Chris Cowl
Y3908147

Q1. Validity

Construct validity

In this paper, the research question "Do behaviours known to relate to social anxiety in the physical world also manifest in gaming contexts?' is a broad question. After reading the paper further, it is apparent that the goal of the study is measuring participants' movement patterns in a digital environment to reveal if physical world behaviours in anxious people such as avoidance transmit to the digital world. The research question should be more defined, it's technically not wrong but if it were my paper I would write a more defined and informing research question such as 'Does avoidance, a known physical world behaviour, also manifest in people behaviours in digital gaming contexts?'. This would have been a lot clearer for the research community to instantly understand what the entire study was going to be investigating. Therefore, I'd say that the construct validity could be better on this one but it's not wrong either, it could just be more transparent.

Previous research suggesting links between anxiety in both the real world and gaming world

The construct validity is good because the researchers draw upon existing research (Mandryk & Birk, 2017) who recommend that placing participants in a game situation and monitoring their responses can elicit behavioural responses that are linked to mental health. The researchers also build on existing studies which champion the link between anxious behaviour in the real world and anxious behaviour displayed in the digital world. Other studies noted that real world biases, in respect to the way players interact with one another, also manifest in a gaming situation between player avatars (Yee & Bailenson, 2007). Also, other studies by (Kang & Gratch, 2010) describe how socially anxious people respond to digital characters and how avatar-based interactions relate to social anxiety in a virtual reality, non-gaming context (Lange & Pauli, 2019). Finally, further research (Dechant et al., 2020) shows social anxiety is achieved in-game in a massively multiplayer online role player game player sample.

The research therefore is valid in terms of construct validity as previous research suggests behaviours associated with social anxiety in the real world can manifest in gaming contexts (and also in virtual reality) and can therefore be explored further in this study.

Measurements

To further validate the study in relation to construct validity, The Liebowitz Social Anxiety Scale (LSAS) (Liebowitz, 1987), was used to measure anxiety after the task.

This is used, amongst other techniques, by health professionals to measure anxiety in the real world so this is a suitable way to measure anxiety here too. Although it has to be noted that this was just one method of measurement of anxiety, to further validate the findings, another method of measurement could have been used.

Manipulations

Test manipulations (Biomarkers)
Based on previous literature, if the participant is anxious during the task, they will respond to the following test measures in various ways.

- **Target error** – Participants are expected to rush the task and therefore accuracy will be reduced (Mullins & Duke 2004).
- **Time in the room** – Participants again will rush through the task and spend less time in the room (Mullins & Duke 2004).
- **Kurtosis** – A higher kurtosis is expected (fewer outliers being far from or close to non-player characters) (Mullins & Duke 2004).
- **Final distance to non-player character** – Participants are expected to have a higher final distance to non-player character because of avoidance behaviours (Hartanto et al., 2014).
- **Minimum distance to non-player character** – Participants are expected to have a higher minimum distance to non-player character because of avoidance behaviours (Hartanto et al., 2014).
- **Mean distance to non-player character** – Participants are expected to have a higher mean distance to non-player character because of avoidance behaviours (Hartanto et al., 2014)..
- **Skew** – Participants are expected to have a higher skew because of avoidance behaviours (Hartanto et al., 2014).
- **Path length** – This could be short (in order to get the task done quickly) or long (to avoid the non-player character) (Hartanto et al., 2014).

Personalised Avatars, Social Awareness and Social Presence
Increased social awareness occurs if the participants' avatar represents themselves closely (Vasalou et al., 2007) and therefore it was relevant to have a personalised avatar manipulation in the experiment.
Also, social anxiety is likely to be increased if a participants avatar closely resembles the participant in real life (Aymerich-Franch et al., 2014).

Dechant et al. enabled the participant to conduct the task from a third-person perspective, a first-person perspective and with a mirror at the player start point for both perspectives. This was to increase the participants' self awareness and therefore the feeling of social presence. The participants were also asked to answer the Player Identification Scale (PIS) which was backed up by previous research (Van Looy et al., 2012), to see how close both the custom avatar and predefined avatars were identifiable with themselves.

Participants were also asked to create the personality of their avatars, to increase the participants identification with their avatar by adjusting five 7-point Likert scales with each one describing one personality trait. This was based on the 10-item short version of the Big Five Inventory (BFI-10) (Rammstedt et al., 2007).

The above are therefore all suitable choices and increase the construct validity, as they are backed up by previous research in the domain.

Conclusion

The above points are suitable ways to explain and measure the concept of avoidance in a digital game and the concept of real world avoidance traits manifesting in a digital game. Although the research question isn't as defined as it should be, it is also not incorrect either. Therefore, the experiment is suitable to answer the research question (although a more defined research question would have been better) and to inform which biomarkers can be used to measure which biomarkers can be embedded in digital games for the use of anxiety assessment in the health sector, but only in relation to movement and avoidance. The manipulations fit with the research question and goals of the study as they have been informed by previous research. Potential biomarkers have been identified and so tangible results can be used to aid the designing of game based anxiety assessments.

Maybe the researchers could have used another method of measuring anxiety as well as The Liebowitz Social Anxiety Scale to confirm anxiety levels as this would only strengthen the research findings.

Because the researchers have used knowledge from previous studies to inform their choices of test manipulations, measurements and to define the concepts. All things considered the construct validity is high.

Internal validity

There are a few areas which could have had an effect on the results of the study, both in positive and negative ways.

Confound minimisation

Dechant et al. reduced confounds by using a custom online task instead of an off the shelf game. This was a wise choice in terms of internal validity as it meant they were in control of the study in this aspect. They also gave all participants enough time to configure their avatars which was crucial to gaining social presence. People who needed to adjust their avatar to more closely match their real world aesthetics were able to. It would be unfair if some people in the real world had

more complicated looks, and therefore took longer to recreate them in their avatar. This could therefore affect results.

Potential confounds

Controlling for time
The researchers don't control for time in the study, or they don't mention it so we won't know for sure. The gaming tasks could have been easily set up to be at 1pm on any given day or even controlled as much as 1pm on a Wednesday if so desired so all participants were assessed during a consistent window of time in relation to the week. So this potentially affects the study outcomes in a negative way by adding an unwanted variable and therefore is a confound.

Participants behaviour change
Participants could act out of character when playing in the digital world, they can be whoever they want to be with little or no consequence to themselves or others (Yee & Bailenson, 2007), I think it's reasonable to accept this is part and parcel of research limitations. In a perfect world the researchers could have screened for participants who are very likely to reflect their real world anxiety through online play but it's reasonable to assume that some participants may change their behaviour and some won't. This would be equally true if they were playing a custom game designed to assess anxiety. Nevertheless, this could be a confound.

Recruiting participants
Previous research has suggested the MTurk method of recruiting participants is a reliable research tool (Mason & Suri, 2012). Because participants were screened for anxiety before the experiment we know we're testing participants susceptible to anxiety. Unfortunately, the participants were mainly made up of males approximately 29–45 which is not the demographic where anxiety in the real world is most prevalent.

Measurement of anxiety
There is always going to be a risk of social response bias when asking participants to fill in a survey or form asking for responses. They could easily tell researchers what they think they want to hear or what they think their friends may think is ok. The Liebowitz Social Anxiety Scale is used by medical professionals as a tool to measure anxiety in the real world, this is only one way to measure anxiety. This is noted in the limitations section but I feel that this was probably due to constraints of time or funding.

Player equipment
As participants were playing from the comfort of their own homes, which adds ecological validity, it also means all participants are using different specifications and standards of monitor, computer, keyboard, mouse and importantly, broadband connection.

Conclusion

The above points do show a few confounds but as with most studies we have to be flexible in our judgement as we weren't the ones doing the research and I guess some information could be lost in the writing up of the paper. The minimisation of confounds was attempted well here so that adds to internal validity.

Although we can't say for sure or not, I would have liked to see a control for time in the experiment. Maybe there was a control for time but it wasn't mentioned in the paper. This is obviously a problem because participants could execute the task when they are either already stressed or very relaxed which we can therefore not account for in the results. This is a confound if time was not controlled for.

Participants' behaviour change and the measurements of anxiety issues – I think we can assume fairly that this is acceptable due to real life compromises in any research that is undertaken. This could be accounted for with the sample size being large enough which it is so I don't think this is a confound but something to be aware of.

The confound of recruitment however, is a problem. There were too many males in relation to females and non-binary individuals and the ages aren't spread evenly enough showing a high standard deviation. Although the participants are randomly assigned to condition groups, and assuming they are checked and the ratio of gender is equal in all condition groups, the experiment is still experimenting on more males than anyone else and therefore not reflecting the general population. The research question and goal of the study is not specific in terms of the variables of age and gender, for example, if the research question was 'Do behaviours of men aged 25-45, known to relate to social anxiety in the physical world also manifest in gaming contexts?' then there would be no confounds. But as it is the research question is non specific on these variables of age and gender and so this is a confound because we are assuming that the research is to be generalised to the general population. This could have been addressed quite easily by either just taking a number of males out and accepting a smaller sample or waiting to recruit more females and non-binary participants to increase external validity and be able to generalise the results more. The study could have also focused on males aged 25-45. I think this is still a confound even if gender ratios are the same between condition groups.

For the measurement of anxiety, I think this is probably a reasonable compromise to just use one well used, validated clinician administered scale. Therefore this is not a confound.

Player equipment, because participants were conducting the tasks at home, it was impossible to know what specifications and speeds of computer equipment

and broadband they were using. Some participants could have an advantage with better broadband or bigger monitors for example. But hopefully there is enough natural variation in the sample for this to even out.

Overall, the internal validity is ok but should be better. I think the participant recruitment issue which has been mentioned previously, means they're not testing a fair representation of the population and therefore reduces the ability to generalise the findings which lets down the data, study and results. Ultimately the research question has not been tested as thoroughly as it should have been. However, the results still illuminate the knowledge base further and can be used to inspire other researchers to do better.

External validity

Dechant et al. make reasonable claims and generalisations in relation to the findings of this study:

Known real world anxious behaviours transferred to gaming
The paper only claims to show that their findings seem to back up known real world anxious behaviours transferring to the gaming world such as avoidance of the non-player character. It is noted by the researchers that they can't be sure if it is because of an estimation bias or an explicit safety behaviour.

Increased self awareness and social presence
They also only claim to show that increased self awareness and social presence has been shown to increase anxiety in this study. The researchers show that by the use of a customisable avatar and a mirror in the task, the players can see themselves and relate with their character and therefore this explains greater anxiety shown in relation to this study which also backs up previous work. They also generalise that when the participant uses a custom avatar in the third-person perspective, the participant feels greater anxiety due to the fact they are performing in front of the non-player character. However they are quick to state this is only a possible explanation and further research is needed.

Desktop computer
It is also concluded that they have been able to show that anxiety can be shown via just the use of an ordinary desk-top computer. This is accessible by the masses with no need for expensive equipment. Which has potential for the design of custom anxiety assessment games.

Game design
The researchers only recommend the effect of the biomarkers in this study be used as considerations for the custom game design process. They acknowledge they don't know exactly how off the shelf games would be used to measure social

anxiety because basically, the games are developed for other purposes such as enjoyment for example. They argue that custom games are of more use insofar as having the control to implement features such as third-person perspectives and custom avatars along with other means to elicit and measure anxiety.

Participant sample
Due to the issues mentioned in the internal validity section, the study findings cannot be generalised far beyond the effect of the tests on males ages 25-45.

Conclusion

The researchers use inferential statistics to generalise to the wider population showing, with a confidence interval at least 95% ($p < 0.05$, $p < 0.01$ and $p < 0.001$) that the results produced from the experiment can be generalised to the wider population. Via statistical analysis (MANOVA) the researchers account for variables such as gender, similarity identification (players being able to represent themselves effectively with the avatar customiser or not), side, distance, and type of non-player character presence, The Liebowitz Social Anxiety Scale scores and effect of third or first person perspectives. This adds external validity because they have done a lot to account for variables in the analysis.

