating from the categorical exemplar, which has resulted in penalties from the userbase.

7.2 Limitations

This research has aimed to minimise the potential for error through careful consideration within its methodology, nevertheless there were some limitations that could not be avoided.

One limitation is related to the nature of the research process. All analyses were conducted by a single researcher. This may have impact on the reliability of this research, because themes and individual sentiments were interpreted solely based on the perspective of the researcher. This could have resulted in the interpretation of certain themes or contents from the reviews, which are not shared by other researchers.

In addition, due to time constraints, the qualitative analysis was limited to video games that had explicitly “VR” in their titles, which made it evident that they had access to VR functionalities. As a result, important and insightful reviews from other relevant VR video games are possibly excluded from the analysis, which could have contained insightful information. Additionally, the analysis was limited to the two lowest rated genres. A broader analysis of supplementary lower rated genres could have resulted in further insights into challenges facing VR video games. This may have had impact on the validity of this research.

Further research could benefit from a more thorough qualitative analysis for video games with VR functionalities, despite of their title. This could identify new themes that may have been missed in this research. It could also be insightful to reconsider this topic, when VR technology is closer to technological maturity and has diffused more. This could give new insights in how VR technology has further disrupted the gaming industry, through categorical boundary shifts and if categorical penalty remains a dominant factor within VR video games.

7.3 Recommendations

Based on the findings, several recommendations are proposed for developers. These recommendations aim to inform developers about certain pitfalls with regard to VR video games and thereby enhancing user experience.

First, it is advised for developers to be careful when diverging from traditional video games, with regards to genre classifications. Even though, the shift in categorical boundaries opens up possibilities for developers to be more innovative with their video games, due to categorical penalty may not be well received by audience members. Therefore, strong considerations should be made when considering the use of certain genre combinations in VR video games.

Second, technological functionality is crucial for positive user experience. To be more precise, developers should ensure compatibility of their video games with a large portion of developed VR hardware. Limited hardware support may decrease its adoption rate and does not offer sustained benefits. Additionally, the performance and mechanisms of the controls are important. Poorly functioning controls may lead to users getting frustrated, due to the inability to perform intended actions within the video game. This may result in low customer retention.

Given that VR technology adds value through its immersion, it should be a strong point for these video games. Thus, regarding video games that try to depict real life situations have to ensure that their applied physics are realistic. A lack of physical realism can disrupt the immersion of the video game, which has consequences for user experiences. Moreover, it is recommended that the video games do not involve a lot of unnecessary and excessive mobility mechanics, such as flips or rotations. These mobility mechanics may induce motion sickness, when used immoderately, which is not preferred by users.

It is also important for developers to maintain active communication with the user base. The user base experience the video game first hand and therefore are able to identify shortcomings, offer suggestions and highlight points of improvement. While not all feedback can be implemented or should be implemented, it is crucial to maintain transparent communication regarding the reasoning behind the decisions that are made. This builds trust between the developers and user base.

Furthermore, it is advised that, prior to release, developers should test their video games to check for bugs or errors. This can be done either internally or with the help of beta tests. This offer the opportunity to identify bugs and errors before releasing the video game, which enhances user experience. Additionally, the video games should have access to a good tutorial. Tutorials should be clear, concise and comprehensive so that users can understand and engage with the video game. Finally, the video game should offer a variety of gameplay to retain customers. A lack of variety in the video game tasks may lead to repetitive gameplay, which may induce boredom into the users. Thus, it is important the video game has multiple dynamics so that the user is always entertained.

8. Conclusion

This thesis aimed to look at the influence of VR technologies on categorical membership and user evaluation in the video game industry. This has been assessed through the utilisation of both quantitative regression analyses and qualitative analysis of user reviews.

The quantitative analysis utilised regression analyses to validate six hypotheses. The first hypothesis was validated, as it showed that video games that are classified under multiple genres receive lower user ratings. The second and third hypotheses also received support, as it was found that lower ratings for multi-genre video games were especially observed in video games developed by novel producers and VR video games receive lower user ratings than non-VR video games. However, no evidence was found for the hypothesis that VR multi-genre video games would receive lower rating when developed by incumbents compared to newcomers.

