

ACTION I LÄSFÖRSTÅELSE

Hur spelmekaniker kan påverka hur väl spelare tar till sig textinnehåll

ACTION IN READING COMPREHENSION

How game mechanics can affect how well players process the contents of texts

Examensarbete inom huvudområdet
informationsteknologi
Grundnivå 30 högskolepoäng
Vårtermin 2021

Rikard Larsson

Handledare: Erik Sjöstrand
Examinator: Jenny Brusk

Sammanfattning

Denna studie syftade till att kvantitativt jämföra hur dataspelmekaniska skillnader påverkar spelares förmåga att förstå en text. Hypotesen var att försökspersoner som engagerat sig i texten via spel innehållande logikpussel presterar bättre på ett efterföljande läsförståelsetest om innehållet än de som läst texten utan pusselement. För att testa hypotesen skapades tre spelversioner för presentation av samma text: Visual Novel (VN) utan pusselement, samt Phoenix Wright- (PW) och Danganronpainspirerade (DR) versioner med pusselement. Ett bekvämlighetsurval på 29 försökspersoner rekryterades. 8, 11 respektive 10 försökspersoner slumpades att spela version VN, PW eller DR, följt av läsförståelsetestet. Av maxpoängen 23 på läsförståelsetestet var medelvärde och standardavvikelse för respektive grupp: VN 10.6 ± 4.7 , PW 11.3 ± 2.9 , DR 13.4 ± 4.0 . Vid parvis jämförelse mellan grupperna identifierades inga signifikanta skillnader. Poweranalys visade att studien hade krävt fler försökspersoner för statistiskt relevant hypotesprövning. Studien kan fungera som pilot för liknande framtida större undersökningar.

Nyckelord: Digitalt Lärande, Engagemang, Visual Novel, Läsförståelsetest, Quantitative Reading Inventory

Innehållsförteckning

1	Introduktion.....	1
2	Bakgrund.....	2
2.1	Om läsförståelsetest	2
2.1.1	QRI-test.....	3
2.2	Inlevelse som motivation.....	4
2.3	Om spelmekaniker relevanta för denna studie	5
2.3.1	Översikt av PW-systemet.....	6
2.3.2	Översikt av DR-systemet	7
3	Problemformulering	9
3.1	Metodbeskrivning.....	9
3.1.1	Utveckling av manus och spelversioner.....	9
3.1.2	Syftet med användning av QRI-test i denna studie och urval av försökspersoner	10
3.1.3	Utformning av QRI-liknande test	11
3.1.4	Datainsamling och hypotesprövning	11
3.1.5	Etiska aspekter.....	12
4	Genomförande	14
4.1	Spel för tre olika sätt att presentera text.....	14
4.2	Manus till mordgåta i debattformat	19
4.3	Läsförståelsetest om mordgåtan	21
4.4	Pilottester och därpå följande modifieringar	21
4.4.1	Justeringar med avseende på speldesign och buggfixar.....	22
4.4.2	Justeringar av manus till mordgåtan	22
4.4.3	Justeringar av läsförståelsetest.....	23
5	Utvärdering.....	24
5.1	Presentation av undersökningen.....	24
5.2	Analys och resultat: hypotesprövning.....	24
5.3	Retrospektiv poweranalys	32
5.4	Slutsatser.....	34
6	Avslutande diskussion.....	36
6.1	Sammanfattning.....	36
6.2	Diskussion	37
6.2.1	Trovärdighet och koppling till antal försökspersoner.....	37
6.2.2	Val av försökspersoner och språk.....	38
6.2.3	Tekniska förbättringsmöjligheter	39
6.2.4	Forskningsetiska aspekter och genusaspekter	39
6.2.5	Samhällelig nytta	40
6.3	Framtida arbete	40
6.3.1	Kortsiktiga möjligheter	40
6.3.2	Långsiktiga möjligheter	41
	Referenser	43

1 Introduktion

Läsförståelse är att kunna ta till sig betydelsen av lästa texter, och är en förutsättning för att kunna lära sig genom det skrivna ordet. Om man inte kan ta till sig texter eller förstå frågor relaterade till en text är det svårt att ge rätt svar även om man sitter inne med kunskaperna. Eftersom det i dagens samhälle är viktigt med en god läsförståelse är det ett viktigt område inom skolundervisningen (Oakhill, Cain & Elbro 2015).

Engagemang och motivation är viktiga drivkrafter för inläring. Dataspel är konstruerade för att skapa engagemang (Jennett et al. 2008). Förhoppningen är att dataspel anpassade för undervisning ska kunna användas för att höja studiengagemang. En del studier, som de av Rosas et al. (2003), Gros (2007), och Faizal (2016), tyder på att dataspel kan vara bättre för inläring jämfört med traditionell bokbaserad undervisning.

Det finns bra exempel på spel där läsförståelse är en avgörande komponent. Phoenix Wright: Ace Attorney (förkortat PW; 2001) och Danganronpa: Trigger Happy Havoc (förkortat DR; 2014) är båda kända exempel på spel (specifikt "visual novel"-spel) där spelaren följer en förberedd mysterieberättelse genom att hitta lögnar och brister i andra karaktärs påståenden. Spelmekaniskt sett måste spelaren använda sin läsförståelse för att avgöra vilka av deras bevis som kan motbevisa de andra karaktärernas påståenden i långa dialogtunga scenarion som kan liknas vid rättegångar.

Spelen använder dock tydligt åtskilda mekaniker när spelaren ska presentera ett bevis mot ett felaktigt påstående. Medan PW inte kräver mer än att du parar ihop ett bevis med ett påstående från en meny använder DR ett system där spelaren skjuter en "beviskula" på den markerade "svaga punkten" i påståendets dialogtext, samtidigt som denna text rör sig på skärmen i olika mönster.

Detta arbete undersökte ifall spelmekaniken bidrar med en noterbar påverkan på spelarens förmåga att ta till sig textens innehåll. En mordgåta skrevs och implementerades i olika spelsystem som passar för ett rättegångsscenario. Dessa system framställdes i Unity 3D (Unity Technologies 2005) och använde mekaniker från PW och DR. Sedan rekryterades försökspersoner som delades in i grupper för att försöka lösa mordgåtan via ett av dessa mekaniksystem, eller att bara läsa mordgåtan som text. Effekten på spelarens läsförståelse som beror på systemens skillnader mättes genom spelarnas totala speltid och genom deras svar på ett läsförståelsetest de fick fylla i efter att de löst gåtan.

2 Bakgrund

Läsförståelse är själva meningen med att läsa. Att kunna ta till sig betydelsen av skrivna texter är en förutsättning för att kunna lära sig genom det skrivna ordet (Oakhill, Cain & Elbro 2015). Oakhill med flera inleder sin bok "Understanding and Teaching Reading Comprehension" med det passande citatet från den brittiska filosofen Edmund Burke (1729–1797), "Reading without reflecting is like eating without digesting". Likaledes, om man inte förstår innebörden av en skriftlig fråga är det svårt att ge rätt svar även om man sitter inne med kunskaperna.

Utöver att i direkt bemärkelse kunna kommunicera på ett språk har läsförståelse en mycket bredare betydelse. En god läsförståelse är basen för lärande även i andra ämnen utöver språk och av betydelse för framgång med studier och i arbetslivet. Det finns en tydlig korrelation mellan socioekonomisk ställning och läsförståelse (National Research Council 1998). Eftersom läsförståelse är en så grundläggande byggsten för lärande så är det viktigt att under skoltiden identifiera elever med svag läsförståelse för att kunna ge dem extra stöttning. Ett flertal olika metoder för att testa läsförståelse har därför utvecklats. Några av dem finns beskrivna nedan.

2.1 Om läsförståelsetest

Testmetoder för läsförståelse inkluderar grovt sett att muntligt sammanfatta innehållet, att korrekt svara på flervalsfrågor, att välja ord som fyller luckor i texten och att associera texten med bilder. Vid studier där elever fått genomgå flera av dessa testmetoder var det i alla årskurser få elever som presterade väl på alla fyra sorters test (Keenan & Meenan 2014), vilket tyder på att testmetoden i sig kan ha stor inverkan på vilka testtagare som får tillfredsställande resultat respektive diagnostiseras för bristande läsförståelse.

Keenan & Meenan (2014) listar fyra väl använda exempel på läsförståelsetest, kallade efter förkortningarna GORT, PIAT, WJPC, och QRI. I dessa fyra sorters test läser de testade först texten innan de får chansen att visa att de förstått textens innehåll. Testmetoden som används för att visa denna förståelse är den största skillnaden mellan de fyra testen.

Kort sammanfattat är det karaktäristiska för respektive test:

GORT - Grey Oral Reading Test (Wiederholt & Bryant 1992 se Keenan & Meenan 2014, s. 5). Den testade läser en mängd kortare textstycken och får svara på flervalsfrågor efter varje stycke. Frågorna läses högt av den testansvarige, vilket innebär att den testades förmåga att läsa text inte påverkar hans förmåga att förstå frågorna. Denna typ av test är anpassade för barn som nyligen lärt sig läsa.

PIAT - Peabody Individual Achievement Test (Dunn & Markwardt 1970 se Keenan & Meenan 2014, s. 5). Den testade läser en enskild mening. Därefter tas texten bort och den testade får sedan titta på fyra olika bilder och ska sedan välja den som bäst representerar meningen hen läst.

WJPC - Woodcock Johnson Passage Comprehension (Woodcock, McGrew & Mather 2001 se Keenan & Meenan 2014, s. 5). De testade läser en till två meningar långa textstycken som innehåller tydliga ordluckor. Ordluckorna ska fyllas i med rätt ord för att visa att testdeltagaren förstått texten.

QRI - Qualitative Reading Inventory (Leslie & Caldwell 2021); beskrivet i mera detalj i sektion 2.1.1. eftersom ett QRI-inspirerat test användes i denna studie. Den testade läser en text som ofta är längre (cirka 250 till 1000 ord) än för de andra testmetoder som nämnts ovan. Hen återberättar sedan med egna ord vad texten handlar om. Vid återberättandet noterar den testansvarige hur många "idea units" (ungefär meningsfulla informationsbärande punkter) den testade lyckas återberätta. Testtagaren svarar även skriftligt på ett flertal ordkunskapsfrågor och faktafrågor baserade på textens innehåll. Frågorna är valda så att det finns ett tydligt rätt svar så att det inte ska vara svårt att rätta testet. Jämfört med de andra test som nämnts här kan svårighetsgraden på ett QRI-test lätt anpassas, vilket gör att det fungerar för olika åldrar och läsförmågor (Keenan & Meenan 2014, s. 5).

Läsförståelse är en komplex process men kan grovt delas in i två komponenter. Den första är "decoding", vilket innebär förmågan att läsa enstaka ord utan kontext. Den andra är den språkliga förmågan att förstå betydelsen av de individuella orden och meningarna de bildar (Oakhill, Cain & Elbro 2015). För yngre barn är "decoding" av stor betydelse, men betydelsen minskar allt eftersom de blir äldre. Några av läsförståelsetesten, som PIAT och WJPC, fokuserar till stor del på förmågan till "decoding" (Keenan, Betjemann & Olson 2008). Generellt så var överensstämmelsen mellan de fyra olika testen rätt dålig. I genomsnitt var överensstämmelsen mellan de olika testen 43% vad gäller identifiering av de 10% lägst presterande eleverna. Korrelationen var något bättre för de yngre eleverna (Keenan & Meenan 2014).

I slutet av varje kapitel i DR, utsätts spelaren för en så kallad "Closing Argument"-sektion. Dessa har likheter med faktafrågorna i ett QRI-test med spelets innevarande kapitel som ämne och är en anledning till att ett QRI-inspirerat test valts för denna studie.