Therefore, Dechant et al. can make reasonable generalisations to the possibilities of using the biomarkers identified in the study to develop custom anxiety assessment games. They generalise that because they have shown real world anxiety behaviours can manifest in a gaming situation, the use of customised avatars coupled with a third-person perspective increases anxiety, the use of inexpensive, everyday desktop computers can be used to elicit anxiety, that there is great potential for game designers to create custom, anxiety assessing games in order to compliment other means of clinical assessment of an individual.

They are also reflexive in the limitations and state more research is needed in regard to non-custom games and that the task in this study only creates a simplistic gaming environment compared to off the shelf games. They champion more research to understand the potential of how more richly designed off the shelf games could help assessing anxiety.

All these generalisations are fair but only within the demographic of males aged 25-45 however. A further study with equal gender and age groups would enable the researchers to generalise the results from 20-40s, 41-60 and 61 plus to get a real overview of the effects of the experiment on the manifestation of real world anxiety behaviours in a digital game. External validity is overall quite good, because of the insights above they have uncovered with the experiment and generalised to a wider population (of 25-45 males only) through inferential statistics, but let down by the fact they can only generalise to a small section of the population, therefore external validity could and should be better.

Ecological validity

The study is not really valid in terms of ecological validity, reasons for this are:

Realism
The test is done using a bespoke online game, whilst this gives the researchers control and reduces confounds, which is very important and a good way to work, in turn it also loses its ecological validity because it looks like a game from 15 years ago. This may have reduced the participants' behavioural reactions to the non-player character due to its lack of realism and therefore it tells us that whilst we can make the link between anxious behaviours in the real world being manifested in a game environment, we don't know by how much. Would more realistic custom games elicit these behaviours more, less or about the same? Would commercial games have more or less of an impact on the elicitation of anxiety? We won't know until a suitable study is done. But still, obviously these are worthy studies to conduct nonetheless.

Simple and repetitious
The task itself was simple and repetitious, there could be an argument that by say the twelfth go, that a participant may become desensitised to the task and run on autopilot. We know it wouldn't happen in the real world otherwise it would be very easy to 'get over' social anxiety by just trying to socialise repeatedly in a short space of time.

Non-player character behaviour
The non-player character also does not move around or speak – it's unlikely that other people in a social situation will remain still, stand in an odd place, face away looking at a wall or stare angrily at you but not speak to you. There is a comedic element to the way the non-player character behaves which again, reduces realism. This could reduce the feelings of anxiety of the participants.

Player environment
The players played the gaming task at home so this adds a little ecological validity to the study as players are being tested in the comfort of their homes, as they would if they were using a digital game as therapy or assessment.

Conclusion

Whilst the ecological validity is increased by players conducting the tasks at home reflecting in the wild nature of therapy or assessment, and while the study is a worthwhile addition to knowledge in the domain, for the above reasons it's definitely not realistic enough to reflect 'in the wild behaviour.' It's just too simplistic and unrealistic.

(Q2. on a new page).

Q2. Creepiness of Technology

In the paper (Woźniak et al., 2021) define a scale for measuring the creepiness of technology. This was done by the use of focus groups to attempt to outline what characteristics technology has to display in order to be considered creepy. I will look at the procedure, participant selection, quality of data and the reporting of the findings in the study in order to assess the suitability of the use of focus groups to answer the research question.

Procedure

The procedure of using a literature review then 10 focus groups to collect data was suggested by previous research (Boateng et al., 2018), because of this, Woźniak et al. started off with a solid base on which to build their own research. By using stimuli to encourage debate and using a semi-structured approach during the focus groups, the researchers made another good choice, increasing their chances of eliciting interesting and meaningful insights on the topic of creepiness. Participants can then comfortably get involved as there is something to really talk about. So far so good in terms of procedure.

Next is the use of the ladder interview approach where the moderator keeps asking "why?" to a participant in order to get deeper meanings about thoughts or perceptions of the technology in question. This potentially digs out real quality data.

During the focus groups, the moderators explored the use of adjectives associated with the feelings of creepiness, for instance, creepy, threatening, frightening, unpleasant and strange. The adjectives were taken from and therefore backed up by a previous research (Langer & König, 2018), (McAndrew & Koehnke 2016), (Ni et al., 2011), (Watt et al., 2017) and (Yip et al., 2019).

Up to this point the researchers made good decisions but in my opinion, there is an issue with the focus group duration. The focus groups appear to be too short. In the first group the exploratory phase is 25 minutes to explore four technology products leaving only 20 minutes afterwards for the follow-up part of the group session. 45 minutes in total may not be enough for participants to get into it and bounce thoughts off each other and fully develop their opinions on the task at hand. It could take participants most of the 20 minutes to warm up and feel more confident to interact fully in the session. The subsequent eight groups only had 10 minutes to explore the chosen EMG-based device and a further 20 minutes to participate in the focus group. Again, 30 minutes in total is somewhat short. As Braun & Clark (2013) suggest "The discussion should take approximately one hour, but may go a little longer." (p.127).

I feel the focus group sizes were much too small also. There were only two groups of four participants and eight groups of three participants. Maybe you could get away with it in the first two groups with there being four participants, but three participants in the remaining eight groups just isn't enough, coupled with the sessions not being long enough, the researchers really missed out in creating an environment to elicit far more high quality data. I think that reducing the number of groups but increasing the size of the groups would have given the participants far more opportunity to debate and bounce off each other.

Participant selection

The researchers included a fairly evenly spread sample group when looking at gender. Across the ten groups it is very evenly split at 15 women and 17 men. The sample is weighted towards the younger demographic however with 20 of the 32 participants (62.5%) being in the 24-31 age range. There is then a huge age gap of around 42 years minimum where the remaining 12 participants (37.5%) are in the 73-79 age range. Because there are no participants from the 32-72 age range, this omits a rich tapestry of people with their opinions and observations about the creepiness of technology which is a big loss in my opinion.

The sample is also skewed in favour of highly educated professionals where 26 of the 32 participants (81%) had either a bachelor's degree or masters. This seems a bit like 'othering' in the sense that only educated people could possibly understand the feelings of creepiness toward an artifact such as those tested. I'm sure it isn't intentional and both points above could be because of some kind of constraint but the sample is still skewed and is almost sure to affect the outcome of the research, although we'll never know for sure exactly by how much.

We also don't know much about the participants except for gender, mean age and education levels. We don't know any background information that could be related to the study such as whether the participants are used to buying, using or being around new technologies similar to those being used as stimuli. This would have been very interesting to know would have potentially had an effect on the study.

I think participant selection could have therefore been much better when you consider that these participants are where you get your data from.

Quality of data

As I have mentioned in the procedure section, the ladder interview approach was used to find deeper meanings to initial comments or perceptions. This is a fantastic way to gain rich data which therefore massively improves the quality of data.

The method of collecting the data is also sound as suggested by previous research (Boateng et al., 2018), 10 focus groups, and looking at previous literature are great ways to collect, justify and guide you through the data collection stage of a study. So in theory the researchers are on the right track with their decisions at this point.

Unfortunately I think a couple of aspects of the study have had a negative impact on data quality. As noted in the procedures section, due to the sizes of focus groups and the shortness of these group sessions, the researchers could and should have had far more rich data from participants. Coupled with the fact that the participant selection doesn't reflect enough of a broad demographic of people I think that the quality of data could have also been much higher. It seems that the scale could be a scale of the experiences of the twenty and seventy somethings rather than a fair representation of the population.

In conclusion I'd say these negative issues let the study down somewhat and sadly, affect the quality of the Perceived Creepiness of Technology Scale.

Reporting of findings

The dimensions that the researchers defined from the combined literature review and the data gathered from the focus groups is indifferent and doesn't seem the strongest part of the research. The three dimensions that shaped the Perceived Creepiness of Technology Scale were implied malice, undesirability and unpredictability.

Implied malice
This dimension is inspired by creepiness in privacy research and interactions with autonomous systems which is related but Woźniak et al. do not reference any of the research so it's hard to comment on how solid an area this was to focus on. This seems to inform the generalisation that creepiness is a violation of privacy which again, is hard to know whether it is worthy of shaping a dimension that the Perceived Creepiness of Technology Scale is based upon.

Undesirability
The researchers drew upon previous research (McAndrew et al., 2016) to guide them in the focus groups. This research mainly focuses on whether men are more likely to be creepy therefore more likely to pose a sexual threat to women, what professions elicit the feeling of creepiness and that unpredictability of human behaviour is creepy. The researchers state that they adapted the outcomes of the study to inform the focus groups and therefore this dimension, but I can't see a strong link in that respect. If anything the research (McAndrew et al., 2016) makes more of a case for unpredictability rather than undesirability. Despite this,

undesirability seems a logical dimension, which is backed up by the comments from P7 and P9, but it could be more solid if supported by previous work a little more clearly.

Unpredictability
I feel this is the strongest dimension in terms of how previous research guided the focus groups. Lack of control was found to be associated with threat (Paredes et al., 2016), not being able to understand a device and perhaps what it does leads to a feeling of being out of control which is backed up by further research (Yip et al., 2019) and finally, not knowing what an object does can create a feeling of creepiness (Oozu et al., 2017). The knowledge from these papers is a great way to help guide the focus groups and when coupled with the use of the Oxford English Dictionary seems a really sound dimension to inform the design of the Perceived Creepiness of Technology Scale.

Conclusion

Overall I think the procedure was, in theory, a good one. The combination of using a literature review and 10 focus groups to collect data was suggested by previous research (Boateng et al., 2018) and so had been used before successfully.

Subsequently, the use of stimuli in the focus groups, a semi-structured flexible approach and a ladder interview technique also set the researchers on a good path. Unfortunately, the short duration of the focus groups and small numbers within these groups potentially affected the quality of data that was created from the sessions. The data could have been more in depth and there would have been more of it, potentially changing the outcome of the Perceived Creepiness of Technology Scale.

The gap in age range and the fact most participants were highly educated also had a negative effect on the quality of data produced by the focus groups and therefore the outcome of the study.

The reporting of the findings is also a bit disappointing. I have the benefit of hindsight, but the three dimensions that inform the final scale are overall not as good as they should be, unpredictability being the only dimension that appears to be the most robust with undesirability maybe counting as a half dimension due to the lack of a strong link to previous research. Implied malice seems like it was generally made up as no research was stated.

All in all, in terms of focus groups, much of the design of the study is suitable to inform the Perceived Creepiness of Technology Scale, but I think the execution in too many areas of the study, mainly the quality of the data (so the part that everything is built from), isn't nearly as good as it could have been.

Ethical considerations

During the evaluation section of the paper by Woźniak et al., the researchers use images of one participant twice who can clearly be identified by full side on images. This is of course an ethical consideration of anonymity and should have been picked up on. The fact that it's an actual face of a person is careless. There's no point in being anonymous as far as data gathering is concerned, if your face is on the study paper. As the head of the participant in the bottom image needed to be visible to show the EEG device, the head could not be cropped out. It could have been easily blurred out or an illustration could have been used which would have illustrated the methods equally well.

I think the second ethical consideration is the amount that the participants of the survey were paid. Yes, the actual survey only took 3-4 minutes (not 5 minutes after the remuneration) but $0.80 is still a derisory amount even if it has then been adjusted to work out at $14 an hour. The time the participants probably took out of their lives probably wasn't considered in the equation either. There will be a small amount of time taken during the process of being accepted in the study. Even though it was probably only small it will have been equivalent at the very least to the 3-4 minutes the study took. Considering that the participants are supposed to be valued in the part they are playing to evaluate the researcher's work, and their contribution to the knowledge base, it seems like they are being taken advantage of. Then you have to question how much thought went into the answers perhaps, as $0.80 is so low it doesn't seem worth the effort to worry about the quality of answers given. It may be that because most of the participants were highly educated that they were motivated to participate to do something worthy but they still need compensating fairly. Another way to have approached this is to give participants a flat fee of around $5 or $6 so the hourly rate is high but it accounts for time spent in the process of participant acceptance. Then the participants would probably be more likely to participate more fully than if they felt otherwise.