In addition, it was assessed whether video games that have access to VR technologies tend to be classified under multiple genres, which was validated. These genres that are being used in VR video games were also less commonly seen in traditional video games, therefore VR video games are seen as categorical disruptive technologies, as VR video games challenge the existing categorical structure, by stretching them which causes them to become more fuzzy. This confirms the notions of Zuckerman (1999) and Frenken and Punt (2023), that categorical penalty occurs among VR video games, due to their deviation from the traditional categorical norms.

In order to add more nuance to the findings from the quantitative analysis, a qualitative analysis was conducted. This analysis identified recurring themes for the two lowest rated genres; sports and adventure. The findings from this analysis showed that VR video games often struggle with three sets of issues that persist. These three issues that were identified are technical, immersive and developer issues. These issues contained problems such as, hardware incompatibility, lack of tutorials, motion sickness and limited engagement with the user base. These issues currently inhibit the development of VR video games and must be addressed in order for VR video games to mitigate innovation penalty.

Even though VR technology opens up opportunities for more creativity among developers, through innovative combinations of genres, they still are exposed to legitimacy constraints due to categorical penalty. Therefore, it is crucial that developers maintain a healthy balance between creativeness and user expectations through genre classification and user feedback.

References

Anne-Marie, T., Chau, N., & Kai, K. K. (2017). Ethical questions related to using netnography as research method. *The ORBIT Journal*, *1*(2), 1-11.

Backlinko (2024). *Steam users.* Accessed on 10-01-2025, from: https://backlinko.com/steam-users

Bian, S. (2023). Research on the Application of VR in Games. *Highlights in Science, Engineering and Technology*, *39*, 389-394.

Boghossian, J., & David, R. J. (2021). Under the umbrella: Goal-derived category construction and product category nesting. *Administrative Science Quarterly*, *66*(4), 1084-1129.

Bove, V. M. (2018, April). Immersive, Social Applications for 8K Displays. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 136-139). IEEE.

Brunt, C. S., King, A. S., & King, J. T. (2020). The influence of user-generated content on video game demand. *Journal of Cultural Economics*, *44*, 35-56.

Christensen, C. (1997). *The innovator's dilemma*. Harvard Business School Press, Cambridge, Mass.

Cudennec, A., & Durand, R. (2023). Valuing spanners: Why category nesting and expertise matter. *Academy of Management Journal*, *66*(1), 335-365.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, *49*(10), 1407-1424.

Durand, R., & Boulongne, R. (2017). Advancing research on categories for institutional approaches of organizations. *The SAGE handbook of organizational institutionalism*, 647-668.

Durand, R., Granqvist, N., & Tyllström, A. (2017). From categories to categorization: A social perspective on market categorization. In *From categories to categorization: Studies in sociology, organizations and strategy at the crossroads* (Vol. 51, pp. 3-30). Emerald Publishing Limited.

Frenken, K., & Punt, M. B. (2023, August). A new view on radical innovation. In *International Sustainability Transitions conference, Utrecht, the Netherlands*.

Guzsvinecz, T., & Szűcs, J. (2023). Length and sentiment analysis of reviews about top-level video game genres on the steam platform. *Computers in Human Behavior*, *149*, 107955.

Hannan, M. T. (2010). Partiality of memberships in categories and audiences. *Annual Review of Sociology*, *36*(1), 159-181.

Harris, T., Yang, Z., & Hardin, J. W. (2012). Modeling underdispersed count data with generalized Poisson regression. *The Stata Journal*, *12*(4), 736-747.

Henkel, J., Rønde, T., & Wagner, M. (2015). And the winner is—Acquired. Entrepreneurship as a contest yielding radical innovations. *Research Policy*, *44*(2), 295-310.

Hsu, G. (2006). Jacks of All Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production. *Administrative Science Quarterly*, 51(3), 420– 450.

Hsu, G., Hannan, M. T., & Koçak, Ö. (2009). Multiple Category Memberships in Markets: An Integrative Theory and Two Empirical Tests. *American Sociological Review*, 74(1), 150–169.\

Khan, M. K. (2020). Technological advancements and 2020. *Telecommunication Systems*, *73*, 1-2.

Kovács, B., & Hannan, M. T. (2010). The consequences of category spanning depend on contrast. In *Categories in markets: Origins and evolution* (pp. 175-201). Emerald Group Publishing Limited.

Kozinets, R. (2010). Netnography. Doing Ethnographic Research Online. By Robert. *Canadian Journal of Communication*, *38*(1).