2.1.1 QRI-test

Ett QRI-test (Leslie & Caldwell 2021) är till för att bedöma flera olika aspekter av försökspersoners, vanligtvis skolelevs, läsförståelse för att kunna anpassa vilken svårighetsgrad på texter eleven i fråga bör få i framtida undervisning och om extra stöd eventuellt behövs.

Enligt QRI-metoden graderas vilken förståelse en elev har av en specifik text som "independent" (självständig), "instructional" (handledd) och "frustration" (frustration). Det huvudsakliga syftet med ett QRI-test är att fastställa vilken svårighetsgrad på texter som motsvarar "instructional", det vill säga den nivå eleven klarar av att förstå med stöd av en lärare. "Independent" kan direkt översättas till att eleven självständigt kan ta till sig textens innehåll. "Frustration" motsvarar som namnet beskriver en för svår nivå på texten som gör att eleven inte har möjlighet att tillgodogöra sig den.

Vanligtvis inleds ett QRI-test med att en elev blir testad för hur många ord hen kan från en ordlista så att testledaren grovt kan uppskatta vilken text-svårighetsgrad som är lämplig att starta det egentliga testet med. Testledaren väljer en text som är graderad till samma svårighetsgrad som den svåraste ordlista eleven har minst 90% rätt på.

QRI-texter inkluderar både berättande och förklarande texter från förskolenivå upp till gymnasienivå. Bland de förklarande texterna finns både sociala studier och vetenskapliga texter. En text från QRI-7 (level 4), "Amelia Earhart", finns kopierad i Appendix C för att visa på upplägget i en typisk text till ett QRI-test (Leslie & Caldwell 2021, s. 240). Exempeltexten är en faktatät narrativ biografi. Det finns också texter i QRI-7 som är av annan karaktär. Till

exempel "A Special Birthday for Rosa", som är en påhittad historia och där dialoger mellan karaktärer förekommer (Leslie & Caldwell 2021, s. 210). Texter på de högre nivåerna är typiskt informationstäta förklarande texter. Se till exempel "World War I" från QRI-7 (Leslie & Caldwell 2021, s. 345). Ett QRI-test är vanligtvis från 250 ord för de lättare nivåerna upp till cirka 1000 ord på högsta nivån.

Innan läsningen av en text inleds ställer vanligtvis testledaren tre till fyra konceptfrågor om texten för att skaffa sig en uppfattning om försökspersonen känner till ämnet sedan tidigare. Läsning av en text som handlar om ett bekant ämne gör vanligtvis så att personen presterar på toppen av sin förmåga.

Efter att eleven läst texten återberättar hen innehållet med egna ord. Testledaren noterar hur många informationsbärande punkter eleven korrekt lyckas beskriva. För att det inte ska bli subjektivt så finns rättningsmallar i form av fördefinierade listor med poänggivande punkter listade. Varje rätt informationsbärande punkt ger en poäng. Nästa steg är att eleven svarar på direkta (explicita) och härledda (implicita) faktafrågor. Vanligtvis skriver försökspersonen in svaren som text på öppna frågor. Frågorna är valda och formulerade så att det finns tydligt korrekta (respektive felaktiga) svar. Varje rätt svar ger en poäng. Poängen räknas sedan samman (inklusive en bedömning av uttal för yngre elever) och jämförs med en skala för att bestämma om texten är på "independent", "instructional" eller "frustration" nivå för eleven i fråga.

2.2 Inlevelse som motivation

De allra flesta dataspel är designade för att innehålla någon form av utmanande moment (Juul 2003). Det finns en stor variation vad detta utmanande moment kan bestå av. Detta har gett upphov till olika genrer, alltifrån spel som fokuserar på ren motorisk skicklighet (till exempel Guitar Hero; 2005) till intellektuell kontextberoende problemlösning (till exempel Myst; 1993). Utmaningen kan i sig användas som en stark motivator för att spelare ska engagera sig i upplevelsen, men det är inte det enda sättet att uppmuntra en spelare att tackla en utmaning. Andra sätt inkluderar till exempel att vädja till spelarnas nyfikenhet genom att presentera teman, mysterier, eller berättelser som väcker deras intresse (Bateman 2009), utöver den grafik, ljud och musik som använts till att förstärka upplevelsen.

Konceptet med motivationshöjande tillägg för att öka engagemang har flera gånger använts i studiesammanhang, då mycket undervisning och inläring också går ut på att utmana de studerandes problemlösningssförmåga. Ett område där inlevelse (på engelska immersion) länge använts för att förhöja inläringen är inom undervisning av främmande språk (Johnson & Swain 1997). Förenklat placeras eleverna/studenterna i en miljö där undervisningen helt och hållet sker på det främmande språket.

Förhoppningen är att dataspel, rätt anpassade för undervisning, ska kunna användas för att höja studieengagemang. Detta kan antas förbättra inläring jämfört med traditionell bokbaserad undervisning, och studier som de av Rosas et al. (2003), Faizal (2016), Gros (2007), Jere-Folotiya (2014) och Liu och Chu (2010) tyder samtliga på att så är fallet. Till exempel så jämförde Faizal (2016) hur studenters förmåga att konversera på engelska utvecklades när de antingen använde ett "visual novel"-spel eller en bok som läromedel. Efter studierna gavs studenterna i de två grupperna samma test för att utvärdera deras kunskapsutveckling. Resultaten visade att gruppen som använt "visual novel"-spelet presterade bättre än kontrollgruppen som använt traditionell bokbaserad undervisning. Detta

resultat står i kontrast mot studien av Buser och Peter (2012) som visar på att arbeta med flera uppgifter parallellt är mindre effektivt än en i sänder. Denna kontrast tyder på att de extra aspekter som utgjorde spelet i Faizals (2016) studie, så som ljud, bild och musik, inte är en tung uppgift som distraherar från textläsning, utan fungerar som ett tillägg som ger texten relevant kontext.

Baserat på efterföljande intervjuer spekulerar Faizal (2016) att utmaningarna och äventyret i "visual novel"-spelet stimulerade studenterna att arbeta igenom spelet och även associera utbildningen till något underhållande, men kunde inte observera en skillnad i motivation hos studenterna i sin studie. I en separat studie (Lopez-Morteo & López 2007) utvecklades och användes en virtuell studiemiljö för undervisning av matematik. Syftet var att etablera en motiverande miljö kallad "Interactive Instructors of Recreational Mathematics" som skulle göra eleverna mer positivt inställda till matematik. Den virtuella studiemiljön och matematikspelen testades på mexikanska gymnasieelever och resultaten visade positiva effekter på elevernas attityd till matematik. Studien testade inte effekten på förbättring av kunskaper i matematik, men författarna spekulerar att den motivationshöjande aspekten har potential att förbättra inlärningsprocessen.

De flesta studier som undersökt dataspel som läromedel pekar på positiva effekter för antingen studenters inställning till ämnet eller resultat som betyg. Bland utmaningarna nämns dock att det kan ta mycket tid i anspråk både för lärare och elever att lära sig hantera undervisningsspelet i fråga (Gros 2007). Det finns också exempel där det inte kunnat visas positiva effekter på prestation men de tenderar att vara äldre. Exempel inkluderar en studie med matematikspel för Sony Playstation (Din & Caleo 2000).

2.3 Om spelmekaniker relevanta för denna studie

"Visual novel"-spel är dataspel som huvudsakligen består av textbaserade berättelser som stöds av interaktiva moment och bilder, ofta tecknade. Som namnet antyder kan dessa spel beskrivas som bokromaner förstärkta av andra medier (Cavallaro 2010). Liksom inom litteraturen finns spel inom visual novel genren med många olika typer av berättelser. Vart och ett designat med spelmekaniker anpassade för sin specifika berättelse. De kan till exempel innehålla prompter som låter spelaren påverka berättelseflödet, värden som graderar vad de olika spelkaraktärerna tycker om spelarens karaktär, eller minispel menade att simulera specifika delar av upplevelsen.

Liksom andra typer av spel behöver visual novel-spel ha grundläggande egenskaper för att göra dem intressanta nog för att skapa inlevelse hos en spelare. Flera delar ingår i detta. Bland det mest grundläggande är att spelet behöver ha ett tydligt, lättbegripligt mål (Schell 2008, s. 113-128). Som exempel på detta kan nämnas att lösa en deckargåta. Utöver ett tydligt mål behöver spelet också vara lagom utmanande för att spelaren ska känna tillfredsställelse och vilja fortsätta spela spelet. En lagom nivå av utmaning innebär den svåra balansgången mellan att det är frustrerande svårt och att det är för uttråkande lätt. Rätt balans i svårighetsgrad kombinerat med direkt feedback hjälper spelaren att hålla fokus och vilja att fortsätta spela (Schell 2008, s. 118-128). Allt eftersom spelaren blir bättre behöver svårighetsgraden anpassas så att spelet inte uppfattas uttråkande lätt.

För att göra visual novel-spel intressanta används ofta kontextberoende problemlösning på narrativa texter. I spelsammanhang används ofta ordet pussel på sådan problemlösning där spelaren behöver stanna upp för att tänka (Schell 2008, s. 207-219). Schell (2008) beskriver

att det för bra pussel generellt är lätt att förstå målet med pusslet och att det är lätt att komma igång med det. Vidare behöver spelaren känna att det går framåt och i slutändan kommer vara lösbart för att motivationen ska hållas uppe.

I studierna av Faizal (2016) och Liu och Chu (2010) användes visual novel-liknande spelmiljöer för att stimulera studenter till bättre inläring av språk. Båda studierna pekade på att de virtuella miljöerna kunde förbättra studenternas resultat i engelska språktest. Miljöerna och spelen i dessa studier är komplext uppbyggda och går utanför ramarna för ett examensarbete. Icke desto mindre inspirerar resultaten till att med hjälp av immersionshöjande visual novel-spel undersöka om engagemang i en text via logikpussel har betydelse för hur försökspersoner förstår och tar till sig information.

Det finns bra exempel på spel där läsförståelse är en avgörande komponent. Phoenix Wright: Ace Attorney (förkortat PW; 2001) och Danganronpa: Trigger Happy Havoc (förkortat DR; 2014) är båda kända exempel på visual novel-spel där spelaren måste använda sin läsförståelse för att klara sig vidare i spelet. Båda dessa spel är relevanta för denna studie eftersom det utöver elementen av läsförståelse är görbart att utveckla spel liknande dessa inom ramen för ett examensarbete.

Phoenix Wright: Ace Attorney (PW; 2001) är en spelserie skapad av Capcom, och Danganronpa: Trigger Happy Havoc (DR; 2014) är en spelserie skapad av Spike Chunsoft. De är båda "visual novel"-spel som utgår ifrån liknande designmässiga grundprinciper för att ge spelaren invecklade mordgåtor att lösa. Utmaningen kommer i form av långa rättegångar som ger spelaren möjlighet att lösa mordgåtorna via en sekvens av isolerade logikpussel med fokus på läsförståelse, då spelaren lyssnar på de olika karaktärernas påståenden och med hjälp av sina bevis försöker avgöra vad som är sant eller falskt.

Gemensamt för både PW och DR är att spelaren ges rollen av en karaktär som behöver reda ut mordgåtor. Först får spelaren chansen att utföra en brottsplatsundersökning där de samlar förberedda bevisföremål (bestående av en förklarande text och tillhörande bild som de senare kan överblicka i en meny) genom att klicka på diverse objekt i sin omgivning. När spelaren har hittat alla bevis som behövs kan en rättegång börja. Under denna rättegång kommer det finnas speciella dialoger (kallade "Cross Examinations" i PW och "Non-Stop Debates" in DR) där spelaren måste utmana en specifik replik med ett specifikt bevis för att komma vidare i berättelsen. Det som skiljer spelen åt är inte logiken i deras pussel, utan hur många alternativ spelaren har att välja bland och vad som behöver göras för inlämning av svar.