(Q3. on a new page).

Q3. Exergames

The experiment involved 500 participants split into three groups, which were allocated to three condition groups. The groups were exposed to an exergame for a month, a regular game for a month and a placebo group who did nothing for a month. At the end of the month all participants were asked to run the 100m and their times were recorded.

It was hypothesised that participants' exposure to the exergame will have a positive effect on their times when running 100 metres, therefore the null hypothesis (H_0) is that there will be no effect or relationship between the exergame condition group and the time participants took to run the 100m.

Distribution of the data sample

If we look at the distribution of the times across the whole dataset (Fig 3.1), we can clearly see that the data has a unimodal positively skewed distribution. This tells us that something within the data is affecting the central tendencies. One of the variables is skewing the data but at this point we don't exactly know which variable that is.
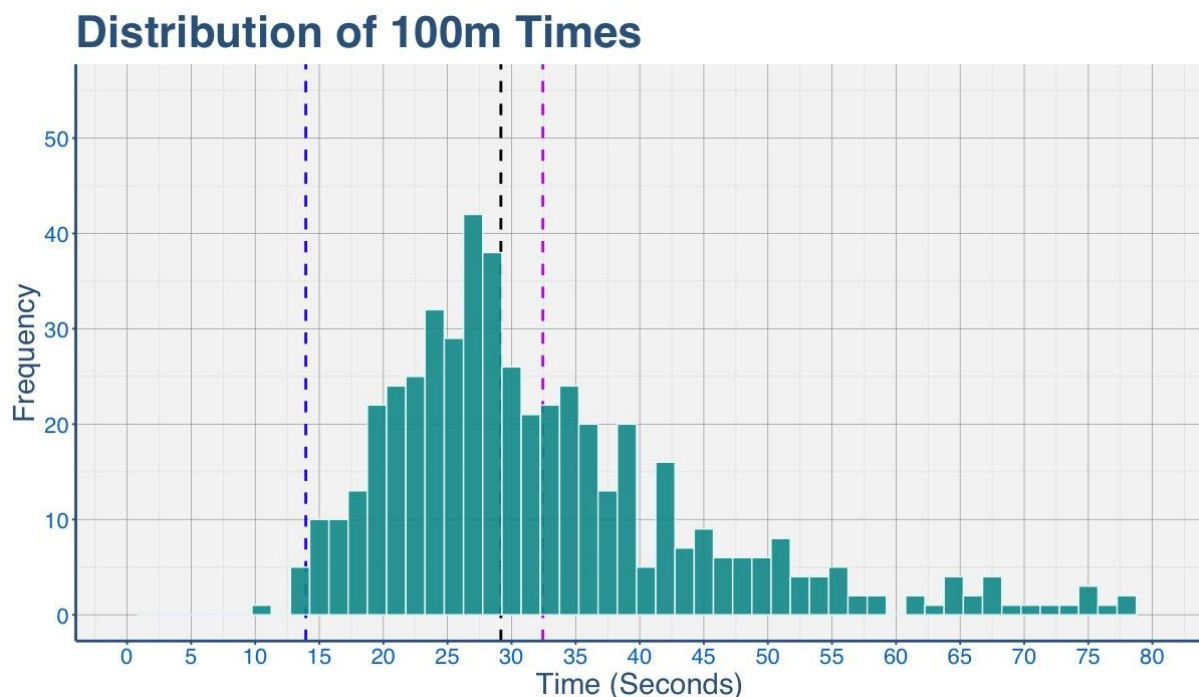

Fig. 3.1

From Table 3.1 we can see that the mean time run is 32.44 secs and the median is 29.17 secs. The range is quite large at 67.92 secs indicating a difference of 67.92 secs between the quickest and slowest times run. What is also interesting is just

from looking at Table 3.1 alone, we could have determined a positively skewed distribution because the mean is greater than the median across the whole dataset.

Because the distribution is not symmetrical, this is one reason that we should apply a non-parametric test to the data, therefore find out where the significance is.

Distribution of Times						
	Central Tendency		Spread			N
	Mean	Median	Range	IQR	SD	
Total	32.44	29.17	67.92	13.94	12.54	500

Table 3.1

Other variables affecting distribution

Relationship between age and time

We can look at how age affects participant run times (Fig. 3.2). As you would expect, the variable of age affects how quickly the participants ran the 100m. The scatterplot shows a trend that the younger the participants were, the faster their times were. This is shown by the cluster towards the bottom left of the graph and the dissipation to the top right of the graph.
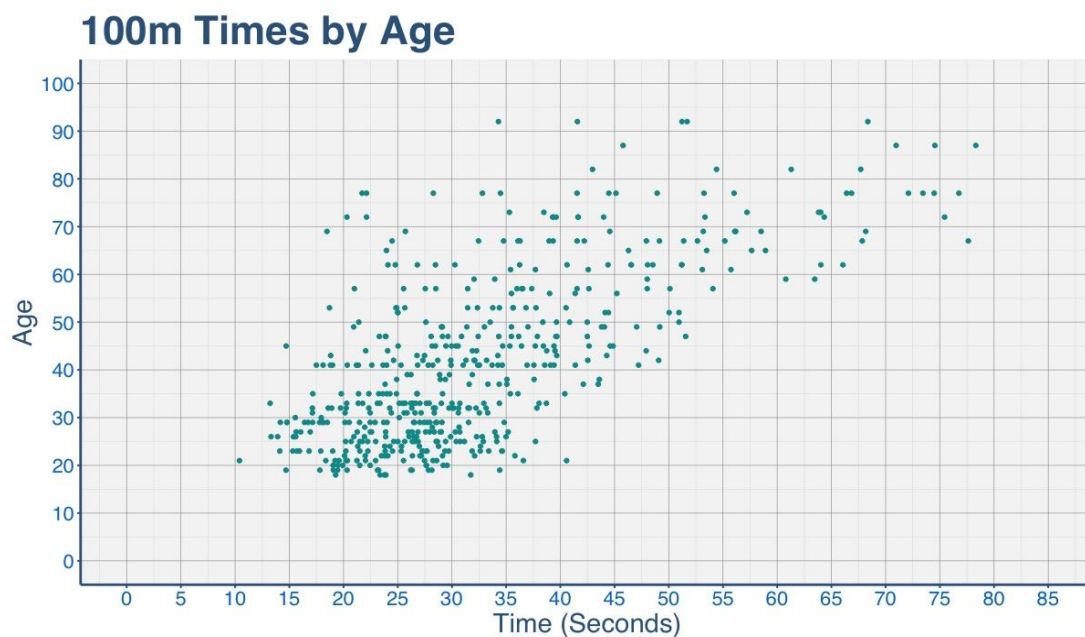

Fig. 3.2

Relationship between gender and time

We can see that gender doesn't drastically affect the time taken to complete the 100m (Fig.3.3) and times are all pretty evenly spread according to gender.
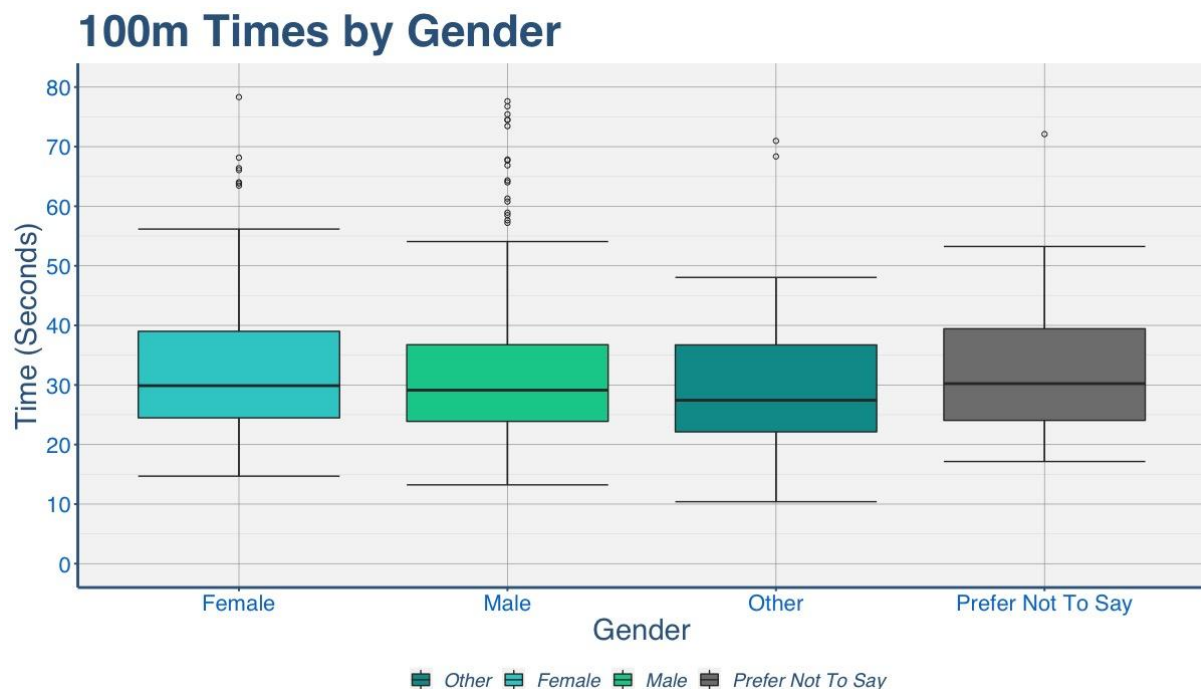

Fig. 3.3

Analysing the effect of the experiment

To test the null hypothesis (H_0) and therefore the effect of the condition groups on the time the participants took to run the 100m, we have to take into a account a few things first in terms of the design of the experiment:

1. A non-parametric test is to be used because as shown earlier (Fig. 3.1) the distribution of the data sample is non symmetrical (not normally distributed) and therefore the central tendencies are also skewed.

2. We have three condition groups with participants being exposed to only one condition group each, therefore a between subjects design of a Kruskal-Wallis Test is to be used inorder to find if a condition group is affecting the times run by participants. The significance level is set at $p < 0.05$ therefore the probability that the effect of any of the condition groups is the product of pure chance, only 5% of the time.

3. If a significant result is shown, a Dunn's Test with post-hoc pairwise comparisons with Bonferroni corrections is run afterwards. This is to find out which condition group is having the most effect. Because we are running multiple comparisons we have to account for the increased chance of attaining a false positive which may push the p-value over $p < 0.05$ causing a type II error. Our significance level for the Dunn's Test has to

be adjusted to p-value < 0.017 to avoid insignificant results being misinterpreted as false positives

Results

A box plot showing the relationship between time run and condition group is shown in Fig. 3.4. Table 3.2 also shows central tendencies and spreads for each variable.

Looking at the central tendencies and spread in Table 3.2, we can see clearly that the condition group that ran the 100m in the fastest times and therefore performed better, are the exergame condition group. The mean and median are both well under the mean and median for the entire dataset. The spread, interquartile range and standard deviation of the exergame group is also much smaller than the other groups and therefore describes a tightly clustered group in terms of consistent times recorded. The other groups have times recorded in a more spread manner from the mean. It appears that the exergame is having a consistent effect on times recorded with them being tightly clustered and also faster according to the mean.

