

Project 2 Qualitative – *Safety measures on vide-sharing platforms survey*

Produced by: Yonder

Fieldwork: 21st September – 8th October 2020

Definitions and Clarifications

Video Sharing Platforms: This research explored a range of websites and apps used by people in the UK to watch and share videos online. Although the term ‘video sharing platforms’ (VSPs) is used, this research does not seek to identify which services will fall into Ofcom’s regulatory remit, nor to pre-determine whether any particular service would be classed as a VSP under the regulatory definition. It should also be noted that the platforms included in this research operate at different scales. This should be taken into consideration when comparing results from users of smaller VSPs against those from users of larger platforms.

Site and App Content: This research explored a range of sites and apps which have video sharing functionalities. Many of these platforms also contain a mix of video and other types of content and allow users to view and participate in a range of ways, of which video sharing is one element.

Sources of Evidence: Evidence in this research is self-reported by respondents who have shared their experiences, recollections and feelings. All respondents participated voluntarily and were free to withdraw their participation at any point during the research process. As such, the evidence is limited by respondents’ freedom to decide whether to participate, their ability to recall events, accuracy of that recall and which experiences they chose to disclose.

Objectives and Methodology

Objectives

- Understand VSP users' attitudes towards existing appropriate measures, how they could be improved and how they define appropriate social etiquette on VSPs

Sample

- 24 adults aged 18+ in the UK
- Quotas set on region, gender, usage of platforms (light, medium, heavy) and working status
- Mix of used platforms: YouTube, Instagram, Snapchat, TikTok, Facebook, Twitter, LiveLeak, Vimeo

Data collection

- Online interviews and online workshops (with online pre-task)
- Conducted by Yonder
- Fieldwork from 21st September to October 8th 2020

Overview of users – similarities and differences



Insight	Light	Medium	Heavy
Usage and engagement	<p>Light users visit VSPs a few times a week and they only sometimes share content, without creating their own – they tend to believe that they lack the knowledge and awareness of how they can stay protected on VSPs</p>		<p>Medium and Heavy users visit VSPs multiple times a day and regularly share and upload their own content. As a result, they tend to feel more confident when talking about how the platforms operate, what is and what is not possible and how they can stay protected on VSPs</p>
Awareness of safety tools	<p>Light users tend to have very little knowledge of the tools available to them, and can typically only highlight the ‘report’ function. Otherwise, they sometimes ‘reduce’ their usage in order to avoid harmful content</p>		<p>Medium and Heavy users tend to have a higher awareness of the tools available to them, and highlight their experiences as uploaders of original content as to why they have significant knowledge. They tend to feel that as both an uploader and a viewer, they are more exposed to information from VSPs about their respective rules and policies</p>
Platform responsibility	<p>The most significant similarity between Light, Medium and Heavy users of VSPs is the perception that platforms are not taking enough responsibility for harmful and inappropriate content that is being uploaded and as a result, the protection of its users. They all tend to agree that more action must be taken to reflect what other entertainment mediums are doing</p>		

Section 1

Attitudes towards existing measures

Users have a universally accepted view of what ‘harmful’ content is, but also acknowledge that VSPs are a difficult place to ‘police’

- When discussing harmful or inappropriate content, users generally agreed on the following themes:
 - **Violence (e.g. death, injury, fighting)**
 - **Illegal activities**
 - **Abuse directed towards other users (which was heightened for minorities within platforms)**
- These areas are generally deemed ‘not allowed’ on platforms – users do not want to see them on the platforms they use, but ultimately, *have* seen such things and so believe that no platform is free from this
- All users tend to feel it is difficult to put a blanket ban on these themes when some content requires context – i.e. some content can use the above themes in relevant and appropriate ways to highlight issues (e.g. animal abuse and the work that charities do to help tackle this). As a result, medium-heavy users are more vigilant of the uploader and description of content as a means of identifying whether something is ‘safe’
- Ultimately, users tend to believe that every piece of content must be viewed before any decision can be made against it, as ‘context’ is perceived as ‘key’

Awareness of existing measures and tools is high amongst some users, and most users have taken advantage of them in order to protect themselves

- Medium and Heavy users are most aware of the various measures that are available across multiple platforms, whereas Light users tend to only be aware of one or two. The most recognised are:

Report

Block

Age gates

Safety
centres

- Additionally, Medium and Heavy users tend to say that they have used one or more of the above tools when they have identified harmful or inappropriate content with the desire to have it removed from 'their timeline'. They feel confident in using them and believe they are 'anonymous' actions – they can use them multiple times with no public repercussions
- However, most users were unaware of tools and measures such as 'machine learning' and using 'experts' to develop new policies for platforms. Ultimately, all users tend to believe that this *appears* to be a positive step in keeping people safe, but do not believe there is enough 'tangible evidence' to know whether they are having an effective impact