2.3.1 Översikt av PW-systemet

I PW-systemet skrivs dialogen upp i form av dialogrutor. Därefter kan spelaren fritt bläddra igenom alla dialogens repliker fram och tillbaka efter eget behag. Hen har då även tillgång till enkla knappar som ger möjlighet att undersöka replikerna djupare. När knappen för att göra detta trycks spelas en speciell dialog upp (som i enstaka fall lägger till kompletterande repliker jämfört med den ursprungliga dialogen), och spelaren har möjlighet att presentera bevis mot en given replik.

När spelaren ska använda ett bevis i PW-systemet väljer de först vilken replik de vill säga emot från alla repliker som finns i den givna dialogen, och väljer sedan ett bevisföremål bland de som spelaren har i sin samling, se figur 1. Effekten blir ett test av logisk slutledning där spelaren först måste kunna avgöra vilket av de givna påståendena som är fel och sedan presentera bevis för varför.



Figur 1 Exempel från PW på normal textuppspelning (överst), fritt bläddrande bland repliker (mitten), samt presentation av bevis (underst).

2.3.2 Översikt av DR-systemet

I DR-systemet läses dialogen upp genom förberedda ljudinspelningar. Istället för att presenteras i textrutor spelas varje replik upp i en bestämd hastighet samtidigt som dialogtexten färdas genom rummet i dynamiska mönster menade att matcha ljudinspelningarnas samtalsflöde. Dialogen upprepas fram till en bestämd maxtid, och innan tiden går ut har spelaren chans att presentera ett bevis mot replikerna.

För att presentera ett bevis måste spelaren först välja ett bevisföremål från en bestämd delmängd ur sin totala bevissamling. Därefter kan spelaren avfyra en så kallad "truth bullet" som representerar det bevis hen valt. Om denna sedan träffar ett av de nyckelord som finns markerat i den omkringfarande texten så presenteras beviset mot den associerade texten, se figur 2. Detta tillagda actionmoment representerar en kaotisk debatt och kan tänkas distrahera spelaren från logikmomentet. För att spelet inte ska bli för svårt begränsas antalet bevis spelaren behöver välja bland.



Figur 2 Exempel från DR på text som åker runt i realtid (vänster), ett markerat nyckelord (mitten), samt en “truth bullet” (höger) som avfyras mot punkten.

3 Problemformulering

Rosas et al. (2003), Gros (2007) och Faizal (2016) menar att spel kan användas för att öka motivation för inläring, till exempel inom läsförståelse. PW och DR är två existerande spel vars utmaningar bygger på läsförståelse och är relativt populära inom denna nisch. Därmed kan det vara rimligt att använda situationer och mekaniker från PW och DR som inspiration vid utveckling av undervisningsspel med fokus på läsförståelse.

Någon som spelat både PW och DR vet att koncepten de utgår från har tydliga likheter: Rättegångar där spelarna använder läsförståelse och sitt logiska tänkande för att räkna ut vilka av spelkaraktärernas påståenden som kan motsägas med bevis. En stor skillnad mellan spelen är att spelmekanismen i DR innehåller ett tillagt motoriskt moment i samband med dessa logikuppgifter. Denna studie designades för att testa hur dessa spelmekaniska skillnader påverkar spelares förmåga att ta till sig och förstå textens innehåll. Frågeställningen var alltså: Hur påverkar olika sätt att presentera en text och interaktions-/distraktions-moment spelarens läsförståelse?

Mer specifikt fick varje försöksperson en text presenterad för sig på ett av tre olika sätt.

1. Som ren text utan interaktionsmoment eller utmaningar, liknande en bok eller ett typiskt "visual novel"-spel.
2. Med logikpussel som måste lösas för att komma vidare i texten, liknande PW.
3. Med logikpussel som utöver kravet att lösas innehåller ett motoriskt moment, liknande DR.

Hypotesen var att de som aktivt engagerat sig i texten via logikpussel kommer att prestera bättre på ett läsförståelsetest inspirerat av QRI-metoden med textens innehåll som ämne. Detta i analogi med vad Faizal (2016) observerat angående studenters utveckling av förmågan att konversera på engelska. Målet var att kvantitativt testa hypotesen jämfört med nollhypotesen att engagemang i texten via logikpussel saknar betydelse för resultatet i läsförståelsetestet. En ytterligare spekulation var att de som distraheras av mer än endast logiken i utmaningen (DR-mekanismen) kommer lösa logikpusslet långsammare och prestera sämre i samma läsförståelsetest än de som endast löser logikpusslen. Detta eftersom multitasking leder till att man löser uppgifter sämre (Buser & Peter 2012).

3.1 Metodbeskrivning

3.1.1 Utveckling av manus och spelversioner

Ett manus till ett mysterium skrevs, inklusive en mängd bevismaterial med tillhörande beskrivningar. I manuset, inspirerad av spelet Among Us (2018), har en besättning på ett rymdskepp samlats för att diskutera mordet på en av deras medlemmar och de ska räkna ut vem av dem som är mördaren. Manuset följer karaktärernas diskussion och innehåller tydligt markerade sektioner för logikpussel där en av bevisbeskrivningarna motbevisar en av karaktärernas påståenden. Manuset konverterades till spelbara format. I spelen fortsätter berättelsen inte vidare till nästa sektion förrän spelarens karaktär påvisat den felande logiken.

Strävan var att skriva ett manus som tillsammans med tillhörande testfrågor har en svårighetsgrad liknande den högsta nivån (gymnasienivå) för ett QRI-test (Leslie & Caldwell 2021).

Engelska användes i manus och efterföljande läsförståelsetest eftersom den berättelse som adapterats till manuset ursprungligen är på engelska (Nutrin 2020). Då undersökningen utfördes i Sverige förväntas att de flesta försökspersonerna har svenska som modersmål och engelska som andraspråk. Det är möjligt att val av språk för manus och test påverkade resultatet. Information om nivå på engelskakunskaper samlades inte in och eventuella samband med detta kunde därför tyvärr inte undersökas.

Uti från manuset skapades tre versioner av ett mysteriespel, kallade version VN, PW, och DR. I version VN (Visual Novel) visas texten i en dialogruta i taget. Spelaren kan växla till nästa dialogruta efter eget behag och när som helst pausa för att läsa bevisbeskrivningarna, men har utöver detta inte någon ytterligare kontroll över spelet. I version PW (Phoenix Wright) har spelaren samma kontroll som i version VN, förutom i de markerade pusselsektionerna, där de fritt kan bläddra fram och tillbaka bland dialogens repliker, men måste klicka på det motbevisbara påståendet med rätt bevisobjekt för att komma vidare i spelet. Version DR (Danganronpa) utgår ifrån samma koncept som version PW, förutom att det är svårare att klicka på påståendena under de markerade pusselsektionerna då texten rör sig över skärmen i olika mönster och dialogen spelar upp sig själv efter bestämda tidsintervaller.

3.1.2 Syftet med användning av QRI-test i denna studie och urval av försökspersoner

Försökspersoner söktes bland studenter inom Spelutvecklingsprogrammet på Högskolan i Skövde och bland deras vänner. Utöver enkelheten gjordes detta bekvämlighetsurval av försökspersoner eftersom kännedomen om spelet Among Us är hög i denna grupp, vilket ger ett fokus på läsning och lösning av mysteriet snarare än att spelarna ska ta reda på hur Among Us fungerar. Detta kan jämföras med hur ett QRI-test genomförs, där testledaren innan läsförståelsetestet börjar ställer konceptfrågor om en text för att skaffa sig en uppfattning om försökspersonen känner till ämnet sedan tidigare. Läsning av en text som handlar om ett ämne försökspersonen känner till gör vanligtvis så att personen presterar på toppen av sin förmåga (Leslie & Caldwell 2021). Då Among Us ursprungligen är ett socialt spel utan ett bestämt händelseförlopp finns det ingen risk att någon försöksperson har kännedom om handlingen som skrivits för studien även om de känner till Among Us.

Det egentliga syftet med ett QRI-test är att bedöma flera olika aspekter av försökspersoners, vanligtvis skolelevs, läsförståelse för att kunna anpassa vilket stöd och vilken svårighetsgrad på texter eleven i fråga bör få i framtida undervisning. Syftet med testet i denna studie var inte att bedöma läsförståelsenivån hos enskilda studenter. Läsförståelsetestet användes här istället för att jämföra hur lätt det är för försökspersoner att ta till sig en text beroende på hur den presenteras. För att skillnader beroende på presentationsmetod ska kunna fångas upp krävs ett tillräckligt stort antal testdeltagare. Generellt är ett slumpmässigt urval av försökspersoner att föredra vid vetenskapliga undersökningar (Aczel 2008). I denna studie har strävan inte primärt varit att försökspersonerna ska efterlikna befolkningen i stort så länge grupperna har samma genomsnittliga läsförståelse. Prioriteringen har varit att rekrytera en så stor mängd testdeltagare som möjligt med kännedom om Among Us. Rekryterade testdeltagare placerades slumpmässigt in i olika försöksgrupper. I denna studie antas att den genomsnittliga läsförståelsen är på samma nivå i de olika grupperna och har formellt inte testats här.

Varje försöksperson fick spela en av de tre spelversionerna. Strävan var att jämnt fördela antalet försökspersoner mellan de olika versionerna. Efter att ha spelat färdigt förväntades

varje försöksperson delta i ett läsförståelsetest om handlingen i spelet de just spelat. Då varje försöksperson redan löst mordgåtan under sin spelsession fokuserade frågorna i testet istället på att sammanfatta händelseförloppet för det mord som spelkaraktärerna diskuterar. Läsförståelsetestet var identiskt oavsett vilken spelversion försökspersonen använt. Antalet poäng på testet användes för att pröva hypotesen (med student t-test, Lantz 2020) om engagemang i texten via logikpussel i en stimulerande spelmiljö har betydelse för resultatet i läsförståelsetestet.

3.1.3 Utformning av QRI-liknande test

Läsförståelsetestet har följt det generella upplägget i ett QRI-test där testdeltagare efter läsning av en text med egna ord återberättar innehållet och sedan svarar på direkta och härledda faktafrågor. Det var dock inte praktiskt genomförbart inom tidsramen för detta projekt att, som i ett QRI-test, utföra individuella muntliga sammanfattningar i form av intervjuer. Försökspersonerna ombads istället att med egna ord skriftligt sammanfatta textens huvudsakliga innehåll. Det finns dock vissa svårigheter förknippade med återberättande i form av fritext. Försökspersonerna kan uppleva det som krångligt och för krävande, vilket riskerar att orsaka bortfall av deltagare. Det kan också vara svårt att objektivt poängsätta svaren. För att hantera dessa utmaningar sattes en gräns på max 2000 tecken (cirka 300 ord) fritext. Rättningen underlättades genom sökning efter nyckelord, även om manuell bedömning krävdes för kontroll att nyckelorden är satta i rätt sammanhang. Varje sådan korrekt informationsbärande punkt ger ett poäng.

Angående efterföljande faktafrågor i ett vanligt QRI-test så skriver försökspersonen in svaren som text på öppna frågor. Frågorna är formulerade så att det finns tydligt korrekta (respektive felaktiga) svar. Sådana frågor finns också inkluderade i testet i denna studie. Nackdelen med den typen av frågor är att rättningen blir komplicerad och tidskrävande. För att förenkla rättningsprocessen implementerades i de flesta fall istället flervalsfrågor. Jämfört med de öppna frågorna i ett QRI-test blir försökspersonerna på grund av flervalsfrågorna i denna studie mer styrda till givna alternativ. Till varje fråga skrevs många svarsalternativ (minst nio) för att undvika att försökspersonerna för lätt ska kunna gissa på rätt alternativ (utan att veta rätt svar). Varje rätt svar på faktafrågorna vare sig de är i fritext eller som flervalsfråga gav ett poäng var.