Fig. 3.4 confirms the exergame condition group performed better against the variable of time. The other two condition groups are quite similar, although the placebo group (mean = 35.93, median = 33.36, SD = 13.00) seemed to slightly outperform the regular game group (mean = 36.83, median = 34.91, SD = 13.48) also with a similar standard deviation .
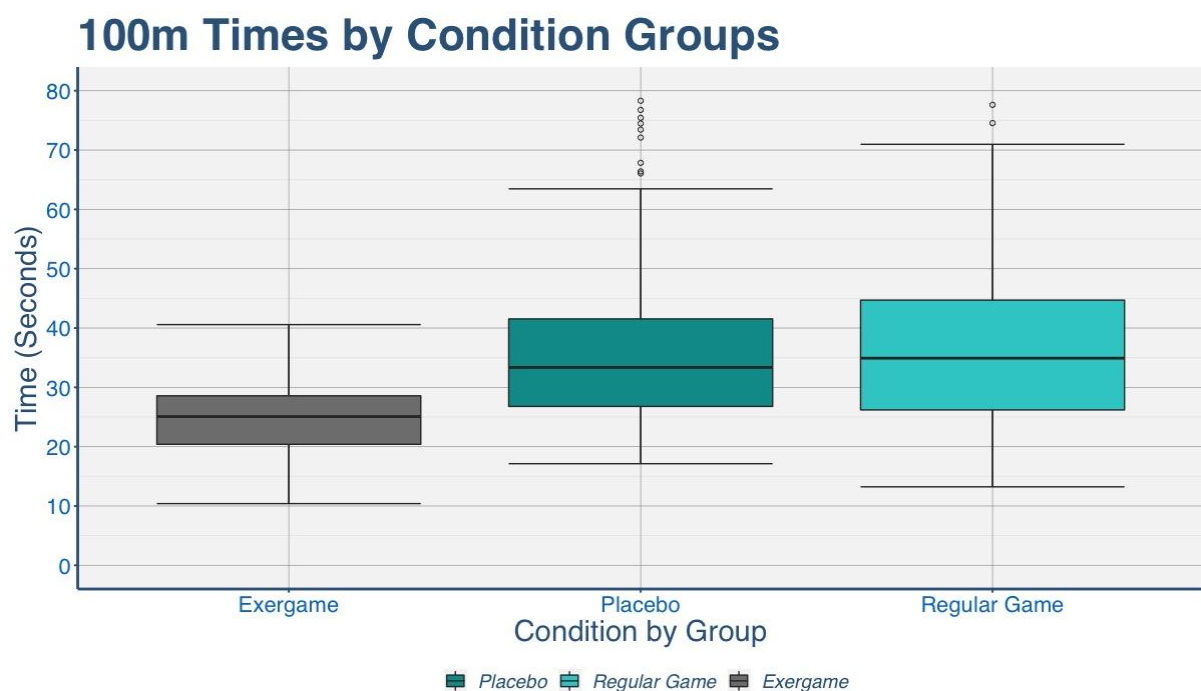

Fig. 3.4

100m times by condition group						
	Central tendency		Spread			N
	Mean	Median	Range	IQR	SD	
Exergame	24.78	25.08	30.17	8.18	5.80	170
Placebo	35.93	33.36	61.19	14.76	13.00	162
Regular game	36.83	34.91	64.40	18.51	13.48	168
Total	32.44	29.17	67.92	13.94	12.54	500

Table 3.2

The result of the Kruskal-Wallis Test was highly significant as the p-value < 0.001 (2.2^{e-16}), so safely below the p-value < 0.05 we set. To find out where the significance was, a Dunn's Test post-hoc pairwise comparisons with Bonferroni corrections was applied.

Because we applied Dunn's Test post-hoc pairwise comparisons with Bonferroni corrections, the adjusted significant p-value < 0.017 to avoid insignificant results being misinterpreted as false positives due to increased likelihood of gaining a significant result purely by chance. The Dunn's Test post-hoc pairwise comparisons with Bonferroni corrections shows us that the comparison of exergame and placebo condition groups was highly significant with adjusted p-value < 0.001 (1.68^{e-18}). The difference between exergame and regular game condition groups was also highly significant with adjusted p-value < 0.001 (2.98^{e-20}). The final comparison group of placebo and regular game with adjusted p-value = 1.000 which was above the adjusted p-value < 0.017 showing no significant difference between condition groups.

Pairwise comparison groups against time	p-value
Exergame vs. placebo	< 0.001
Exergame vs. regular game	< 0.001
Placebo vs. regular game	1.000

Table 3.3

At this stage I would have to reject the null hypothesis (H_0) and agree with the hypothesis that exposure to the exergame will make participants faster at

running 100 metres but we need to fully explore the effects of the other variables in the experiment.

Alternative explanations for results

As mentioned previously in Fig. 3.2, age has a big impact on participant times recorded. This is something that should be cross examined before we conclude the analysis of the experiment.

In order to check for any confounding variables in the condition groups, I plotted age against both gender and condition, and tabulated condition and gender.

Fig. 3.5 shows no obvious trends between age and gender with central tendencies and spreads quite similar.

Fig. 3.5

Table 3.4 shows the distribution of gender within condition groups, highlighting a very similar distribution of any particular gender across all of the condition groups therefore meaning if there was a relationship between gender and participants times run, it would affect each condition group fairly consistently.

Condition and gender				
Condition	Female	Male	Other	Prefer not to say
Exergame	45	97	15	13
Placebo	50	85	14	13
Regular game	58	92	10	8

Table 3.4

However, Fig. 3.6 clearly demonstrates that the exergame condition group is confounded with younger participants, to the extent that the whiskers on the boxplot do not overlap with the other two condition groups.
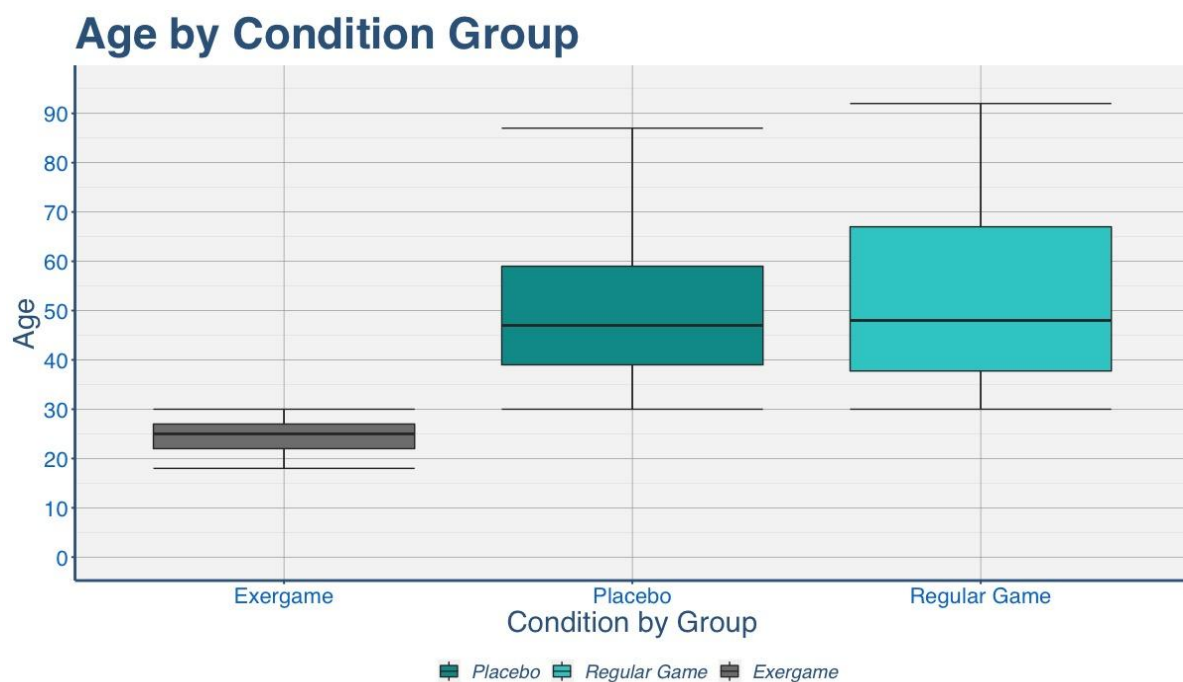

Fig. 3.6

Table 3.5 shows the median of the exergame condition group is 25 and the medians for the regular game and placebo condition groups are 48 and 47 respectively. The standard deviation for the exergame condition group is distributed tightly that almost all participants are approximately in the 21-28 age range. The standard deviation for the regular game condition group is widely spread approximately ranging from 35-69 with participants even in their nineties indicated by the whiskers in Fig. 3.6. The statistics from the placebo condition group indicate the standard deviation being spread from 35-64 with participants even in their mid to late eighties.

Clearly the exergame condition group is highly advantaged in relation to the experiment. They are consistently three times younger than some of those in the other two condition groups indicated by the small standard deviation and lower mean age. When looking at the medians, they are also approximately half as young as the medians in the other two condition groups.

Age by condition group						
	CenTendetral ncy		Spread			N
	Mean	Median	Range	IQR	SD	
Exergame	24.29	25.00	12.00	5.00	3.31	170
Placebo	49.48	47.00	57.00	20.00	13.90	162
Regular game	52.06	48.00	62.00	29.25	16.92	168
Total	41.78	38.00	74.00	26.00	17.91	500

Table 3.5

Interpretation of analysis

The experiment set out to test the hypothesis that exposure to the exergame will make participants faster at running 100 metres. However, age was confounded with the condition groups with younger people overly represented in the exergame condition group.

The researchers failed to design the experiment well enough, if they had ensured that each condition group had evenly distributed participants in terms of age, as they did with gender, then the hypothesis would have been tested thoroughly. The experiment was therefore a failure in relation to the original goal.

We cannot separate the effect of either the exergame condition group or age on the times run by participants. In conclusion, we cannot reject the null hypothesis (H_0) that there was no effect or relationship between the condition groups on the time it took for participants to run the 100m.

Exergames may have an effect on the times participants take to run the 100m. However, due to the execution of the experiment the original hypothesis has not been tested.

(Q4. on a new page).

Q4. Content Analysis

Development of the Coding System

The content analysis that I undertook came from the viewer comments and opinions of the video "Game Therapy - How Can Games Improve Mental Health?" by popular gaming channel Extra Credits.

I coded the viewers comments by using inductive coding taken from the viewers comments themselves. The first pass of the data resulted in 100 codes emerging from the data, this was increased to 107 on the second pass, then drastically reduced to 21 codes on the third pass and finally refined to 11 categories with 13 subcategories on the fourth pass. The code dictionary below (Table 4.1) was created from the final 11 categories and 13 subcategories as below.