Despite the level of awareness and use, most users tend to believe that the tools are not effective enough

- **Ineffective tools:** While most users state that they have taken advantage of the tools available to them in reaction to viewing harmful or inappropriate content, the majority do not believe that they are effective enough in removing the content or stopping future content from appearing on platforms
- **Reporting & blocking:** When ‘reporting’ harmful or inappropriate content, users do not believe that the report is immediately actioned, as the content is still viewable. As a result, most believe that the content can take 24+ hours for it to be removed. This is frustrating, particularly as they can see other users (e.g. via comments) stating that they too have reported the content
 - Consequently, users also ‘block’ the content (or the content creator) in order to have it ‘removed’ from their timeline. However, users acknowledge that this is a very ‘personal’ outcome – the content disappears for them but remains viewable by everyone else
 - Users therefore rely on the report functionality, but just do not believe that it is ‘fast’ enough, and instead relies on bulk reporting by the community for it to attract the platform’s attention. Users state that this is damaging due to the nature of the content and the impact it can have on people’s wellbeing in the time that it is viewable

The process of some of the measures and tools are not perceived to be transparent – most users want greater information and education

- **Lack of transparency:** One of the most frustrating aspects of the platforms' measures and tools for all types of users is the perception that there is a lack of transparency about the process platforms take when reviewing inappropriate and harmful content
- **Reviewing of reports:** This begins with the initial experience of the report functionality and is perceived to vary platform to platform, which includes the ability to refine the nature of the report. Some users state that some platforms do not allow users to 'fine-tune' their view, and instead, must rely on 'generic' drop-down menus, whereas other platforms include 'write-in' options to help the platform understand why the content is being reported. This lack of consistency between platforms helps to drive concerns about how platforms are reviewing reports and whether they are being taken seriously enough
- **Platform responses:** Users have not been satisfied with the platforms' responses to reports. If, for example, a piece of content is not found by the platform to have broken its guidelines or rules, some users have stated that they were disappointed with the result, and this disappointment is strengthened when platforms are perceived to not provide detailed explanations or statistics about the nature of the content and the associated reports (e.g. how many people reported it)

Heavy users rely on the platforms' personalisation 'algorithms' to protect them from harmful content

- **Tailored platform:** Heavy users felt more confident in the platforms' ability to 'shield' them from inappropriate and harmful content due to their belief that the more they use the platform, the more their 'feed' will be populated with recommended content that they are interested in viewing
- **Light users feel more susceptible:** This is in direct contrast with light users, and particularly those who do not sign into a platform, who feel more susceptible to inappropriate and harmful content – they feel that there are a lack of options to protect themselves – they often cannot 'predict' what their viewing experience will be like. Those with profiles though believe that this itself is a protection measure
- **Perceptions can be affected:** However, when heavy users and those with profiles are finally exposed to inappropriate and harmful content, they question the platform's ability to deliver relevant content and as a result, heavy users' attitudes towards these platforms are negatively effected – they in turn question whether the platforms can shield them at all

Section 2

Perceptions of platforms

Users tend to state that the majority of their experiences are positive, but one negative experience can alter their perception of platforms

- **Majority of users are ‘appropriate’:** Users described their experiences of video sharing platforms as largely positive – they can view the content they are interested in, and heavier users and those with profiles are exposed to further content that is relevant to them. There was a consensus that the majority of those within a platform’s community abide by the guidelines
- **The challenges that platforms face:** Users also tend to recognize that platforms do attract people that abuse the guidelines. Users point to the ‘physical’ world as evidence – most people are law abiding, and it is the minority that break the law – the same can be applied to the ‘digital’ world. As a result, most users have a degree of empathy for platforms in attempting to tackle those that upload harmful or inappropriate content
 - Despite this, users state that one ‘bad experience’ can change their perception of the platform they are using. Many feel that the popularity of the platforms equates to large profits, and therefore struggle to understand why certain videos remain on platforms for so long, and why these resources aren’t being ‘better used’
 - This view is strengthened when users can identify multiple occasions where harmful or inappropriate content has been uploaded to a platform

Due to the perception that tools are not effective enough, users tend to believe that there is a lack of appropriate social etiquette