3.1.4 Datainsamling och hypotesprövning

Den viktigaste data som samlats in i denna studie är resultaten från det QRI-liknande läsförståelsetestet. Dessa samlades in via ett inlämningsformulär i Google forms. Antalet rätt och vilken spelversion som spelats noterades för varje försöksperson och resultaten skickades automatiskt till testledaren.

Utöver detta samlades sekundärdata in från själva spelsessionen. Under tiden försökspersonerna spelade räknades antalet felaktiga försök i spelet och speltiden mättes. När en spelare nått slutet av spelet visades denna information och hen uppmanades ta en skärmbild av den och lägga in i formuläret. Denna sekundärdata gjorde det möjligt att undersöka eventuella korrelationer mellan antalet fel, tidsåtgång och prestation på läsförståelsetestet. Då bildmotivet visades först i slutet av spelet fungerade det också som kvitto på att försökspersonen klarat spelet. Även sekundärdata hade kunnat samlas in automatiskt genom att skriva till en fil, men en enklare teknisk lösning valdes.

För att kontrollera för yttre faktorer med eventuell påverkan, som motivation eller spelvana, avslutades formuläret med att deltagarna själva på en femgradig skala graderade sin tidigare erfarenhet av visual novel-spel och deras tillfredsställelse med upplevelsen.

Denna studie är en hypotesprövning (Ejvegård 2009). Målet var att statistiskt testa ifall de personer som aktivt engagerat sig i en text via logikpussel presterar bättre på ett läsförståelsetest än de som bara läst den som text. Nollhypotesen som sattes upp är att det inte kommer finnas någon statistiskt mätbar skillnad i resultatet på läsförståelsetestet för de försökspersoner som spelat VN-versionen jämfört med de som spelat PW-versionen. Motsvarande nollhypotes sattes upp för jämförelse mellan VN-versionen och DR-versionen. Ett tvåsidigt oberoende student t-test (Lantz 2020) användes för att pröva om en eventuell skillnad mellan grupperna var signifikant. Hypotesprövningar / statistiska analyser med student t-test och F-test samt plottning av grafer gjordes med Prism (GraphPad 1994).

För att möjliggöra denna hypotesprövning var det viktigt att testresultaten rättades efter en så objektiv mall som möjligt för att inte favorisera något resultat. Därför formulerades testfrågorna på ett sådant sätt att det strikt kan definieras vad som kan räknas som ett korrekt svarsalternativ. Då det var svårt att i förväg bedöma hur stor variationen i testresultat skulle bli för personer som spelat samma respektive olika spelversioner var det också svårt att veta hur många testpersoner som behövdes för att uppnå (95%) signifikans. Denna studie har likheter i upplägget med studien publicerad av Faizal (2016) vilken hade 27 testdeltagare per grupp. Ett liknande antal personer per grupp söktes i denna studie.

Risken för falska negativa svar vid hypotesprövning kallas typ II-fel. Sannolikheten för detta kallas β , och betyder att felaktigt dra slutsatsen att två grupper inte skiljer sig åt fast de i verkligheten gör det (Aczel & Sounderpandian 2008, s.260-261). Det är möjligt att minska risken för typ II-fel genom att försäkra sig om att studien i fråga har tillräcklig power (Aczel & Sounderpandian 2008, s.264). Power definieras $1-\beta$. Studier eftersträvar ofta en power större än 0.8, vilket betyder 80% chans att hitta en skillnad som finns i verkligheten. Poweranalys för att estimerar gruppstorlekar gjordes med G*Power (Faul, Erdfelder, Lang, & Buchner 2007).

3.1.5 Etiska aspekter

Studien utfördes i enlighet med de grundläggande individskyddskraven; informationskravet, samtyckeskravet, konfidentialitetskravet och nyttjandekravet. Dessa finns väl sammanfattade i Vetenskapsrådets artikel Forskningsetiska principer (2002). Individskyddskravet: "Forskaren skall informera de av forskningen berörda om den aktuella forsknings-uppgiftens syfte". Samtyckeskravet: "Deltagare i en undersökning har rätt att själva bestämma över sin medverkan". Konfidentialitetskravet: "Uppgifter om alla i en undersökning ingående personer skall ges största möjliga konfidentialitet och personuppgifterna skall förvaras på ett sådant sätt att obehöriga inte kan ta del av dem". Nyttjandekravet: "Uppgifter insamlade om enskilda personer får endast användas för forsknings-ändamål".

Konkret betyder individskyddskraven att försökspersonerna i denna studie informerats om syftet med detta examensarbete i samband med att de ombads delta i studien. Det var helt frivilligt att delta i studien och försökspersoner kunde när som helst välja att avbryta sitt deltagande. Datainsamlingen gjordes genom frågeformulär som indexerades allt eftersom de lämnades in. Inga personuppgifter samlades in. Ingen information insamlad i denna studie

har använts eller kommer att användas för några andra syften än att slutföra detta examensarbete.

4 Genomförande

Frågeställningen i denna studie var hur olika sätt att presentera en text påverkar läsarens förmåga att ta till sig och förstå innehållet i texten. Hypotesen var att de som tvingats att aktivt engagera sig i texten via logikpussel kommer att prestera bättre på ett läsförståelsetest jämfört med bokliknande läsning. Studien undersökte också om interaktions- och distraktionsmoment påverkar läsförståelsen.

PW (2001) och DR (2014) är kända exempel på visual novel-spel där läsförståelse är en avgörande komponent för att klara sig vidare i spelet. Båda dessa spel är relevanta för denna studie eftersom det utöver elementen av läsförståelse var görbart att utveckla spel liknande dessa inom ramen för ett examensarbete. DR har ett mera komplext spelsystem än PW. Utgångspunkten för designen var att göra ett spel som till sitt upplägg liknar DR, och sedan modifiera till versionerna PW och VN.

För att presentera en och samma text på olika sätt, inklusive som "visual novel"-spel, har spelmotorn Unity 3D (Unity Technologies 2005) använts eftersom den erbjuder ett användarvänligt sätt att rendera text och bilder. Läsförståelsetexten i denna studie i form av en mordgåta inspirerades av DR (2014) och har adapterats från Among Us (2018). Läsförståelsetestet som har använts i denna studie är inspirerat framför allt av QRI-metoden (Leslie & Caldwell 2021).

Det praktiska arbetet i denna studie delades in i områdena 1) utveckling av tre spelversioner för presentation av text på olika sätt, 2) manus till logikpussel/mordgåta i debattformat att använda som text för att testa läsförståelse och 3) läsförståelsetest.

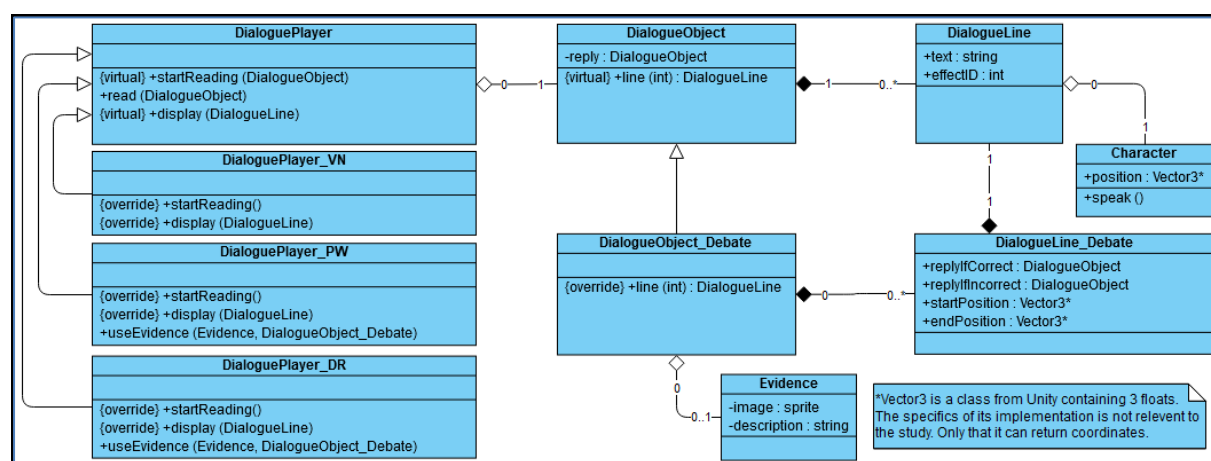
Först etablerades en gemensam spelplattform i Unity 3D för visning av text. Med spelplattformen på plats kunde support för det tänkta specialiserade debattformatet liknande det i DR skapas. Därefter delades detta system upp i de tre olika spelversioner som studien innefattar. Därefter skrevs manuset till mordgåtan som ett fristående dokument. Då texten i manuset kunde kopieras in från ett utomstående dokument var det möjligt att implementera i alla tre spelversioner samtidigt för att minimera risk för variationer och produktionsfel. Med fungerande spel på plats utarbetades testfrågorna inspirerade av QRI-metoden.

4.1 Spel för tre olika sätt att presentera text

Visual novel-spel behöver, som beskrivits i avsnitt 2.3, grundläggande egenskaper för att göra dem intressanta och skapa inlevelse. Grundläggande är att spelet behöver ha ett tydligt lättbegripligt mål, är lagom utmanande och ger direkt feedback (Schell 2008, s. 113-128). Eftersom DR uppfyller dessa grundläggande krav och fungerar väl för att presentera narrativa berättelser i dialogform användes DR som förlaga till den mest komplexa spelvarianten som användes i denna studie.