CODE	N	DESCRIPTION	SAMPLE
SYMPTOM OF A REAL LIFE PROBLEM	**33**		
Gaming Disorder is a Symptom Not a Cause	17	Gaming disorder is a symptom of real life problems not a cause. The condition is misunderstood. Terrible life events can cause gaming addiction and disorder. We have to find the root cause the same as with any other illness.	"You are pretty much right, an addiction and/or obsession are symptoms of a deeper issue, and it doesn't have to be games, any form of escapism could do." (Ivarent)
Gaming to Avoid Problems is Bad	16	Gaming to avoid real world problems is bad because real life always catches up with you. It is what you are escaping from not escaping to.	"…equally so it makes me temporary forget rather than fix my problems, so I used it as an escape, as a way to ignore it all, to flee" (Gokiburi)
GAMES HELP	**27**		

General Help	6	Games help in general ways, they are like medication for minor problems such as staying calm.	"The first set of questionnaires actually revealed how much games help people." (Lex Hover)
Helps Emotional Pain	9	Games help anxiety, stress, depression and negative feelings. Helps keep the mind active and to see positives and find a way out.	"I play games to escape from what would otherwise be crippling depression." (Dicerson)
Coping Mechanism	3	They are a coping mechanism which is better than drugs or alcohol.	"...games aren't' necessary escapism but can be used as a healthy coping mechanism." (Ryan)
Helps Social Pain	4	Games help combat loneliness in the real world.	"It also helped my sister get over her social anxiety, bit by bit." (LordBloodySoul)
Helps Physical Pain	2	Games help with dealing with or reducing physical pain.	"Video games help me deal with chronic severe ankle pain, and also helped my anxiety." (sid2tiger6)
Game Type Helps	3	The type of game matters, critical thinking games, puzzle games, social games help, co-op games and nostalgia in games help.	"When I feel panicked or anxious he tells me to grab a game, one that makes me think critically to solve a problem instead of something that relies on reflexes." (DozerZigashi)
BALANCED INDEPENDENT RESEARCH	**16**	Further understanding of the topic via balanced and independent research that considers both positive and negative effects of gaming akin to the medical	"I'm all for exploring how games can help people, but it's no different than prescription drugs that can also be abused or have side effects..."

		industry. Until then we have to be cautious.	(SuperLotus)
SCEPTICISM OF GAMING	**15**		
Scepticism of Gaming Industry	4	Positive research that ignores the negative effects of gaming will be met with scepticism, lack of trust.	"On the other hand any research into positive psychological effects of gaming leads to suspicion, and more if it's funded by the games industry because it looks like they're trying to imitate Coca Cola's advertisements claiming it to be a health tonic." (Steven Neiman)
Industry Status Quo	3	The gaming industry doesn't care, doesn't want to evolve or change in any way. It is complacent and happy with the way things are.	"The other problem is that considering the positive effects of games would lead to a shift in the ways that games are designed and marketed, and the companies that have the money to fund anything are the ones which are optimized to prosper in the current market, so it's not in their interests to change that market." (Steven Neiman)
Gamers Bias	8	Recent disorders are knee-jerk or over the top and any criticism of gaming won't stop people playing.	"Also, anything bad people say about it sounds like old people whining and makes consumers all the more protective of their beloved games." (Steven Neiman)

GAMES CREATE POSITIVITY	7		
Positive Digital Experiences	3	Games are entertainment with lots of positive effects such as accomplishment and happiness.	"It gives me a feeling of accomplishment just to save the land of Hyrule." (EmeraldDragon)
Positive Real Life Experiences	4	Gaming fuels positive real life experiences such as family bonding and connections with people. Games can nurture transferable skills to real life.	"Games have helped my family a lot. It bond us together. Even years after we all have left our home and lived our lives separate, there is not a Sunday when we are not all in the same house playing video games together." (LordBloodySoul)
REAL WORLD ISN'T RIGHT	6	The real world is not right and so the digital world is better than the real world. People get frustrated with the real world. Gaming fills a void.	"I'm so glad that there are others who sometimes feel there is something "fundamentally not right with the outside world". (Ryan)
TEMPORARY ESCAPISM IS OK	4	If you understand why you game, perhaps for enjoyment or relaxation, escapism is a good thing.	"If you know you are fleeing from things and you just need to be away then good, do that, let the troubles to the side and have fun/relax." (Gokiburi)
REAL WORLD APPRECIATION	3	Games can make people realise the real world is not so bad and to appreciate the real life support and connections someone has. No matter how bad a person feels, life goes on.	"And the reason I play this game, despite its sad and hopeless ton, is that it reminds me that my world isn't so bad, and asks me to look at and value the connections

			and the support that I have." (SocraTetris)
GAMES ARE A PROBLEM	3	Games are bad for your mental health, for example gambling.	"Games are bad for Your mental health." (Dag Backerud)
INTEGRITY OF GAMING INDUSTRY	2	Importance of ethics, transparency and best practices/rules in the gaming industry.	"It would make it seem, as if the game community is trying to ignore the harmful effects of games, which might do serious damage to its image." (Toboter XP)
CRUTCH TO DISORDER	2	Games cannot go from crutch to disorder, dependence to addiction or a casual habit to addiction.	"It would be too easy to turn help into dependence then into full blown addiction without ethics and morals." (Sertan Doom)

Table 4.1

Results

The codes within the original comments from viewers were then counted on a pertinent sentence by sentence basis, unless there were two short sentences clearly linked together, in which case that was counted as one pertinent sentence. In some cases, a sentence would yield two or more codes.

Looking at Table 4.1, four clear code categories emerged from the data which we will discuss below.

(Gaming Disorder is a) Symptom of a Real Life Problem
The content analysis of the data shows us that comments relating to the sentiment of gaming disorder being a symptom of a real life problem rather than a cause are the most prevalent (Symptom of a Real Life Problem, N = 33 (28%) (a sum of subcategories Gaming Disorder is a Symptom Not a Cause, N = 17, Gaming to Avoid Problems is Bad, N = 16)).

Comments from the first subcategory, Gaming Disorder is a Symptom Not a Cause, N = 17, show people feeling that gaming disorder is misunderstood and is a symptom of much larger real life problems akin to any other disorder, *"You are pretty much right, an addiction and/or obsession are symptoms of a deeper issue, and it doesn't have to be games, any form of escapism could do."* (Ivarent). Comments from this subcategory account for nearly 15% of all comments made.

This was closely linked to the second subcategory of Gaming to Avoid Problems is Bad, (N = 16) where people felt that avoiding real life problems by gaming, and therefore potentially becoming addicted, was a bad idea and that it doesn't make real life problems go away *"...equally so it makes me temporary forget rather than fix my problems, so I used it as an escape, as a way to ignore it all, to flee."* (Gokiburi) and *"Addiction is rarely about the thing you're escaping into. It's about what you're escaping from."* (Patrick Reding). Comments here accounted for nearly 14% of all comments in the text.

As can be seen, this code category makes up just under a third of all of the comments in the data.

Games Help
The next most popular sentiment was that games help people in all sorts of ways and can therefore be looked upon as a positive way to help deal with a variety of issues whether they are mental or physical (Games Help N = 27 (23%), (a sum of subcategories General Help, N = 6, Helps Emotional Pain, N = 9, Coping Mechanism N = 3, Helps Social Pain, N = 4, Helps Physical Pain, N = 2 and Game Type Helps, N = 3)).

Looking at the subcategory General Help, (N = 6), people were upbeat about the positive effects of games, that they can help people in a general sense *"...games as therapy can definitely help people that need the balance and release."* (Sped Wyrm).

Helps Emotional Pain, (N = 9), was a strong subcategory as this included remarks that relate to help with anxiety, stress, depression and negative feelings. *"My therapist told me that games can be great to deal with anxiety."* (DozerZigashi) and *"I've actually been using "Don't Starve" to help me with my depression/executive disfunction."* (Draekn Jensen).

Coping Mechanism (N = 3), shows an opinion for games being used as a healthy coping mechanism *"Games are a coping mechanism, they are one of the most entertaining and enjoyable activities (for many, at least.) there are..."* (Dicerson).

Helps Social Pain, (N = 4), gives insight into the help that games give to people suffering with social issues such as social anxiety *"She also managed to approach*

people in Real Life more easily, without fearing it too much." (LordBloodySoul) and loneliness *"And well I don't feel so lonely when I play games."* (Irma Trubura).

Helps Physical Pain, (N = 2) only had a couple of comments but still gave us some meaningful insight into gaming helping with physical pain, *"Video games help me deal with chronic severe ankle pain, and also helped my anxiety."* (sid2tiger6).

Game Type Helps (N = 3), whilst only small, gives us a few fantastic insights into what types of games help people, *"When I feel panicked or anxious he (therapist) tells me to grab a game, one that makes me think critically to solve a problem instead of something that relies on reflexes."* (DozerZigashi) and *"From what I know about the research, social games help a lot, and co-op games help, too. There's also evidence that nostalgia helps people who are lonely, at least for a little while."* (Kyle).

Balanced Independent Research
The third most popular category that emerged from the data was there is a need for research that explores both the positives and negatives of how gaming affects the mental health of people, (Balanced Independent Research, N = 16 (14%)).

The comments mainly focus on caution until further research on the effects of gaming on people with mental health issues is conducted, *"I'm all for exploring how games can help people, but it's no different than prescription drugs that can also be abused or have side effects[...]."* (Toboter XP).

The research has to be balanced and cover positive and negative effects and not be funded by the gaming industry, *"On the other hand any research into positive psychological effects of gaming leads to suspicion, and more if it's funded by the games industry because it looks like they're trying to imitate Coca Cola's advertisements claiming it to be a health tonic."* (Steven Neiman).

Scepticism of Gaming Industry
The last of the most popular sentiments running through the narrative of the comments is people were sceptical of gaming (Scepticism of Gaming Industry N = 15 (13%), (a sum of subcategories Scepticism of the Gaming Industry, N = 4, Industry Status Quo, N = 3 and Gamers Bias, N = 8)).

The subcategory of Scepticism of the Gaming Industry, (N = 4), comments seem to display a lack of trust towards the gaming industry, *"[...]the scientific analysis of commercial games, or really anything commercial[...]always brings with it the danger of contamination via indirect bribery (aka donations), so that major game studios could just get someone to sign of their new Brain Explosion Simulator 2022 or whatever as psychologically valuable, which again might lead to a distrust in games."* (Toboter XP).

Industry Status Quo, (*n* = 3), shows sentiment that the gaming industry doesn't need or want to change in order to explore the potential of creating games that have positive effects on the gaming community, *"It's enjoying its status as entertainment and the profits that come with that status, and it doesn't care that it could do more by continuing to grow into a beautiful and diverse medium."* (Steven Neiman).

Gamers Bias, (*n* = 8), the sentiment of this category is that the gaming community is biased about the effects of gaming on mental health, "Also, anything bad people say about it sounds like old people whining and makes consumers all the more protective of their beloved games." (Steven Neiman).

Analysis

From the analysis, we can make a few conclusions. The first being that the comments suggest that gaming addiction is seen as a symptom and not a cause of mental illness (N = 17, 14%) and that gaming disorder can be a sign of a gamer avoiding real life problems, rather than dealing with them (N = 16, 14%). Therefore, we can conclude that the majority of comments in the most prevalent category of Symptom of a Real Life Problem indicate that most of the commenters feel that games do not cause mental health issues, rather gaming disorder or addictions are caused by deeper issues in the real world.

The second conclusion that can be drawn from the data is that commenters feel that games do help in a very positive way in some form or another (N = 27, 23%) with many kinds of conditions such as anxiety and depression (N = 9, 8%), physical pain (N = 2, 2%), social anxiety (N = 4, 3%) and it can be a healthy coping mechanism (N = 3, 3%) as well as generally (N = 6, 5%). Its' worth noting that commenters felt that game type did affect the efficacy of games ability to help (N = 3, 3%). Again, because this is the second strongest narrative running through the data, we can conclude that commenters feel strongly about the positive effects of gaming on not only mental health but wider health issues.

A third conclusion from the content analysis is that when the categories and subcategories were analysed for positive, negative and neutral sentiment (Table 4.2) we see that as a whole, people were mostly positive in what they thought about digital games as a means to help people with mental health issues. Yes there were concerns but interestingly, around 44% of negative comments were constructive in their nature (Gaming to Avoid Problems is Bad, N = 16) and are a real insight into the minds of people who are sceptical about gaming avoid real life problems. The neutral comments are also very meaningful as they balance the argument that digital games can be used as a form of mental health therapy. We have to be careful that balanced research into the benefits and the possible side effects of digital games is fully understood before medical professionals include them in their toolkit and that we have to be aware of a possible risk of digital

gaming therapy evolving from being an aid, where the patient shows improvement with their symptoms, to becoming a disorder where the patient cannot function without playing digital games.