Lackermair, G., Kailer, D., & Kanmaz, K. (2013). Importance of online product reviews from a consumer's perspective. *Advances in Economics and Business*, *1*(1), 1-5.

Laerd (n.d.) *Spearmans rank order correlation statistical guide.* Accessed on 24-04-2025, from: https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide-2.php#:~:text=The%20Spearman%20correlation%20coefficient%2C%20rs%2C%20can%20take%20values%20from,the%20association%20between%20the%20ranks.

Leung, M. D., & Sharkey, A. (2009). Out of sight, out of mind: the mere labeling effect of multiple-category memberships in markets. *Work. Pap., Stanford Grad. Sch. Bus., Stanford, CA*.

Lowe, A., Norris, A. C., Farris, A. J., & Babbage, D. R. (2018). Quantifying thematic saturation in qualitative data analysis. *Field Methods*, *30*(3), 191-207.

Machała, S., Chamier-Gliszczyński, N., & Królikowski, T. (2022). Application of AR/VR Technology in Industry 4.0. *Procedia Computer Science*, *207*, 2990-2998.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets?. *Memory & Cognition*, *6*(4), 462-472.

Medium (2021). *Why did Harley Davidson fragrances fail?* Accessed on 16-12-2024, from: https://medium.com/@david.t_34066/why-did-harley-davidson-fragrances-fail-642b124566e0

Melnikava, D. V., Ponyatov, A. A., & Murzich, L. F. (2021). AR/VR technology in the gaming industry. *Turk. J. Comput. Math. Educ.(TURCOMAT)*, *12*, 52-53.

Meta (2024). *Meta Quest.* Accessed on 12-01-2025, from: https://about.meta.com/technologies/meta-quest/

Motamedi-Fakhr, S., Moshrefi-Torbati, M., Hill, M., Hill, C. M., & White, P. R. (2014). Signal processing techniques applied to human sleep EEG signals—A review. *Biomedical Signal Processing and Control*, *10*, 21-33.

Ozili, P. K. (2023). The acceptable R-square in empirical modelling for social science research. In *Social research methodology and publishing results: A guide to non-native English speakers* (pp. 134-143). IGI global.

Philp, M., & Nepomuceno, M. V. (2024). How reviews influence product usage post-purchase: An examination of video game playtime. *Journal of Business Research*, *172*, 114456.

Pontikes, E. G., & Kim, R. (2017). Strategic categorization. In *From categories to categorization: Studies in sociology, organizations and strategy at the crossroads* (pp. 71-111). Emerald Publishing Limited.

Prolific Studio (2024). *Indie vs AA vs AAA games.* Accessed on 11-01-2025, from: https://prolificstudio.co/blog/indie-vs-aa-vs-aaa-games/

Researcher (2024. *What is purposive sampling? Methods, techniques and examples.* Accessed on 09-06-2025, from: https://researcher.life/blog/article/what-is-purposive-sampling-methods-techniques-and-examples/

Schumpeter, J. A. (1942). Socialism, capitalism and democracy. *New York: Harper & Brothers.*

Schwab, K. (2024). The Fourth Industrial Revolution: what it means, how to respond. In *Handbook of research on strategic leadership in the Fourth Industrial Revolution* (pp. 29-34). Edward Elgar Publishing.

Scribbr (2023). *Control variables.* Accessed on 17-01-2025, from: https://www.scribbr.com/methodology/control-variable/#:~:text=Control%20variables%20enhance%20the%20internal,and%20helps%20avoid%20research%20bias.

Scribbr (2023). *Understanding confidence intervals. Easy examples & formulas.* Accessed on 11-06-2025, from: https://www.scribbr.com/statistics/confidence-interval/#:~:text=A%20confidence%20interval%20is%20the,another%20way%20to%20describe%20probability.

Shah, S. F. A., Mazhar, T., Shahzad, T., Ghadi, Y. Y., & Hamam, H. (2024). Integrating educational theories with virtual reality: enhancing engineering education and VR laboratories. *Social Sciences & Humanities Open*, *10*, 101207.