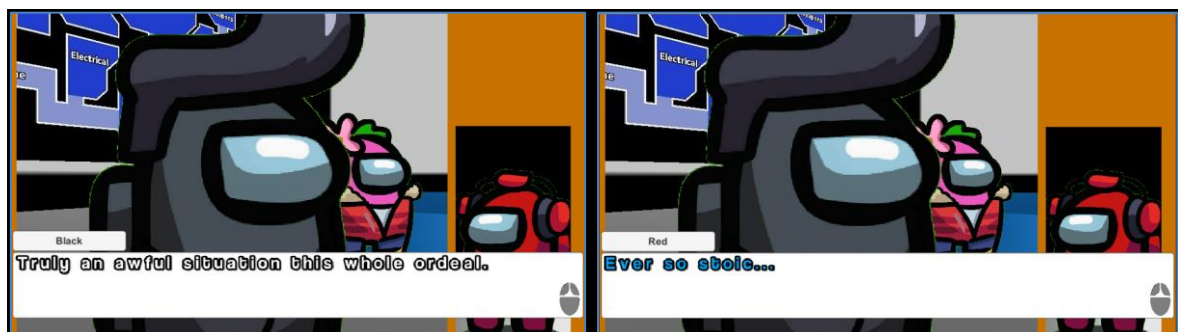
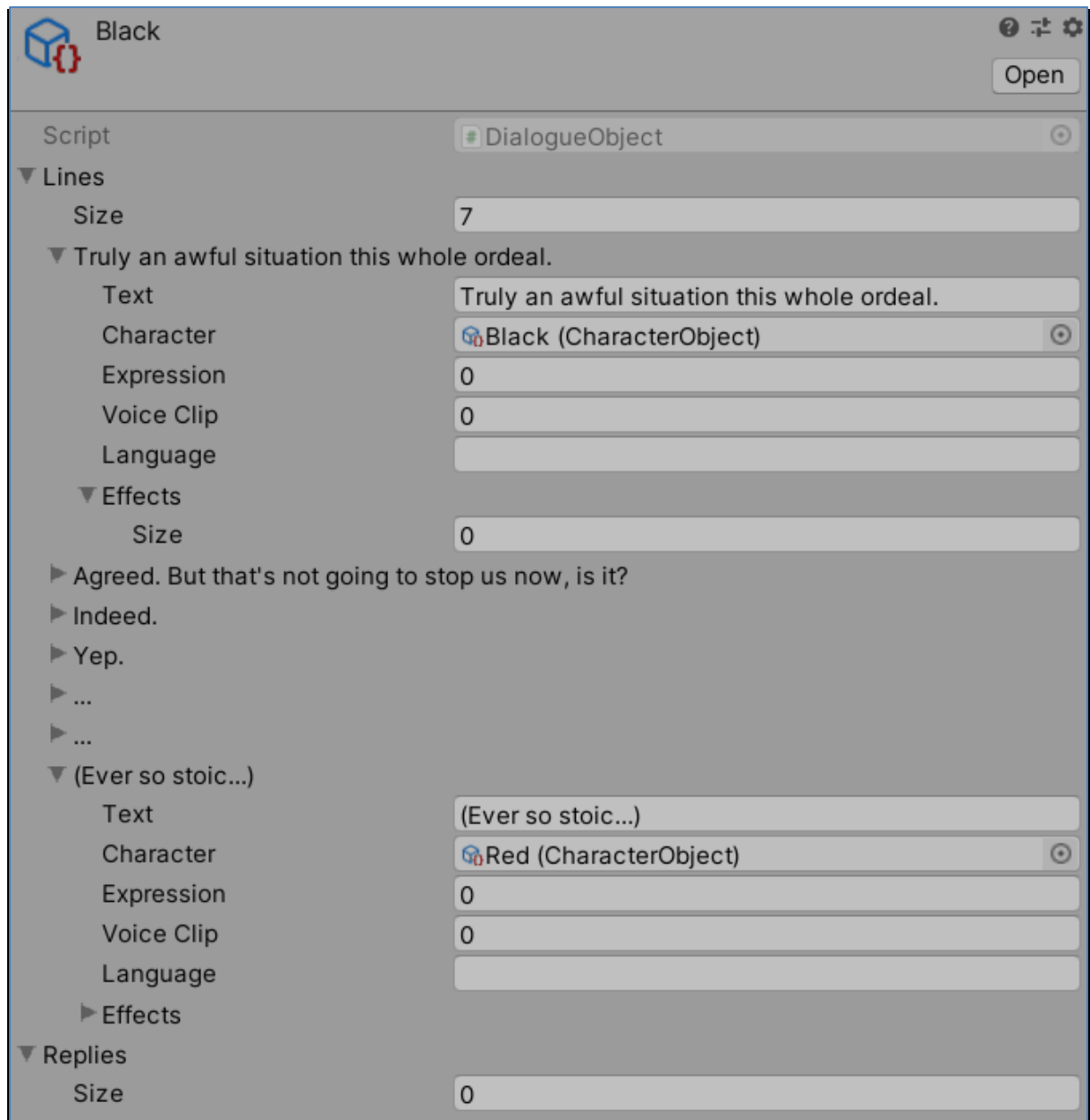
Nedan beskrivs hur ett dialogsystem skapades med hjälp av spelmotorn Unity 3D. De två viktigaste typerna av komponenter i systemet är dialogobjekt innehållande spelkaraktärernas repliker, samt dialogläsare som tar text från dialogobjekt och renderar dem i form av textrutor som spelare kan klicka fram en i taget. Dialogobjekten kan referera till varandra på så sätt att när alla repliker i en dialog har lästs upp börjar nästa dialog spelas upp automatiskt. Dialogsystemets uppbyggnad finns illustrerat i figur 3.

Det dialogsystem som utvecklats i denna studie stödjer möjligheten för ett dialogobjekt att ha flera möjliga alternativa fortsättningsdialoger för spelaren att välja mellan. För varje replik finns också möjligheten att spela olika ljudfiler och byta den talande karaktärens utseende. Dessa funktioner har dock inte använts i högre grad i spelen då tiden att implementera stilistiska detaljer var begränsad och dessa inte nödvändigtvis tillför så mycket extra till studien. Vissa effekter har dock implementerats för att tydliggöra om spelaren gett rätt eller fel svar (gäller version PW och DR).



Figur 3 Ett förenklat UML-diagram över klasserna som bygger upp dialogsystemet.

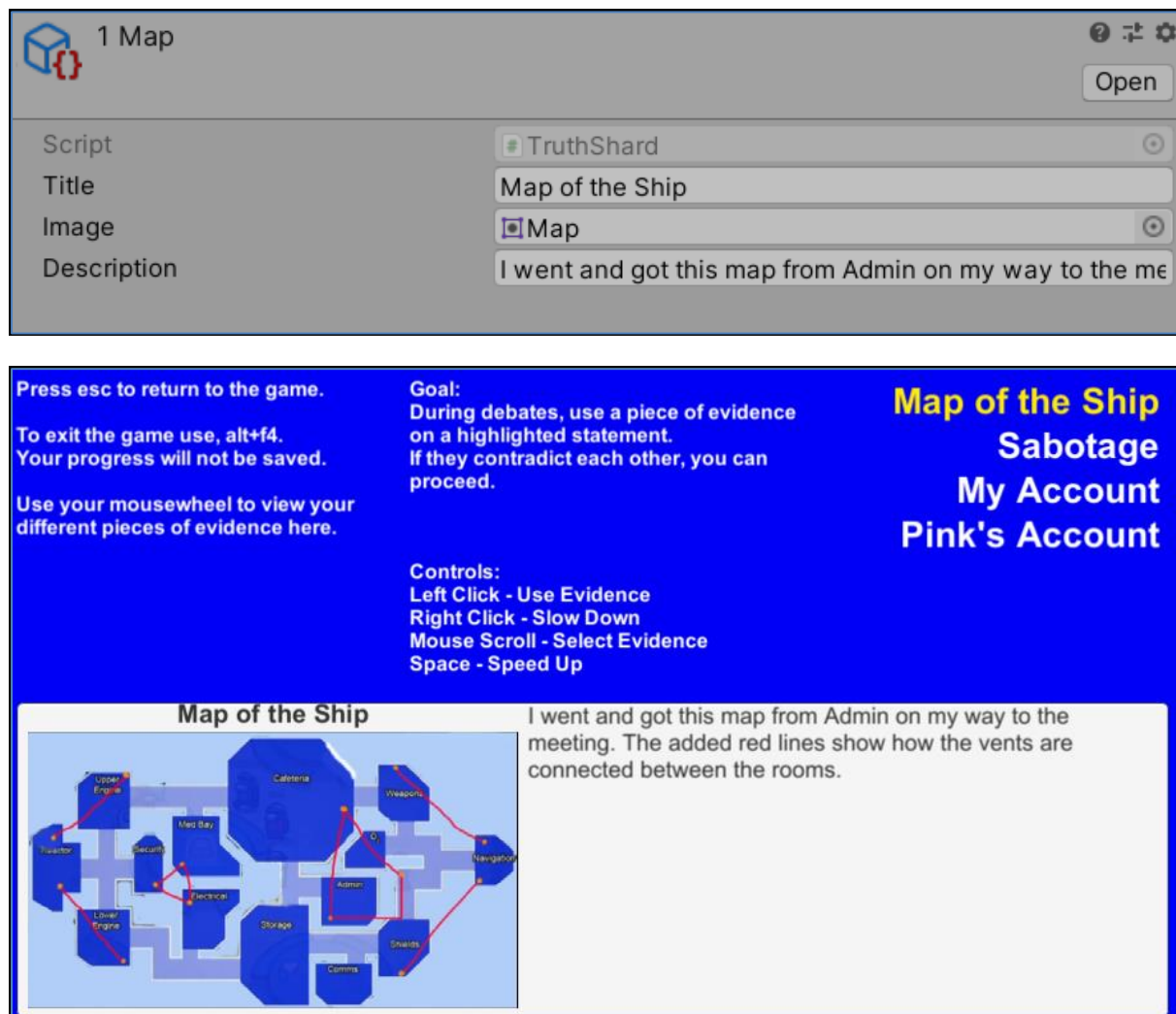
Inspirerat av utseendet i DR så är varje replik i dialogobjekten associerad till en karaktär som säger den repliken. När en karaktär säger en replik kommer kameran att flytta sig så att den talande karaktären är i fokus. Oavsett var kameran befinner sig visar dialogläsaren upp textrutorna längst fram på skärmen och kan applicera olika effekter så som radbrytning och färg på texten, baserat på om den avlästa repliksträngen har markerats för detta med särskilda tecken, se figur 4.



Figur 4 Exempel på dialogobjekt (över) och hur dess repliker visas på skärmen av en dialogläsare (under).

För att efterlikna ett DR-inspirerat scenario har bevisobjekt skapats. Ett bevisobjekt är en klass som innehåller titel, bild och beskrivningstext (förenklat kan en klass definieras som ett

avsnitt programkod som samlar en mängd relaterade attribut och funktioner som kan används till generera flera objekt). Spelaren har tillgång till en komplett lista med bevisobjekt från spelets början via en pausmeny som spelaren kan titta på vid behag, se figur 5. Utöver det innehåller pausmenyn information som ska påminna spelaren om spelets kontroller och regler.

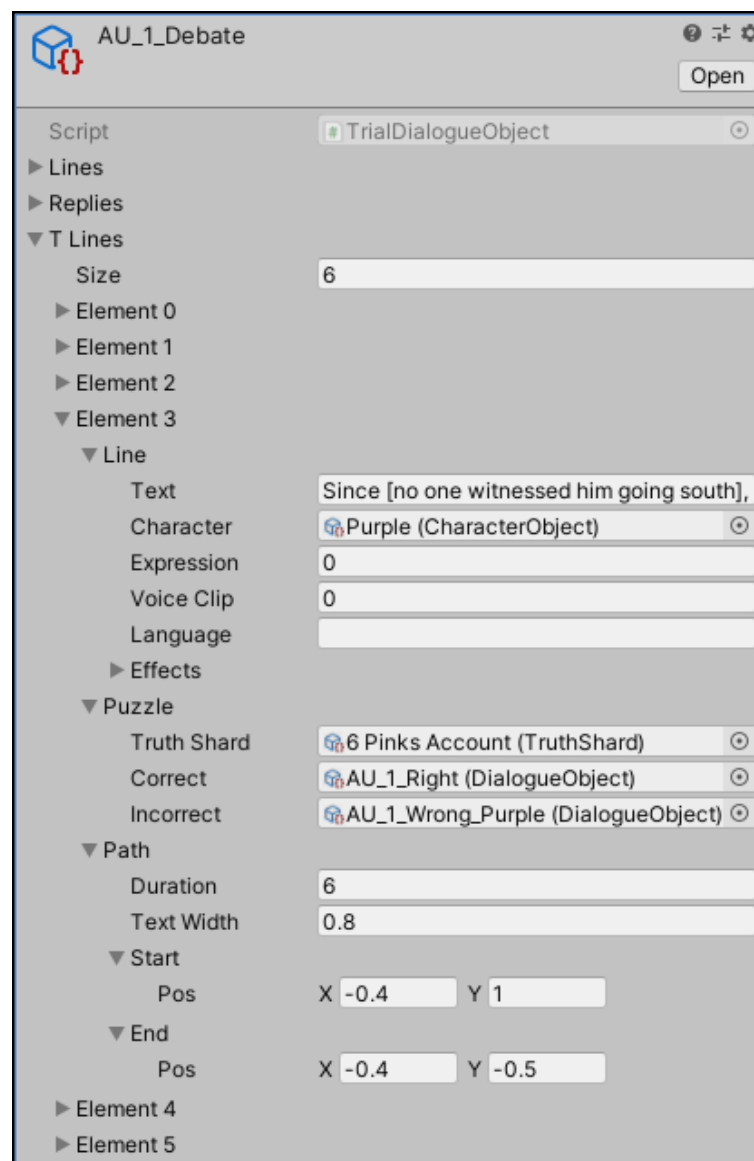


Figur 5 Exempel på bevisobjekt (över) och hur det visas i pausmenyn (under).