Following on, the research has to also be independent as the gaming industry is generally looked upon with a lack of trust and so gaming industry backed research that concludes games have only positive effects on mental health will be looked at in a similar way to other large industries and companies. So similar to the medical industry bringing out a new form of medication, the positive and negative effects of gaming need to be fully understood before gaming is used as a form of medication in the wider population.

CODE CATEGORY SENTIMENT	N	%
Positive	64	54%
Gaming Disorder is a Symptom Not a Cause, N = 17, Games Help, N = 27, Games Create Positivity, N = 7, Real World Isn't Right, N = 6, Temporary Escapism is OK, N = 4, Real World Appreciation, N = 3.		
Negative	36	31%
Gaming to Avoid Problems is Bad, N = 16, Scepticism of Gaming, N = 15, Games Are a Problem, N = 3, Integrity of Gaming Industry, N = 2.		
Neutral	18	15%
Balanced Independent Research, N = 16, Crutch to Disorder, N = 2.		
Total	118	100%

Table 4.2

(Q5. on a new page).

Q5. Literature Bricks

Rowe, J. W., & Kahn, R. L. (1997). Successful aging. *The Gerontologist, 37*(4), 433–440. https://doi.org/10.1093/geront/37.4.433
Research shows we can affect the quality of later life, once thought to be purely down to uncontrollable effects of ageing. Specifically for cognitive function, the results of a good sized 1,189 participant (70-79 years old) longitudinal study show education and self-efficacy are major indicators of cognitive function in later life. Therefore, keeping the brain active via continual testing and education during our lives can improve or maintain cognitive function. This tells me digital games can aid in these cognitive function goals and help shape the mental health of the population. The concept of successful ageing has three aspects, one of which relates particularly to my study - 'The maintenance of high physical and cognitive function'. Ensuring relatively high cognition via digital games throughout the lifecycle could reduce the likelihood of cognitive decline and therefore related diseases like depression. The paper suggests interventions related to successful ageing should be explored.

Lockwood, K., Alexopoulos, G., & van Gorp, W. G. (2002). Executive dysfunction in geriatric depression. *The American Journal of Psychiatry, 159*(7), 1119–1126. https://doi.org/10.1176/appi.ajp.159.7.1119
40 depressed and 40 healthy participants (20 20-60 years and 20 61+ years for both groups) showed depressed participants performed poorly during executive function tests suggesting depression and executive dysfunction are not independent. Old age compounded these effects, motivating my study insofar as digital games could help with improving the later life cognitive deficits and possibly depression symptoms. A good sample size was used, stringent exclusion criteria, a well balanced healthy comparison group, balanced depression groups measured by a recognised depression scale, a control for education and health status and an array of modalities covered in tests balanced the limitations of a small sample. A larger sample could provide insights about the ages of the onset of depression. If a digital game intervention was used, a longitudinal study would result in more meaningful data. A more equal gender balance is needed. These study limitations show potential for a research gap.

Story, T. J., Potter, G. G., Attix, D. K., Welsh-Bohmer, K. A., & Steffens, D. C. (2008). Neurocognitive correlates of response to treatment in late-life depression. *The American Journal of Geriatric Psychiatry: Official journal of the American Association for Geriatric Psychiatry, 16*(9), 752–759. https://doi.org/10.1097/JGP.0b013e31817e739a
The study shows participants with Late Life Depression and low verbal memory and processing speed test scores can have a lower response to antidepressant medication. This relates to my study because other therapy solutions are needed for older people with cognitive deficits due to potential for low response to

medication. Digital games could act as therapy to those suffering with depression, providing improved cognitive health and well-being. Inclusion was based on strict criteria by a psychiatrist using an established scale, depression was evaluated by self report and a psychiatrist who was blind to neuropsychological performance. Neuropsychological functioning was evaluated with established scales and a psychometrician who was blind to depression status. The sample size could be bigger and be more gender balanced, a placebo group could have been included and the neuropsychological tests could be more extensive to produce greater insights. These study limitations show great potential for further research.

Griffiths, M. D., Kuss, D. J., & Ortiz de Gortari, A. B. (2017). Videogames as therapy: An updated selective review of the medical and psychological literature. *International Journal of Privacy and Health Information Management (IJPHIM), 5*(2), 71-96. https://doi.org/10.4018/IJPHIM.2017070105
The literature review shows a plethora of research related to digital games being used as therapy in various scenarios. In relation to my study the following scenarios particularly motivate it; video games as physiotherapy and occupational therapy, video games and cognitive rehabilitation, videogames and therapeutic benefits in the elderly, video games in psychotherapeutic settings, video games and healthcare and video games and psychological well being. If digital games can act as a therapeutic tool for good in many different scenarios, then I am motivated to find commercial digital games that could provide positive outcomes for depressed individuals without the need for medication. The literature review is mainly based on older studies and does not include the criteria of inclusion. However, the broad applications shown in the review suggest digital games as a therapeutic tool, more specifically as a therapeutic depression tool is not a far fetched concept.

Yee, N. (2006c). The psychology of MMORPGs: Emotional investment, motivations, relationship formation, and problematic usage. In R. Schroeder & A. S. Axelsson (Eds.), *Avatars at work and play: Collaboration and interaction in shared virtual environments* (pp. 187-207). London: Springer-Verlag.
The study highlights valid concerns regarding potential problematic usage of players in four popular MMORPGs (EverQuest, Dark Age of Camelot, Ultima Online and Star Wars Galaxies). The study was large (N = 30,000 over four years) and survey based. Limitations are the survey could have been completed by mainly experienced players as the demographics of the players was anonymous and the survey is non-random. The paper describes a concept of problematic usage which cannot be ignored if digital games are going to be used as a tool for depression (and other mental diseases). Whilst digital massively multiplayer online role playing games or real-time strategy games could deliver increased or maintained cognitive function or reduced depressive symptoms, or both, digital game therapy has to be administered knowing a small addiction risk to a small section of the public is possible.

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., Kong, E., Larraburo, Y., Rolle, C., Johnston, E., & Gazzaley, A. (2013). Video game training enhances cognitive control in older adults. *Nature, 501*(7465), 97-101. https://doi.org/10.1038/nature12486
Study shows multitasking declines with age, older adults improve on multitasking training tasks and most importantly, the training benefits transfer to untrained cognitive tasks with benefits lasting months after intervention with some abilities comparable with young people in their 20s. This is study motivation as if digital games train people in the right cognitive areas, the benefits are generally transferred causing improved cognitive abilities over time. Therefore, regular digital serious game play can steadily improve or maintain cognitive function during the lifecycle, particularly in older adults. The study was thorough with a challenging and adaptive fast paced digital game, a balanced gender spread, with all right handed non-gamers, with participants encouraged to play at the same time of day, not talk to other participants and to avoid playing other digital games. A commercial game could be used now, with an even more longitudinal study design revealing a gap.

Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychology and Aging, 23*(4), 765-777. https://doi.org/10.1037/a0013494
Rise of Nations commercial real time strategy game improved participants' cognitive functions in four of five executive control tasks and one of four visio-spatial tasks showing skill transfer from playing to individual cognitive abilities. In relation to my study it means cognitive functions relating to control in older adults can be improved by cheap, available, commercial real time strategy games. The study design motivates my study. Sample size was good and study was rigorous as participants were screened for visual acuity, gender was balanced, they were all right handed, were assessed for dementia using a valid scale (mMMSE). Limitations that suggest a gap; increasing the training time on RON could lead to increased transfer effect or improvement on a wider range of cognitive abilities, having a control group that played a different genre of digital game to confirm Rise of Nations or real time strategy game efficacy on cognitive transfer.

Allaire, J. C., McLaughlin, A. C., Trujillo, A., Whitlock, L. A., LaPorte, L., & Gandy, M. (2013). Successful aging through digital games: Socioemotional differences between older adult gamers and non-gamers. *Computers in Human Behavior, 29*(4), 1302-1306. https://doi.org/10.1016/j.chb.2013.01.014
The study shows compared to non-gamers, regular and occasional gamers scored significantly better for well-being and negative affect, regular gamers had significantly less depression and occasional gamers had less depression. Older adult game players scored higher for socioemotional functioning generally. This motivates my study as regular gaming improves socioemotional levels in older adults which could be used as therapy for depressed patients or to manage levels

throughout life. Validated scales (SF-36, CES-D and PANAS) are used for socioemotional functioning and age and education were consistent across the gamer groups. Limitations were sample size, participants were aware of the digital game intervention therefore may be more inclined to sign up, differences in gamers versus non-gamers may be down to pre-existing differences that drive gaming interest or not. Games were mostly digital traditional games like card games. More immersive games could test the effects of complex digital games on socioemotional functioning.

Russoniello, C. V., Fish, M., & O'Brien, K. (2013). The efficacy of casual videogame play in reducing clinical depression: A randomized controlled study. *Games for Health Journal, 2*(6), 341–346. https://doi.org/10.1089/g4h.2013.0010
The study shows casual video games played on mobile phones significantly reduce the symptoms of depression. The study design of training sessions conducted at home in between lab sessions motivates my study. The importance of commercial digital games improving depression because they are cheap and easy to access also motivates my study insofar as commercial digital games as therapy could provide benefits to those suffering from depression without the need for medication. The study is a small sample size with randomised groups, depression was measured by a validated scale (PHQ-9) and depression was evenly matched in both groups. The limitations however are the experimental group has a mean age of 29 which is a limitation in relation to my study. Other limitations are the self report nature of the intervention, participants could lie about how much they played in between lab sessions.

Kaufman, D., & Zhang, F. (2015, May). *Can playing massive multiplayer online role playing games (MMORPGS) help older adults?* 7th International Conference on Computer Supported Education, CSEDU 2015, SCITEPRESS - Science and Technology Publications, 527–535. https://doi.org/10.5220/0005551405270535
The study indicates World of Warcraft, a massively multiplayer online role playing game (MMORPG), affects depression. Quality of guild play (involving satisfaction with guild mates and social interactions), time spent playing and social motivation for playing all had significant effect on levels of depression. This motivates my study. It highlights time spent and the general game mechanisms of MMORPGs reduce depression. Loneliness (ULS-8), depression (CES-D) and social support (MSPSS) were measured with valid scales. Limitations are the study relied on self reporting, possible social desirability bias, participants were heavy gamers and most likely are experts in World of Warcraft. The sample could have been bigger (n=176) more balanced in terms of gender, age, relationship status, work status and education levels. A research gap could involve non MMORPG players participating in MMORPG in a longitudinal intervention style study with depression levels measured before and after intervention and multiple months after.

(Q6. on a new page).

Q6. Study Design

I have shown via my literature review that we can age successfully by affecting the way we age, in particular by striving for high cognitive functioning during our lives (Rowe & Kahn., 1997). According to research, there is a relationship between depression and low cognitive function, with depressed participants displaying low cognitive function but even lower if they are older (Lockwood et al., 2002). It is a concern that older adults with low cognitive function who suffer from Late Life Depression are also at risk from potentially poor responses to antidepressant medication (Story et al., 2008) and so new avenues need to be explored to help people in later life with poor cognitive functioning and depression. It has been shown that older adults, when exposed to a serious game regularly, can improve their cognitive functioning in relation to multitasking to levels seen in untrained 20 year olds (Anguera et al., 2013) which is exceptionally interesting as there is real hope that digital games can be used as therapy, although this isn't an entirely new concept (Griffiths et al., 2017).

However, in order to be more effective, the digital game therapy has to be both financially viable and easy to administer. This is where research into the effects of commercial digital games as a tool to improve cognitive function is important. When young regular gamers were compared to non-gamers in relation to socioemotional levels such as depression, well-being and negative affect, they were shown to have less depression, negative thoughts and greater well-being suggesting there could be ways to help improve symptoms of depression by playing commercial digital games regularly (Allaire et al., 2013).