- **Physical world vs. digital world:** There is a perception amongst all types of users that how people correspond with each other in the ‘physical’ world is different to how that is done in the ‘digital’ world – i.e. users tend to believe that a minority of people can be more hostile when engaging with people on VSPs than they would be if face-to-face
- **Lack of effective tools is the driver:** Users tend to feel that this is the case due to a lack of effective tools on VSPs, ‘allowing’ people to ‘get away’ with inappropriate behavior. They tend to believe that if face-to-face, consequences would be harsher, thus acting as a deterrent. As a result, VSPs provide a degree of anonymity that ensures some people can act inappropriately without repercussions
 - Users state that this inappropriate behavior and lack of social etiquette is displayed through: offensive or combative comments and sharing of inappropriate content. Furthermore, some users tend to believe that some people do not take into consideration the audiences that may see this type of content (e.g. children) – again, they tend to believe that if face-to-face, people go through more of a ‘regulation’ of their words and comments if in the presence of children. When using VSPs however, users tend to believe that the same consideration is not given

Users tend to believe that platforms are operating in a reactive fashion, and are not doing enough to proactively remove content

- **Reactive, not proactive:** One of the biggest frustrations that users of all types have with platforms is the perception that they are operating in a reactive fashion regarding harmful or inappropriate content, as opposed to proactively identifying the content and attempting to stop it from being viewed by a significant amount of users
- **Long-term effects:** Many users highlighted the suicide that was live-streamed on TikTok (September 2020), and remained on the platform for 24 hours before being removed. Users believe that stories such as this highlight the platform's apparent reliance on the community to identify harmful content. The issue this exposes for our users however, is that people (including children) will have viewed it, potentially causing long-term negative effects on mental health
- **Therefore, the biggest question users had was: 'what will be done in the future to stop this happening again?'**
 - Currently, users do not feel confident in the platforms' ability to stop this type of content appearing in the future, and are frustrated at the length of time it can take for situations to be resolved

Users state that there is a lack of responsibility and accountability, and that too much damage is being done to platforms' communities

- **Lack of responsibility:** Ultimately, users tend to believe that platforms are not taking enough responsibility for the harmful and inappropriate content that can be uploaded, and point to the platforms' 'focus' on providing tools to the community as 'absolving' their responsibility – many users feel that platforms are trying to equip communities to *protect themselves*, as opposed to the platforms taking the lead
 - This view is strengthened when focusing on light users of platforms, who do not feel that they belong to a 'community' – they are less familiar with the tools available on these platforms and want the platforms to be taking a greater role in protecting users
 - Medium-heavy users appreciate the tools they have at their disposal, but agree with light users that too much emphasis is being placed on the 'community' to keep everyone else safe
- **Lack of accountability:** Users also tend to believe that there is little accountability for when harmful or inappropriate content is uploaded. There is a perception that platforms do not provide avenues for people to complain or seek help. This is worsened when accounts are anonymous, and the content is particularly violent or inappropriate – users are concerned about the long-term mental health implications and the perceived lack of support from platforms

Section 3

Appropriate measures for the future

Ultimately, all types of users want platforms to co-operate with each other in order to identify and remove content

- **Working together:** When harmful or inappropriate content is uploaded to a platform, one of the concerns that users have is that the piece of content can ‘multiply’ – despite it being tackled by one platform, the content can very quickly appear on *another platform* – meaning a whole new community of people can be exposed to it
 - This prolongs the damage being done to users – the piece of content almost becomes ‘mythical’, and users are aware that some people want to try and view it before all platforms become aware and remove it. Users can point to multiple instances where this has happened, and they perceive this to mean that there is a lack of dialogue and co-operation between the various platforms
 - Users tend to believe that if platforms were more ‘aligned’ and engaged in helping each other, they could ensure that the content does not ‘multiply’ and platforms can attempt to identify the content before it gets uploaded

Users tend to feel strongly that a combination of machine learning and human review is the best way to review content

- **Machine learning and human review is ideal:** Out of all of the platforms' measures, most users were particularly positive of YouTube's approach to identifying and removing harmful and inappropriate content, which can be seen below:
 - *“With hundreds of hours of new content uploaded to YouTube every minute, we use a combination of people and machine learning to detect problematic content at scale. Machine learning is well suited to detect patterns, which helps us to find content that is similar to other content that we've already removed, even before it's viewed.”*
 - Users tend to believe this approach is best due to the combination of both machine and human. For the former, users acknowledge the large scale of content that is uploaded, and feel generally confident that a machine could identify universally accepted harmful and inappropriate content, thus allowing humans to focus on more 'nuanced' pieces of content
 - Users refer to the term 'nuanced' when discussing content that may require context (e.g. a news story), and may require a detailed response from a human if found to violate guidelines

Most users tend to want upfront visibility around help and support on platforms for vulnerable audiences that may be exposed to content