Från klassen för dialogobjekt utvecklades arvsklassen debattdialogobjekt. Dessa fungerar fortfarande som dialogobjekt och förblir möjliga att länka med övriga dialogobjekt men innehåller extra data för varje replik. Datan anger huruvida repliken kan bli motbevisad och i sådana fall av vilket bevisobjekt, samt vilken dialog som ska spelas upp om spelaren använder detta bevisobjekt.

Samtliga tre spelversioner (VN, PW eller DR) använder sig av samma dialogobjekt. Detta gjorde det möjligt att implementera samma text i alla tre spelversioner samtidigt. Förutom att säkerställa att texten är identisk i de olika spelversionerna effektiviserades kodningen eftersom dialogobjekten återanvändes i alla versioner. För de tre spelversionerna utvecklades dock var sin unik variant av dialogläsare. När en av dessa läsare visar upp en replik från ett vanligt dialogobjekt är det ingen betydande skillnad i utseende. Skillnaden mellan versionerna

är istället hur de hanterar och visar upp repliker från ett debattedialogobjekt. Nedan följer beskrivningar av vad som skiljer spelversionerna åt, vilket även ses i figur 6.



Figur 6 Exempel på debattedialogobjekt (över) och hur den ser ut visad på skärmen (under) i de olika spelversionerna (VN, vänster; PW, mitten; DR, höger).

Version VN hanterar dialogobjekt och debattedialogobjekt identiskt, förutom om en replik är markerad som motbevisbar. Det som då händer är att spelet förutsätter att spelaren korrekt

motbevisat repliken och direkt går vidare till att läsa upp dialogen som motsvarar att delfrågan är avklarad.

Version PW hanterar debattdialoger på ett helt annat sätt. Väl inne i en sådan dialog kan spelaren inte längre klicka fram repliker på normalt vis med enkla klick, utan måste använda sig av designerade knappar som tar spelaren framåt eller bakåt inom debattdialogen en replik i taget. Spelaren kan också välja att klicka på en markerad text för att med ett valfritt bevisobjekt försöka motbevisa dess replik. Spelaren kan inte lämna den pågående dialogen förrän hen korrekt motbevisat en replik med hjälp av ett bevisobjekt. Sammantaget gör dessa tekniska skillnader mellan PW- och VN-versionerna att spelaren är tvungen att aktivt lösa ett intellektuellt problem för att kunna ta sig vidare i texten i PW-versionen.

Version DR hanterar också uppläsning av debattdialoger på ett sätt där spelaren inte kan lämna den pågående dialogen förrän en replik korrekt motbevisats med ett bevisobjekt. I likhet med version PW motbevisas fortfarande en replik genom att klicka på dess markerade text. Skillnaden är att version DR innehåller ett motoriskt moment. Momentet går ut på att texten i sig inte är bunden till en textruta utan replikerna rör sig individuellt över skärmen i en loop tills spelaren genom ett kikarsikte skjuter på den motbevisbara repliken.

4.2 Manus till mordgåta i debattformat

Manuset till läsförståelsetexten skrevs i dialogform eftersom detta är ett berättelseformat som fungerar väl i ett visual novel-spel. Med målet att skapa en stimulerande spelupplevelse (och därigenom även en stimulerande upplevelse av texten) introducerades pusselelement (Schell 2008, s. 207-219).

Manuset till mordgåtan baseras på ett scenario från "Among Us", se appendix A. Among Us (2018) är ursprungligen ett spel som går ut på att genom sociala interaktioner räkna ut vem i gruppen av deltagande spelare som fått rollen som mördare.

Initialt var planen att manuset skulle innehålla en historia inspirerad av DR (2014). Karaktäristiskt för DR är att historien är komplex. Den har en ingående bakgrunds- och miljöbeskrivning, innehåller många karaktärer och det tar rätt lång tid tills mordgåtan presenteras första gången och diskussion om den kan starta. Detta skulle ha blivit en för lång text respektive för långt spel för försökspersonerna att arbeta igenom innan läsförståelsetestet. Ett fullödigt DR-scenario hade inte heller varit rimligt att skapa inom tidsramen för denna studie.

Among Us fungerar väl som en förenklad version av berättelse uppbyggd på samma sätt som en DR; en mordgåta placerad i ett isolerat sammanhang. Among Us innehåller dock inte samma långa bakgrundshistoria och saknar komplexa karaktärer och relationer.

En sökning på internet visade att likheterna mellan Among Us och DR tidigare uppmärksammats och använts till att skriva enklare DR-inspirerade mordgåtor presenterade som videos. En av dessa mordgåtor adapterades och användes som textmanus i denna studie (Nutrin 2020). Texten från videoserien kopierades först ner, varefter ett flertal anpassningar av texten gjordes. Långa stycken med upprepningar och fruktlösa diskussioner som inte förde gåtan framåt togs bort. För att förenkla både mordgåtan och spelmekaniken skrevs historien om så att kravet på bekräftande bevis och falska bevis togs bort. Fokus lades istället helt på bevis som motbevisade karaktärers inblandning i mordet (friade dem från mordet).

En sammanfattning av manuset lyder:

- Det har nyss skett ett mord i rummet “Electrical” ombord på ett rymdskepp.
- De nio besättningsmedlemmarna, som alla har namn efter färger, har samlats för att lösa denna mordgåta.
- Besättningsmedlemmarna inser att det måste vara någon av dem som är den skyldige.
- Ett strömavbrott gjorde att Green gick till Electrical och då hittade kroppen.
- Purple misstänker Green, då hon tycker att den som hittat kroppen är den första man borde misstänka. Hon påstår att Green gick till Electrical långt innan strömavbrottet.
- Pink vittnar att hon hörde när någon tog den väg Green beskrev sig själv ta. Eftersom ingen annan tog den vägen så bekräftar det att Green inte ljög.
- Orange anser att detta inte är tillräckligt som bevis. Han anser att Green kan ha orsakat strömavbrottet från Reactor för att locka ett offer till Electrical.
- Red vittnar då att Yellow inte blev lockad till Electrical, utan var där innan strömavbrottet utan tecken på att hon försökt åtgärda det.
- Insikten bevisar att sabotaget gjordes först efter mordet, vilket innebär att Green är oskyldig.
- Black anklagar Red med argumentet att Red var i rummet Storage, som är nära Electrical, och var den sista som såg Yellow i livet.
- Red förklarar hur mördaren inte kunde ha använt dörren till Electrical för att ta sig ut därifrån efter sabotaget. Detta eftersom dörren behöver fungerande el för att kunna stängas, och den var stängd när Green kom dit. Red bör därför inte vara misstänkt bara för att han var nära dörren.
- Pink frågar hur mördaren i sådana fall tog sig ut.
- Besättningen går igenom sina alibin och det visar sig att de alla var där de borde vara, vilket White inte gillar.
- Red kommer på att det finns en väg ut ur Electrical via ventilationssystemet. Detta gör att Blues alibi, att han satt ensam i rummet Security (som har en ventilationskanal kopplad till Electrical), inte längre håller.

Utöver dialogtexten så har spelarna tillgång till ett “Inventory” med bevismaterial. Det finns fyra olika bevis som alla är tillgängliga redan vid starten av spelet.

Spelmiljöerna i version PW och DR gör förstås så att läsningen av texten blir annorlunda jämfört med läsning av en boktext eller version VN. Viktigaste skillnaden är pusselelementen som spelaren behöver klara i versionerna PW och DR, att visa upp rätt bevis för att komma vidare till nästa steg. Det fanns risk för att spelare skulle välja att avbryta sin medverkan i studien om hen inte klarar pusselelementen och fastnar i spelet. Målet har varit att försöka balansera svårighetsgraden till en nivå som behåller spelaren motiverad snarare än frustrerad (Schell 2008, s. 113-128). Detta hanteras till exempel genom att första problemlösningssituationen i form av en fråga bara har fyra svarsalternativ att välja bland, varav ett är rätt. Om man svarar fel, får man dessutom en ledtråd (en replik från spelarens avatar Red) så att det blir lättare vid nästa försök.

Texten i denna studie är anpassad för att fungera bra i en spelmiljö. En text i dialogformat valdes eftersom det då är smidigt att implementera karaktärer som är möjliga att interagera med. Dialoger förekommer i även i QRI-test, som exempelvis “A Special Birthday for Rosa”

(Leslie & Caldwell 2021, s. 210). En annan likhet är att berättelsen i denna studie liksom “A Special Birthday for Rosa” är fiktiv.

Ett QRI-test är vanligtvis från 250 ord för de lättare nivåerna upp till cirka 1000 ord på högsta nivån som är tänkt för elever på gymnasienivå. Den slutliga läsförståelsetexten i denna studie är cirka 2000 ord lång. Det är längre än en typisk QRI-test-text. En risk med detta kan vara att det blir mycket ny information att hålla reda på innan man ser frågorna. En skillnad mellan texternas karaktär är dock att texter från QRI-test på de högre nivåerna typiskt är informationstäta, se till exempel “World War I” från QRI-7 (Leslie & Caldwell 2021, s. 345), medan det informationsbärande innehållet är glesare utspritt i texten i denna studie. Texten implementerad i spelen designades på detta sätt för att skapa en komplett spelupplevelse. En svårighetsgrad liknande den högsta nivån (gymnasienivå) i ett QRI-test eftersträvades för att testet skulle bli utslagsgivande och undvika att många försökspersoner får alla rätt, eftersom det minskar möjligheten att se skillnader mellan grupperna som spelat de olika spelen.

4.3 Läsförståelsetest om mordgåtan

Läsförståelsetest enligt QRI-metoden (Leslie & Caldwell 2021) har fungerat som inspiration för läsförståelsetestet i denna studie. QRI-test är utformade för att bedöma flera olika aspekter av läsförståelse inklusive uttal och igenkänning av enskilda ord. Uttal och ordigenkänning är inte relevanta för denna studie och har därför inte undersökts. De aspekter som tagits fasta på är istället den språkliga förmågan att förstå betydelsen av de individuella orden, meningarna de bildar, och även information som indirekt kan härledas från texten.

Läsförståelsetestet har följt det generella upplägget i ett QRI-test där försökspersoner efter läsning av en text med egna ord återberättar innehållet och sedan svarar på direkta och härledda faktafrågor.

För återberättandet implementerades ett enkelt formulär i Google forms (Google Inc. 2006) med instruktionen “Describe the incident discussed by the in game characters”. Formuläret begränsades till max 2000 tecken (motsvarande i storleksordningen 300 ord) fritext. För rättning av fritexten skapades en rättningsmall innehållande informationsbärande punkter. Dessa finns samlade i Appendix B.

Faktafrågorna i testet implementerades som flervälsfrågor i ett frågeformulär i Google forms där svaren automatiskt rättades och gjordes tillgängliga för testledaren (Google Inc. 2006). Exempel på en flervälsfråga är:

- ”Who did Green walk past during the blackout?” [Explicit]
Alternativ: Ett alternativ var för var och en av de nio besättningsmedlemmarna
Rätt svar: Pink

Samtliga testfrågor finns samlade i Appendix B. Strävan har varit att faktafrågorna ska efterlikna ett QRI-test med direkt koppling till texten. Frågor med explicita och implicita svar har blandats. Svaren på frågorna finns utspridda över hela texten.