Research has also shown that a commercial off the shelf digital real-time strategy game Rise of Nations has shown an improvement in cognitive function in older adults with only 23.5 hours play over the course of a month (Basak et al., 2008). Even more critically, those cognitive improvements were displayed, to a lesser extent, six months after playing the game with encouragement that regular digital game play transfers cognitive benefit to the individual's cognitive skill set rather than just improving an older adult's ability to play that particular digital game better (Basak et al., 2008). Another commercial off the shelf game, this time the massively multiplayer online role-playing game World of Warcraft, was shown to reduce depression in older adults (Kaufman & Zhang, 2015).

Unfortunately, a small but significant percentage of people may be susceptible to the possibility of addiction to a prescribed digital game therapy (Yee, 2006c) and so doctors and medical experts need to also be aware of the risk that a digital game therapy may lead to unwanted side effects as with any other prescribed medication.

It's worth noting that in this experiment I will not be examining the relationship between cognitive functioning and depression, I will be making sure that depression is spread evenly in all gaming groups, representing a likely distribution in the general population.

Research questions

1. What are the effects of a commercial real-time strategy (RTS) game and a commercial massively multiplayer online role playing game (MMORPG) on cognitive function in older adults?
2. If positive effects are shown, will those effects be noticeable 12 months after intervention?
3. Do older adults find a digital game intervention a useful tool to improve cognitive function?

I propose three research questions above in order to 1) examine the effect of each game genre (real-time strategy and massively multiplayer online role playing game) on cognitive function, 2) examine if any positive effects will be noticeable 12 months after the digital game intervention and 3) what older adults actually think about the usefulness of a digital game intervention.

Hypothesis

1. I hypothesise that both intervention groups will have a positive effect on the levels of cognitive function.
2. I hypothesise that the positive effects, if any, of a digital game intervention will be lasting and be noticable 12 months after the digital game intervention has taken place.
3. I hypothesise that older adults will be sceptical about the worth of digital games as a tool for combating cognitive decline but they will be more receptive to them after the experiment.

Method

This will be a longitudinal, controlled, mixed methods, between subjects study design.

Participants

The participants;
- Will be recruited via various UK depression help forums such as dealingwithdepression.co.uk, carersuk.org and carenity.co.uk, noticeboards in community centres, flyers in local businesses windows, local newspaper ads and by reaching out to local depression groups.
- Will be aged 65-80 years old.

- Will be screened for gaming activity with a very short survey and only non-gamers will be included in the study.
- Will be asked to fill in a demographic survey for general demographic information such as age, gender, relationship status, living status and education levels (Kaufman & Zhang, 2015), although we don't need the survey part relating to work status as most will be retired.
- Will be asked about any regular medication they are taking.
- Will all suffer from depression and be screened with a validated PHQ-9 scale, (Russoniello et al., 2013).
- Will be assessed and excluded for ethical reasons if they have had 1) significant medical illness (i.e., history of metastatic cancer, brain tumour, myocardial infarction, or stroke), 2) history of head trauma or epilepsy, 3) delirium, dementia, Huntington's chorea, or Parkinson's disease, 4) history of substance abuse, 5) hearing, vision, or motor impairment that precluded neuropsychological testing, and 6) an PHQ-9 score lower than 4 and therefore not suffering from depression (Lockwood et al., 2002).
- Will remain anonymous.

Intervention

Once 150 participants that matched the criteria for inclusion are recruited, they will be randomly assigned to one of the three game condition groups. Each game condition group will then be plotted against each of the independent variables (depression, age and gender) to check for potential confounds before participants are informed which group they are in.

Group 1
50 participants will play the popular real time strategy game Age of Empires IV.

Group 2
50 participants will play the most popular massively multiplayer online role playing game World of Warcraft.

Group 3
50 participants will do everything group 1 and group 2 do, except they will play no games for the duration of the experiment.

Participants in group 1 and group 2 will be told to play the intervention game three times a week at home for exactly one hour per session on Monday, Wednesday and Fridays between 9am-10am, for the first four months of the trial. This is similar to a study conducted to find the efficacy of casual digital game play in reducing clinical depression where in that study participants played a minimum of 30 minutes three times a week with 24 hours in between. (Russoniello et al., 2013). However, I want to include more control in this study so

all participants are being exposed to the intervention in the same time windows for exactly the same amount of overall time by the end of the experiment.

Cognitive assessment

The participants will be assessed for baseline (Assessment 1) cognitive function and cognitive function at Assessment 2 (after 1 month), Assessment 3 (after 4 months) and Assessment 4 (after 16 months). The cognitive battery is the same as used to test the effect of the real time strategy game *Rise of Nations* on the transfer to cognitive abilities (Basak et al., 2008). This involves 10 tests;

Executive control tasks
1. Operation Span
2. Task Switching
3. N-Back Task
4. Visual Short-Term Memory
5. Ravens Advanced Matrices
6. Stopping Task

Visuo-spatial Attentional Tasks
1. Functional Field of View
2. Attentional Blink
3. Enumeration
4. Mental Rotation

Recording the responses
Participants will get a trial run at each task before being informed that the next attempt at the task will be recorded for the study. In each task a participant will be allocated a score and predefined pass or failure thresholds will inform the study if a participant passes or fails a task. The total score will then be recorded for the set of 10 tasks, for example 7/10 passes for Participant X. This would then be converted to a decimal of 1, for example 0.7 (with 1 indicating a score of 10/10) for the overall score for cognitive function.

The assessment sessions will be conducted over two days to ensure we respect the participants' energy and focus levels and will be conducted at 9am consistently to control for time. All participants will use the same apparatus for all tests to reduce confounds.

Depression assessment

The participants will be assessed for baseline levels of depression (Assessment 1). Levels of depression will be measured with PHQ-9 scale when used to find the efficacy of casual digital game play in reducing clinical depression (Russoniello et al., 2013).

Insights from older adults on digital game interventions

The participants will fill in a short survey about their thoughts and opinions on the use of digital game interventions for mental health in relation to cognitive function and depression. The participants will fill out a survey before the digital game intervention and after, at the end of month 4.

Survey

This will be a short 5-point Likert scale survey which will include ten statements with answers (in this order) for all; strongly disagree, disagree, neither disagree or agree, agree, strongly agree:

1. Digital games could be useful now for increasing/maintaining my cognitive skills and keeping my mind active.
2. Digital games could be useful now for helping me with my symptoms of depression.
3. I would be open to playing digital games as a way to improve/maintain my cognitive health.
4. I would be open to playing digital games as a way to improve/manage my depression.
5. I see playing digital games as something only people who are younger than me do.
6. I see playing digital games as entertainment only.
7. I am worried I could become addicted to playing digital games.
8. I think digital games would have been a useful tool to improve/maintain my cognitive health before I got to older age.
9. I think digital games would have been a useful tool to improve/manage my depression before I got to older age.
10. I would prefer to use digital games as therapy instead of taking regular medication for my depression.

There will be a very short qualitative section of questions at the end of the survey as follows.

Qualitative questions
1. What do you think about digital games in general?
2. Why haven't you been interested in playing digital games before?
3. Do you have any concerns about using digital games as therapy for managing your mental health (cognitive health and/or depression)?
4. Do you have any other thoughts about using digital games for managing your mental health (cognitive health and/or depression)?

Controlling variables

The study will aim to control as many aspects as possible:

- The length of time participants are exposed to the digital game intervention (3 x week for 1 hour exactly).
- The time and day the participants will be exposed to the digital game intervention (Monday, Wednesday and Friday between 9am–10am).
- Gender spread in all game condition groups (male, female, non-binary and prefer not to say)
- Depression levels will be spread naturally because of a good sized sample for each group of N = 50, although this will be checked for confounds.
- A control group will be used (group 3) to compare to. Group 1 and group 2 will be playing the digital game intervention which could be affecting cognitive function from an hour of relaxing or general benefits of digital games but this is unlikely to produce significant uplifts in cognitive function. I could ask the control group to browse the internet for 1 hour or play a non-digital game three times a week on Mondays, Wednesdays and Fridays but this may add more noise to the data so I would like them to play no digital games at all over the course of the 16 month study.
- The cognitive assessment sessions will be conducted at 9am consistently to control for time.
- All participants will use the same apparatus for all cognitive battery tests which will involve the same computer, monitor, mouse, keyboard, chairs and tables.
- All participants will be right-handed.
- Participants will be asked not to interact with other participants they may meet during assessment days
- Participants will be asked not to play digital games for the duration of the experiment.

Study schedule (Fig. 6.1)

After initial demographic survey and assessment of cognitive function and depression levels (Assessment 1 – baseline), participants will be allocated a game condition group of group 1, group 2 or group 3. Participants in group 1, group 2 and group 3 will fill in a short survey (Survey 1) about their thoughts and feelings towards digital games as a tool for combating cognitive decline and depression symptoms, this will include mostly quantitative questions with some qualitative open ended questions.

Before commencing the intervention, group 1, group 2 and group 3 will be introduced to both games to expose all participants equally to the introduction session. Instructions on how to play the game will be printed out so each participant reads the same instructions and descriptions of each game e.g. what is the purpose of each game and what are the general aims of gameplay. A short video demo will be shown to all participants for both games to familiarise the participants with how the game looks and works.

Participants assigned to group 1 (Age of Empires IV, real-time strategy game) or group 2 (World of Warcraft, massively multiplayer online role playing game) are to play their designated game at home for one hour exactly, three times a week for exactly one hour per session on Monday, Wednesday and Fridays between 9am-10am, for the first four months of the trial. As mentioned before, this is similar to a study conducted to find the efficacy of casual digital game play in reducing clinical depression where in that study participants played a minimum of 30 minutes three times a week with 24 hours in between. (Russoniello et al., 2013). However, I want to include more control in this study so all participants are being exposed to the intervention in the same time window for exactly the same amount of overall time by the end of the experiment.

End of month one
Participants in group 1 and group 2 will play the digital game assigned to them for one month. Group 1, group 2 and group 3 will be again assessed for cognitive function only (Assessment 2).

End of month four
Participants in group 1 and group 2 will play the digital game assigned to them for three months. Group 1, group 2 and group 3 will be again assessed for cognitive function only (Assessment 3). After this point, no more game interventions will occur. All participants in group 1, group 2 and group 3 will fill in a short survey (Survey 2) about their thoughts and feelings towards digital games as a tool for combating cognitive decline and depression symptoms, this will include mostly quantitative questions with some qualitative open ended questions (same survey as before).

End of month 16 (12 months after last game intervention)
Finally, at the end of month 16, so 12 months on from the last game intervention, the final assessment of cognitive function will occur (Assessment 4).