- **Easier access to support:** One of the biggest areas in which users believe platforms are lacking in responsibility is the aftermath of harmful and inappropriate content being removed. Users have either personally been affected or know someone that has been affected as a result of viewing harmful or inappropriate content, and feel that there are little to no support avenues provided by platforms, despite being the ‘host’ of the content itself
 - Therefore, users want platforms to provide visibility of available help and support. Some believed that this could work on an ad-hoc basis (e.g. when a particular piece of content has been viewed by a significant number of people), whereas others believed that platforms should have a clear area for support on landing pages, so that users, and particularly those that identify as vulnerable, can easily access help and support
 - Some people pointed to other mediums of entertainment such as television, where if a piece of content is aired that is particularly sensitive or inappropriate, the channel provides viewers with various avenues to get information or support (e.g. websites, helplines) – users of video sharing platforms believe the same level of responsibility should be adopted

Users tend to feel that platforms ‘hide’ their guidelines behind menus and instead want regular reminders

- **Difficulty in recall of guidelines:** Many users could not recall specific guidelines outside of certain etiquette (e.g. being respectful, ensuring posts/content are legal), and believe their only exposure to a platform’s guidelines and policies is at the initial sign-up process and the requirement to agree to the platform’s rules
- **Mandatory reminders of guidelines:** As a result, users believe that platforms should provide mandatory reminders of its guidelines and measures to help encourage better behaviour and more appropriate content, thus potentially reducing the need for removing content. Users believed these reminders could be done in the following ways:
 - When uploading a new piece of content – users believe that when uploading new content, platforms could provide a reminder of the key rules and measures in place, and that upon uploading new content, they are agreeing to these
 - Quarterly reminders – users also believe that platforms could require its users to watch a short 2-3 minute video every quarter that details rules, measures and how to behave on platforms, as well as the help and support that the platform can provide

As some of the existing measures are perceived to lack effectiveness, users want to see platforms introduce greater authentication

- **Protecting younger audiences:** Users do not believe that platforms currently do enough to protect those aged under 18, with many platforms perceived to provide ‘easy’ methods of ‘faking’ an age. For example, most users felt that their children would be able to set up an account with a platform, and if required, state that they were born in a year that would allow them to view harmful or inappropriate content
- **Better authentication:** Users felt that again, this was a lack of responsibility on the platform’s part and currently, the process of entering a birth date is a ‘box-ticking’ exercise as opposed to showing genuine care for vulnerable audiences. As a result, users believe that platforms should provide greater methods of authentication to ensure that the user is genuinely who they say they are, including the following:
 - Identification required at sign-up – this measure was introduced by users as they believed people would feel more responsible for their actions with a formal identification process
 - Voice activation – introduced by parents as an additional parental control (e.g. if a child tries to access their account or harmful content – voice would be required for access)

Fundamentally, users want there to be greater accountability for what happens on these platforms and a central place to voice concerns

- **Lack of ‘central body’ for VSPs:** Users point to other channels that have ‘central’ places to raise concerns over content, including Ofcom for television and radio. They believe that this level of authority is not available for video sharing platforms, and as a result, they are perceived to be unaccountable to anyone or anything (other than themselves as a private company)
 - Users want this to change – they not only want to see platforms take greater steps to introduce measures that better protect people from harmful or inappropriate content, but when something does occur (and users expect this to be the case), they believe that there should be some type of governing body or organization that acts as both a catalyst for co-operation, but as a place for users to voice concerns or complaints about video sharing platforms

Section 4 Conclusions

Light, Medium and Heavy users are sympathetic to the challenges that video sharing platforms face, but ultimately, users do not believe that existing tools and measures go far enough to protect its communities – they agree that more responsibility must be taken by platforms to help shield people from harmful and inappropriate content. As a result, these users tend to believe six key measures could help achieve this ambition:

**Greater
co-operation**

Users want the leading platforms to work together to help combat emerging content that could be deemed as harmful or inappropriate

**Machines &
Humans**

Users believe the best way to identify and remove content is a combination of machine learning and human review, to ensure fairness and accuracy

**Help &
Support**

Users acknowledge that some content will still 'slip through the net' – in this event, users want platforms to provide greater access to help and support

**Regular
reminders**

Users do not believe that a one-time-only view of a platform's guidelines are enough, and want users to be periodically reminded of how to behave on these platforms

**Greater
authentication**

Users believe that there are not enough strong tools to stop younger audiences from accessing certain content, and want better authentication to prove who you are

**Governing
body**

Users feel that VSPs are not accountable, and look to other mediums for 'best practice' – they want a central place that will listen to their concerns