4.4 Pilottester och därpå följande modifieringar

Under utvecklingen av spelen och läsförståelsetestet gjordes en pilotundersökning för att säkerställa att de var lämpligt utformade för att kunna ge svar på frågeställningarna i studien,

som framför allt var att statistiskt testa ifall de personer som aktivt engagerat sig i en text via logikpussel presterar bättre på ett läsförståelsetest än de som bara läst den som text. Tre personer kallade testpiloter engagerades i pilotundersökningen, och de fick provspela spelen och göra efterföljande läsförståelsetest. Deras testkörningar observerades i realtid. Efter genomförda tester intervjuades testpiloterna och de framförde synpunkter på spelupplevelsen, brister i speldesign, tekniska buggar, manus och svårighetsgrad på läsförståelsetestet.

4.4.1 Justeringar med avseende på speldesign och buggfixar

Testpiloterna återkopplade att det saknades tydliga kontrollbeskrivningar för vilka manövrer spelarna kunde använda sig av. Därför skrevs sådana beskrivningar och lades till i pausmenyn. Detta kan ses i figur 5.

En annan återkoppling var att spelarna inte självklart uppfattade om de valt rätt replik i debatterna i versionerna PW och DR. Det var viktigt att det tydliggjordes, eftersom val av rätt replik i en debatt är den sak som tar en spelare vidare i spelet. Detta tydliggjordes genom enkla visuella effekter (skakande kamera vid fel svar) och ljudeffekter ("pling" och "krash" vid rätt respektive fel svar) som spelare lätt kan associera till rätt respektive fel svar.

Specifikt i version DR infördes en mekanik där spelarna fick möjlighet att justera hastigheten med vilken den rörliga texten rör sig på skärmen. Detta infördes för att underlätta det motoriska momentet att sikta på texten (långsammare hastighet). En annan anledning var att minska väntetiden då spelaren intellektuellt löst frågan och vet vilket svar som ska väljas men motoriskt missat det och väntar på att det rätta svaret ska dyka upp igen på skärmen (högre hastighet).

Vid pilottesterna upptäcktes att renderingen av karaktärer och bakgrund skedde i fel ordning på vissa datorer. Det såg korrekt ut på den dator som användes för att programmera spelversionerna, men på vissa testpiloters datorer renderades bakgrunden över karaktärerna. Detta åtgärdades efter upptäckt.

4.4.2 Justeringar av manus till mordgåtan

Ett antal anpassningar av manuset gjordes efter pilotundersökningen. En justering var att texten kortades ner från drygt 3000 ord till cirka 2000 ord. En anledning till det var att närma sig längden på texter som vanligtvis förekommer i QRI-test. I samband med det togs många detaljer bort som inte är väsentliga för mordgåtans lösning. Syftet med det var att minska komplexiteten hos mordgåtan och ge en bättre sammanhållen berättelse utan mängder av distraherande detaljer och krångliga resonemang.

Som kuriosas kan nämnas att som del av processen att korta ner texten ändrades den skyldige från Orange till Blue. Detta genom att två hela debatter som syftade till att bevisa att Blue var oskyldig togs bort. I den uppdaterade versionen är berättelsen skriven så att Blue är den enda karaktären som saknar alibi samtidigt som ett alibi introducerats till Orange (Purple kunde bekräfta att Orange varit tillsammans med henne när strömmen gick och därigenom inte kunde vara skyldig).

4.4.3 Justeringar av läsförståelsetest

Efter att en första version av manus och läsförståelsetest gjorts, gjordes en noggrann jämförelse mellan dem och befintliga QRI-test (Leslie & Caldwell 2021). Ett antal justeringar gjordes i läsförståelsetestet så att det till högre grad skulle efterlikna ett typiskt QRI-test:

- En sammanfattningsdel infördes. Försökspersonerna ombeds i den finala versionen att med egna ord skriftligt sammanfatta händelseförloppet i berättelsen med max 2000 tecken (cirka 300 ord). I tidigare versioner fanns ett större antal flervalsfrågor, men några av dessa ersattes med den nya sammanfattningsdelen. Rättningen av sammanfattningsdelen gjordes genom sökning efter nyckelord och kontroll att dessa var satta i rätt sammanhang. Varje sådan korrekt betydelsebärande punkt ger ett poäng.
- Frågor infördes där korta manuellt inskrivna svar efterfrågades. Exempel på en sådan fråga är, "Why did the victim go to Electrical?". Exempel på godkända svar är, "she had to calibrate the distributor" och "she had to do her tasks". Svar som inte räknas som godkända inkluderar, "she was going to fix the lights".
- Detaljminnesfrågor togs bort för att ge testet mer karaktär av läsförståelsetest och mindre karaktär av minnestest. Exempel på en fråga (flervalsfråga) som togs bort är, "which rooms do we know Purple has been in?".

I den slutliga versionen av läsförståelsetestet är maxpoängen 23, varav 9 på sammanfattningsdelen, 10 på flervalsfrågorna och 4 på frågor för manuellt inskrivna svar.

5 Utvärdering

Huvudsyftet med denna studie var att testa hur olika sätt att presentera en text påverkar spelares förmåga att förstå innehållet i texten. Texten presenterades som två olika versioner av ett visual novel-spel med olika spelmekanik eller i ett format som liknar löpande text. 29 försökspersoner deltog i studien. Efter att försökspersonerna spelat någon av de tre versionerna fick de göra ett läsförståelsetest med textens innehåll som ämne. Hypotesen är att de som aktivt engagerat sig i texten via logikpussel kommer att prestera bättre på efterföljande läsförståelsetest än de som läst texten i det mera bokliknande formatet. Nedan presenteras och analyseras resultaten från studien.

5.1 Presentation av undersökningen

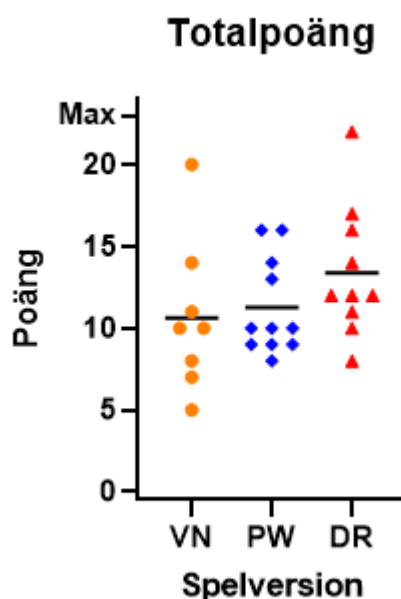
Försökspersoner rekryterades genom upprepade annonseringar på sociala medier framför allt i grupper relaterade till spelutveckling vid Högskolan i Skövde. Försökspersonerna slumpades baserat på födelsedatum till att spela antingen version VN, PW eller DR. En person från grupp PW exkluderas på grund av att hen inte fullföljde hela testet. Resultaten från totalt 29 försökspersoner inkluderades i dataanalysen. 8 personer från version VN, 11 personer från version PW och 10 personer från version DR. Tiden det tog för försökspersonerna att spela igenom spelet mättes. Försökspersonerna gjorde sedan läsförståelsetestet direkt efter att ha avslutat spelet.

Läsförståelsetestet började med en sammanfattningsdel i fri text där varje betydelsebärande punkt från manuset gav en poäng var, totalt max 9 poäng. Sedan följde 14 stycken explicita och implicita frågor (flervals eller korta textsvar) som var och en kunde ge 1 poäng. Maxpoängen på testet var således 23 poäng. Antalet och fördelningen av poäng liknar upplägget i QRI-test (Leslie & Caldwell 2021). Samtliga frågor finns samlade i Appendix B.

I anslutning till testet angav försökspersonerna också på en skala från 0 till 5 (där 0 betyder ingen/inget och 5 är bäst/mest) vilket nöje de haft av spelet, vilken tidigare erfarenhet de har av visual novel-spel och hur bekanta de är med Among Us.

5.2 Analys och resultat: hypotesprövning

Totalt inkluderades 29 försökspersoner i analysen av läsförståelsetestet. Figur 7 visar totala antalet poäng som uppnått på läsförståelsetestet uppdelat på de olika spelversionerna. Medelvärde och standardavvikelsen i antalet uppnådda totalpoäng för försökspersonerna som spelade respektive spelversion är summerade i Tabell 1. Antalet uppnådda poäng användes för att statistiskt testa ifall de personer som aktivt engagerat sig i en text via logikpussel (versionerna PW och DR) presterat bättre på ett läsförståelsetest än de som bara läst den som text (version VN).



Figur 7 Totalpoäng på läsförståelsetestet uppdelat per spelversion. Varje punkt representerar data från en enskild försöksperson. Horisontella streck representerar medelvärdet för respektive grupp.

Tabell 1 Medelvärden (M) och standardavvikelser (SD) för försökspersonernas totalpoäng, poäng på sammanfattningsdelen (Sammanfattning) och poäng på enpoängsfrågorna (Frågor) från läsförståelsetestet. M och SD finns också angivet för den tid i sekunder det tog för försökspersonerna att spela respektive spelversion och nöjet de upplevde på en skala från 0 till 5.

	VN		PW		DR	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Totalpoäng	10.6	4.7	11.3	2.9	13.4	4.0
Sammanfattning	2.4	3.0	2.5	2.4	3.0	2.6
Frågor	8.3	2.9	8.8	2.4	10.4	2.7
Tid (s)	977	253	1802	1108	1639	654
Nöje	3.3	1.0	4.4	0.7	3.4	1.1

Den första nollhypotesen som sattes upp vid planeringen av studien är att det inte finns någon statistiskt mätbar skillnad i resultatet på läsförståelsetestet för de försökspersoner som spelat version VN jämfört med de som spelat version PW. Motsvarande nollhypotes är att det inte heller finns någon statistisk skillnad i resultat för de som spelat version VN jämfört med de som spelat version DR.

Ett oparat två-sidigt student t-test (Lantz 2020) användes för att pröva om nollhypoteserna kunde förkastas. En nollhypotes förkastas om skillnaden mellan medelvärdena för två

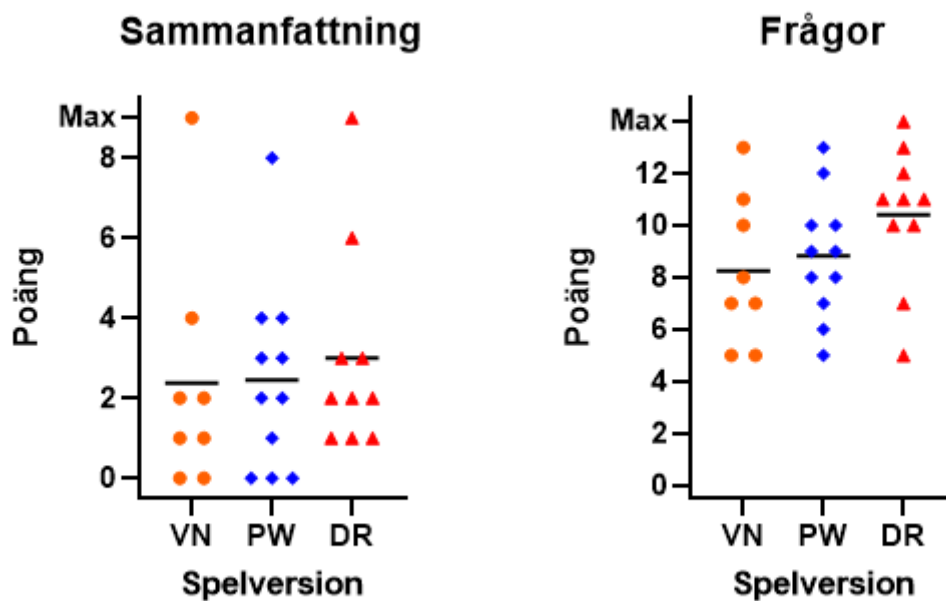
jämförda grupper är signifikant. Signifikansnivån (α) för att betrakta skillnader som signifikanta sattes till 0.05 (vilket innebär en risk på 5% att nollhypotesen felaktigt förkastas).