Spearman, C. (1961). The proof and measurement of association between two things. *New York: Appleton-Century-Crofts.*

Statista (2024). *Annual gaming company investment segment.* Accessed on 12-01-2025, from: https://www.statista.com/statistics/1452453/annual-gaming-company-investment-segment/

Statista (2025). *Share of game developers worldwide working on game projects for select VR/AR platforms in 2024* Accessed on 02-03-2025, from: https://www.statista.com/statistics/1060239/game-developers-vr-ar-platforms/

Statista (2024). *Steam gaming platform – statistics & facts.* Accessed on 10-01-2025, from: https://www.statista.com/topics/4282/steam/#editorsPicks

Statista (2024). *Steam game sales units.* Accessed on 10-01-2025, from: https://www.statista.com/statistics/1333058/steam-game-sales-units/

Statista (2023). *VR gaming penetration rate vs video game penetration global.* Accessed on 12-01-2025 from: https://www.statista.com/chart/29401/vr-gaming-penetration-rate-vs-video-game-penetration-global/

Statsig (2025). *90% vs. 95% confidence interval: which to use?* Accessed on 11-06-2025, from: https://www.statsig.com/perspectives/confidence-interval-90-vs-95

Statstutor (n.d.). *Spearmans*. Accessed on 24-04-2025, from: https://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf

SteamDB (2017). *SteamDB rating.* Accessed on 15-01-2025, from: https://steamdb.info/blog/steamdb-rating/

Theodoropoulos, A., & Antoniou, A. (2022). VR games in cultural heritage: A systematic review of the emerging fields of virtual reality and culture games. *Applied Sciences*, *12*(17), 8476.

Tiwari, I., Solankar, D., & Khande, R. (2023). Virtual and augmented realms: the evolution of video games with VR and AR technology. *Vidhyayana-An International Multidisciplinary Peer-Reviewed E-Journal- 8*(7), 545-561.

Tripsas M, Gavetti G. 2000. Capabilities, cognition, and inertia: evidence from digital imaging. *Strategic Management Journal* 21(10/11): 1147–1161.

Tulane University (n.d.) *What the difference between AR and VR.* Accessed on 12-01-2025, from: https://sopa.tulane.edu/blog/whats-difference-between-ar-and-vr

Valve Corporation (2022). *Applying tags.* Accessed on 27-03-2025, from: https://partner.steamgames.com/doc/store/tags

Van Nguyen, S., Le, S. T., Tran, M. K., & Tran, H. M. (2022). Reconstruction of 3D digital heritage objects for VR and AR applications. *Journal of Information and Telecommunication*, *6*(3), 254-269.

Vergne, J. P., & Wry, T. (2014). Categorizing categorization research: Review, integration, and future directions. *Journal of Management Studies*, *51*(1), 56-94.

Verhulst, I., Woods, A., Whittaker, L., Bennett, J., & Dalton, P. (2021). Do VR and AR versions of an immersive cultural experience engender different user experiences?. *Computers in Human Behavior*, *125*, 106951.

Weitkamp, L., van der Pol, E., & Akata, Z. (2019). Visual rationalizations in deep reinforcement learning for atari games. In *Artificial Intelligence: 30th Benelux Conference, BNAIC 2018,'s-Hertogenbosch, The Netherlands, November 8–9, 2018, Revised Selected Papers 30* (pp. 151-165). Springer International Publishing.

Wilms, R., Mäthner, E., Winnen, L., & Lanwehr, R. (2021). Omitted variable bias: A threat to estimating causal relationships. *Methods in Psychology*, *5*, 100075.

Wry, T., & Lounsbury, M. (2013). Contextualizing the categorical imperative: Category linkages, technology focus, and resource acquisition in nanotechnology entrepreneurship. *Journal of Business Venturing*, *28*(1), 117-133.

XLstat (n.d.). *Ordinary least squares regression ols.* Accessed on 16-01-2025, from: https://www.xlstat.com/en/solutions/features/ordinary-least-squares-regression-ols

Yarnell, D., & McGilvray, G. (2022, March). Behind the scene: delivering 4K olympics games. In *Proceedings of the 1st Mile-High Video Conference* (pp. 128-128).

Zuckerman, E. W. (1999). The categorical imperative: Securities analysts and the illegitimacy discount. *American Journal of Sociology*, *104*(5), 1398-1438.

Zuckerman, E. W. (2017). The categorical imperative revisited: Implications of categorization as a theoretical tool. In *From categories to categorization: Studies in sociology, organizations and strategy at the crossroads* (Vol. 51, pp. 31-68). Emerald Publishing Limited.