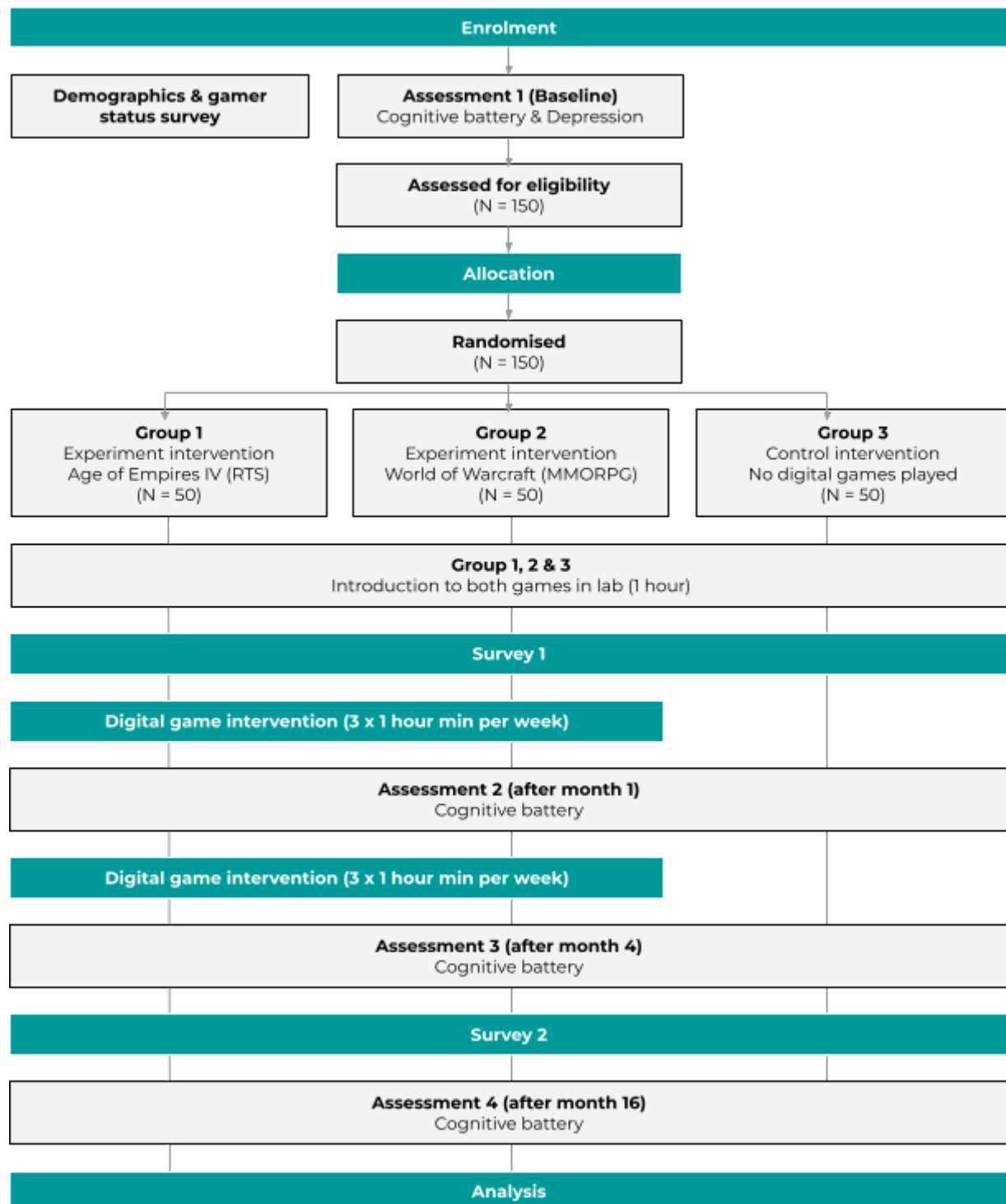Study schedule on a new page (**Fig. 6.1**).

Fig. 6.1 Study schedule

Analysis

I will analyse the data to see if there is a relationship between group 1, group 2 and group 3 and the 10 subtests of cognitive function. I will tabulate descriptive statistics showing the central tendencies for each of the 10 subtests of cognitive

function against the three game condition groups. I will be very interested in the medians as these are not susceptible to outliers skewing the data.

There will be three game condition groups with participants being exposed to only one condition group each, therefore a between subjects design of a Kruskal-Wallis Test is to be used inorder to find which a condition group is affecting each of the 10 cognitive function battery tests. The significance level is set at $p < 0.05$ therefore the probability that the effect of any of the condition groups is the product of pure chance, only 5% of the time. If a significant result is shown, a Dunn's Test with post-hoc pairwise comparisons with Bonferroni corrections will be run afterwards. This is to find out which condition group is having the most effect. Because I will be running multiple comparisons I have to account for the increased chance of attaining a false positive which may push the p-value over $p < 0.05$ causing a type II error. The significance level for the Dunn's Test has to be adjusted to p-value < 0.017 (0.05 / 3 game condition groups) to avoid insignificant results being misinterpreted as false positives.

I will analyse the data from both surveys both quantifiably and for the open ended questions I will conduct a small inductive content analysis to quantify the common insights that emerge from the data.

(Q7. on a new page).

Q7. Limitations of the Study

Limitations

I haven't measured the effect of depression on cognitive function, having two response variables is beyond my experience at this time and so I kept it simple. I will however account for depression by randomising participants with depression into different game groups to ensure a random spread of depression levels in each game group.

I couldn't assess the effect of other real life variables such as relationship status, living status and education levels, although I have ensured a large enough sample size to naturally account for these other variables.

It could also be that the mechanics of playing any digital game could affect the cognitive function battery scores and not the actual digital game.

A further limitation is it's possible that as the intervention takes place entirely at participants' homes that they'll play the digital games; not enough; too much; not at all; at different times or on back to back days. This loses control but due to the longitudinal nature of the experiment it would probably be unrealistic for 100 participants to come into the lab three times a week for four months. All participants will be using their own equipment in their own homes which adds ecological validity but loses control insofar as some participants will have advantages simply because they choose to spend their money on a big monitor or may have fantastic broadband connection.

Survey 1 and Survey 2 are self-report so are susceptible to social response bias. It's also possible that due to the longitudinal nature of the study that there will be some participant drop off for the final month 16 assessment. I have purposely only asked participants to play the digital game intervention up to the end of month 4. This is as far as I think I can realistically push participants to keep to the schedule of game playing.

I want one overall score to represent cognitive function, and not ten individual scores, and so I will allocate a mark out of ten which is determined by predefined thresholds of pass or failure for each task. Although this achieves my goal of simplifying the scoring process, it also means that Participant X may achieve 7/10 at Assessment 1, 7/10 at Assessment 2, 7/10 at Assessment 3 and 7/10 at Assessment 4, but the mark could be attained by passes in different tasks each time. Therefore, the study will lose a bit of insight in terms of how the data breaks down but for the purposes of the study I think this still acts as a consistent measure across all participants so it is still a fair measure.

Another limitation is that because my literature review uncovered mostly North American literature, the scales and tests we're not necessarily the ones that are best suited for this type of experiment and further research would be needed to determine more suitable scales perhaps. In a real life study I would discuss this with my research colleagues or reach out to experts in the fields of psychology.

Further study

As with every study, there are limitations and compromises which inform us on how we can improve in a further study. I haven't measured the effect of depression on cognitive function, I have just accounted for it by randomising participants with depression into different game groups to ensure a random spread of depression in each game group. In a more complex future study, I would take depression assessments at the same time as I assessed for cognitive function (Assessments 1, 2, 3 and 4) and analyse the relationship between cognitive function and depression levels. Therefore having two response variables. I would then categorise participants into depression groups with the PHQ-9 scale (0–4 = none; 5–9 = mild depression; 10–14 = moderate depression; 15–19 = moderately severe depression; and 20–27 = severe depression). It would then be possible to analyse the effects of game condition groups against depression groups and cognitive function. In an ideal world the experiment would also be revisited five or ten years down the line with the same participants. The study could involve updated digital games in the same genre, or if digital games have progressed rapidly, then two appropriate genres will be decided upon at that point. It would be nice to have the study entirely in a lab but the practicalities of 100 participants in the labs at the same time (because I want participants to be exposed to the intervention at the same time on the same days) may be too much. The scale used to measure depression would be researched further, although the scale was validated and has been used in various studies, I would discuss this with my colleagues and perhaps use another scale.

References

Aymerich-Franch, L., Kizilcec, R. F., & Bailenson, J. N. (2014). The relationship between virtual self similarity and social anxiety. *Frontiers in Human Neuroscience, 8.* 1-10. https://doi.org/10.3389/fnhum.2014.00944

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health* 6(June). https://doi.org/10.3389/fpubh.2018.00149

Dechant, M., Poeller, S., Johanson, C., Wiley, K., & Mandryk, R. L. (2020). *In-game and out-of-game social anxiety influences player motivations, activities, and experiences in MMORPGs.* 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376734

Hartanto, D., Kampmann, I. L., Morina, N., Emmelkamp, P. G. M., Neerincx, M. A., & Brinkman, W. P. (2014). Controlling social stress in virtual reality environments. PLoS ONE. https://doi.org/10.1371/journal.pone.0092804.

Kang, S., & Gratch, J. (2010). Virtual humans elicit socially anxious interactants' verbal self-disclosure. *Computer Animation & Virtual Worlds, 21*(3-4), 473-482. https://doi.org/10.1002/cav.345

Lange, B., & Pauli, P. (2019). Social anxiety changes the way we move – A social approach-avoidance task in a virtual reality CAVE system. PLoS ONE 14(12), e0226805. https://doi.org/10.1371/journal.pone.0226805

Langer, M., & König, C. J. (2018). Introducing and testing the Creepiness of Situation Scale (CRoSS). *Frontiers in Psychology, 9.* https://doi.org/10.3389/fpsyg.2018.02220

Liebowitz. M.R., (1987). Liebowitz Social Anxiety Scale. Modern Problems of Pharmapsychiatry.

Mandryk, R. L., & Birk, M. V. (2017). Toward game-based digital mental health interventions: Player habits and preferences. *Journal of Medical Internet Research*, *19*(4), e128. https://doi.org/10.2196/jmir.6906.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. Behavior Research Methods. *Behavior Research Methods 44*, 1–23. https://doi.org/10.3758/s13428-011-0124-6.

McAndrew, F. T., & Koehnke, S. S. (2016). On the nature of creepiness. *New Ideas in Psychology 43*(Dec), 10–15. https://doi.org/10.1016/j.newideapsych.2016.03.003

Mullins, D. T., & Duke, M. P. (2004). Effects of social anxiety on non-verbal accuracy and response time I: Facial expressions. *Journal of Nonverbal Behavior 28(*1), 3–33. https://doi.org/10.1023/B:JONB.0000017865.24656.98.

Ni, T., Karlson, A. K., & Wigdor, D. (2011). *AnatOnMe: Facilitating doctor-patient communication using a projection-based handheld device.* SIGCHI Conference on Human Factors in Computing Systems (CHI '11). Association for Computing Machinery, New York, NY, USA, 3333–3342. https://doi.org/10.1145/1978942.1979437

Oozu, T., Yamada, A., Enzaki, Y., & Iwata, H. (2017). *Escaping chair: Furniture-shaped device art.* Eleventh International Conference on Tangible, Embedded, and Embodied Interaction (TEI '17). Association for Computing Machinery, New York, NY, USA, 403–407. https://doi.org/10.1145/3024969.3025064

Paredes, P., Ko, R., Calle-Ortiz, E., Canny, J., Hartmann, B., & Niemeyer, G. (2016). Fiat-Lux: Interactive urban lights for combining positive emotion and efficiency. 2016 ACM Conference on Designing Interactive Systems (DIS '16). Association for Computing Machinery, New York, NY, USA, 785–795. https://doi.org/10.1145/2901790.2901832

Van Looy, J., Courtois, C., De Vocht, M., & De Marez, L. (2012). Player identification in online games: Validation of a scale for measuring identification in MMOGs. *Media Psychology, 15*(2), 197-221. https://doi.org/10.1080/15213269.2012.674917

Vasalou, A., Joinson, A. N., & Pitt, J. (2007). *Constructing my online self: Avatars that increase self-focused attention.* Conference on Human Factors in Computing Systems. https://doi.org/10.1145/1240624.1240696.

Watt, M. C., Maitland, R. A., & Gallagher, C. E. (2017). A case of the "heeby jeebies": An examination of intuitive judgements of "creepiness". *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement 49*(1), 58–69. https://doi.org/10.1037/cbs0000066

Yee, N., & Bailenson, J. (2007). The proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, *33*(3), 271–290. https://doi.org/10.1111/j.1468-2958.2007.00299.x

Yip, J. C., Sobel, K., Gao, X., Hishikawa, A. M., Lim, A., Meng, L., Ofana, R. F., Park, J., & Hiniker, A. (2019). Laughing is scary, but farting is cute: A conceptual model of children's perspectives of creepy technologies. 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300303

END