För totalpoängen var skillnaden mellan medelvärdena för versionerna VN och PW inte signifikant; $P=0.71$. Därmed följer att nollhypotes nummer 1 inte förkastas. Skillnaden för totalpoängens medelvärde mellan versionerna VN och DR var inte heller den signifikant; $P=0.19$, även om värdet närmar sig signifikansnivån. Därför kan inte heller nollhypotes 2 förkastas. Inte heller någon signifikant skillnad i medelvärde mellan versionerna PW och DR kunde påvisas; $P=0.18$. P-värdena för dessa och andra jämförelser av medelvärden med student t-test finns summerade i Tabell 2.

Tabell 2 Sammanställningar av värden från oparade tvåsidiga t-test för jämförelser av spelversioner. * indikerar att en signifikansnivå <0.05 uppnåts, vilket betraktas som att medelvärdena skiljer sig åt. t, t-faktorn; df, frihetsgrader; P, sannolikhet.

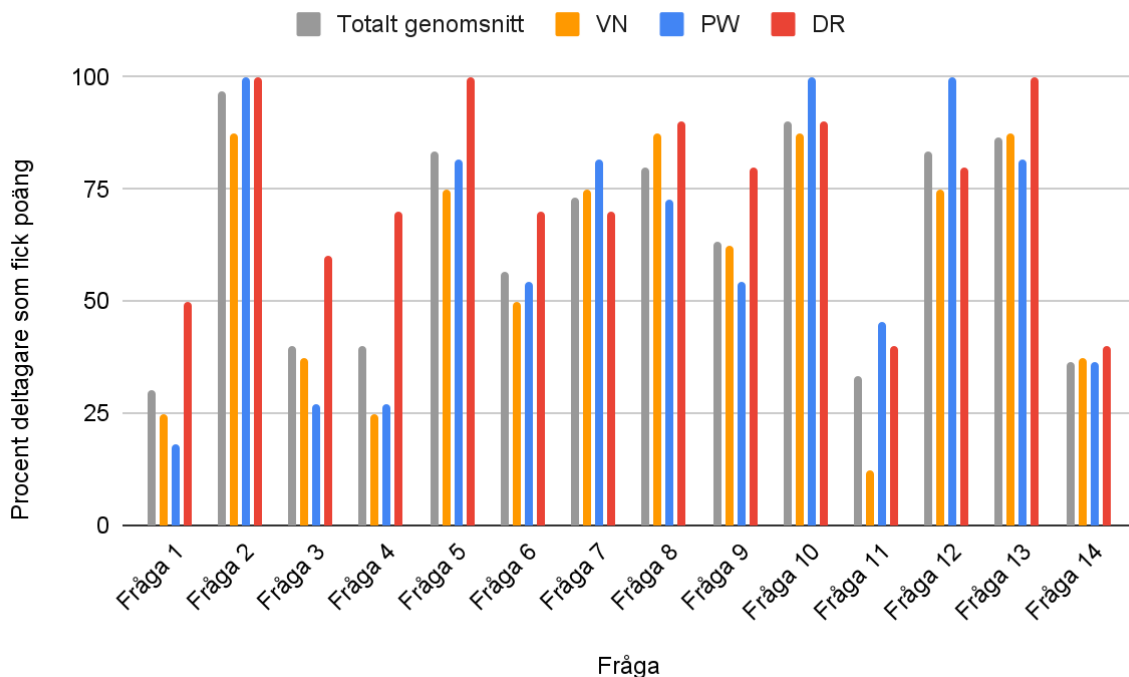
	VN - PW			VN - DR			PW - DR		
	t	df	P	t	df	P	t	df	P
Totalpoäng	0.37	17	0.71	1.36	16	0.19	1.39	19	0.18
Sammanfattning	0.06	17	0.95	0.48	16	0.64	0.50	19	0.62
Frågor	0.47	17	0.64	1.64	16	0.12	1.43	19	0.17
Tid	2.1	17	0.06	2.7	16	0.02*	0.40	19	0.69
Nöje	2.8	17	0.01*	2.5	16	0.77	0.30	19	0.02*

Vidare undersöktes om det fanns skillnader mellan grupperna vad gäller sammanfattningsdelen respektive enpoängsfrågedelen i läsförståelsetestet (Figur 8). I korthet så kunde inte signifikanta skillnader mellan spelversionerna påvisas när sammanfattningsdelen och enpoängsfrågedelen analyserades var för sig (Tabell 2). Intressant att notera är dock att medelvärdet på enpoängsfrågorna för version DR (10.4 ± 2.7) närmar sig en signifikant skillnad både gentemot version VN (8.3 ± 2.9) och PW (8.8 ± 2.4), med $P=0.12$ respektive $P=0.17$.



Figur 8 Poäng på sammanfattningsdelen (vänster) respektive enpoängsfrågorna (höger) på läsförståelsetestet uppdelat per spelversion. Varje punkt representerar data från en enskild försöksperson. Horisontella streck representerar medelvärdet för respektive grupp.

Figur 9 visar andel försökspersoner som svarat rätt på varje enskild enpoängsfråga på läsförståelsetestet. I genomsnitt svarade 64% av försökspersonerna rätt på en enskild fråga. Frågorna uppvisade en stor spännvidd i svårighetsgrad. Till exempel gav 30-40% av försökspersonerna korrekta svar på frågorna 1, 3, 11 och 14 medan fler än 90% gav korrekta svar på frågorna 2 och 10.

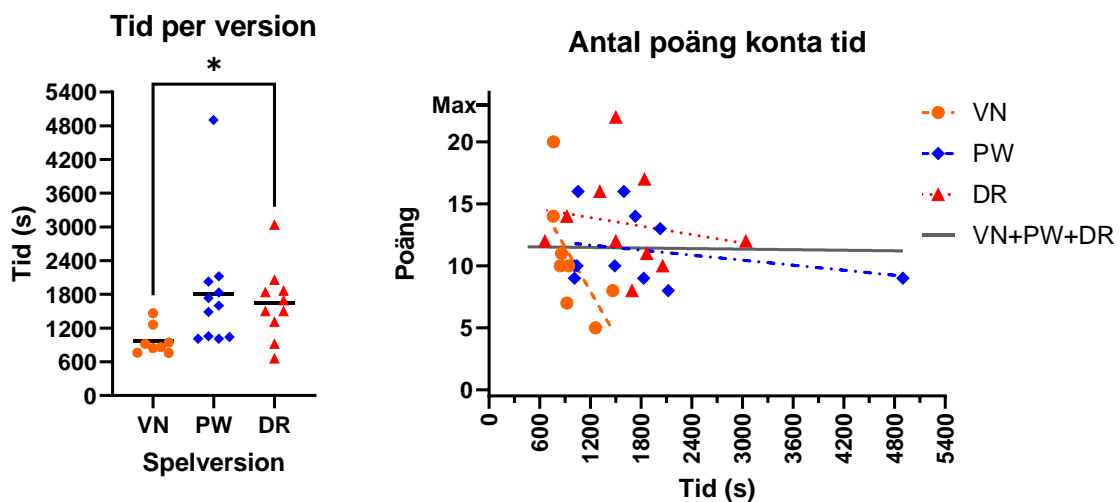


Figur 9 Översikt av andel (%) poänggivande svar på läsförståelsetestet för alla enpoängsfrågor, uppdelat per spelversion och sammantaget.

Ytterligare analyser gjordes för att undersöka samband mellan resultat på läsförståelsetestet och tidsåtgång, hur “bra spelet gick” (antal felval i debatter), nöje av spelet, samt erfarenhet av visual novel-spel i allmänhet och kännedom om Among Us.

VN-versionen tog i genomsnitt drygt 15 minuter (977 ± 253 sekunder) att spela/läsa igenom medan PW och DR versionerna tog cirka 30 minuter (1802 ± 1108 respektive 1639 ± 654 sekunder) att spela (Figur 10 vänster; Tabell 1). Skillnaden i speltid mellan VN och DR var signifikant, $P=0.02$ (Tabell 2). Skillnaden i speltid mellan VN och PW var inte signifikant, men notera den stora variationen i tidsåtgång för olika deltagare för version PW. Värt att notera är också att det var tekniskt möjligt att spela versionerna med problemlösningsmoment (PW och DR) snabbt. Den försöksperson som kom igenom mordgåtan på allra kortast tid spelade version DR.

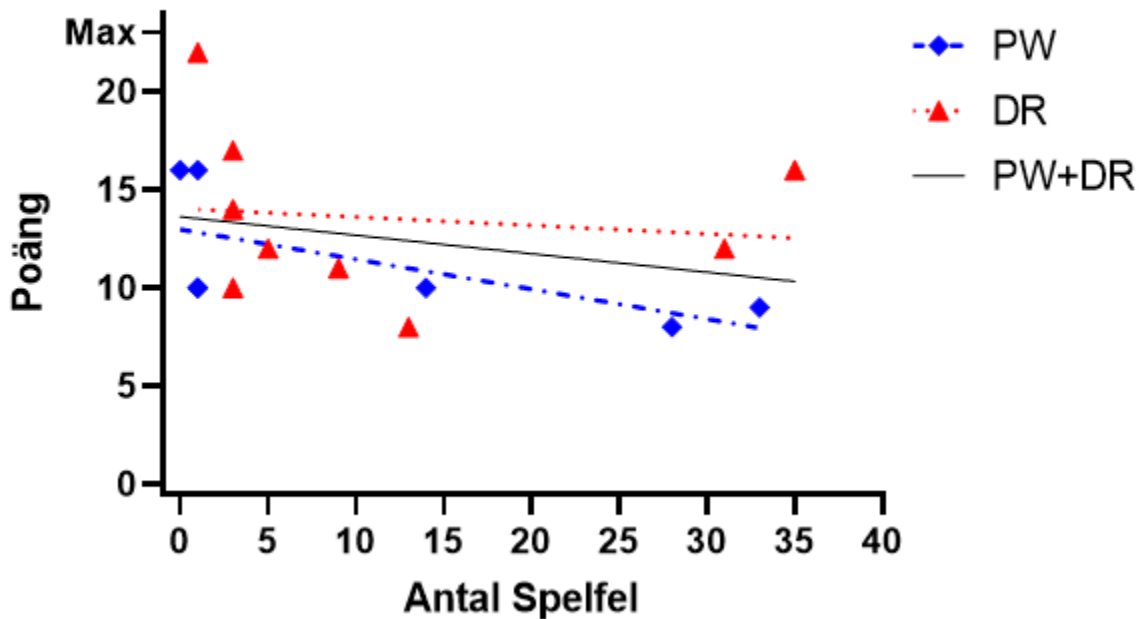
Figur 10 höger illustrerar antal poäng på läsförståelsetestet kontra tid. Linjär regression enligt minsta kvadratmetoden gjordes för de olika grupperna. Genom F-test med Prism (GraphPad) undersöktes sedan om variansen hos datapunkterna var tillräckligt liten för att stödja att linjernas lutningar signifikant skiljde sig från 0. Nollhypotesen är i Figur 10 höger att lutningen är 0 (samma gäller också i övriga exempel i denna studie som analyserats med linjär regression). Om en skillnad från lutningen 0 inte kan visas kan inte nollhypotesen förkastas och heller inget samband mellan variablerna påvisas (Somekh & Lewin 2005, s.266-235). Resultaten från detta och efterföljande F-test finns samlade i Appendix D. Inte för någon av spelversionerna (eller totalt) kunde ett signifikant samband mellan antal poäng på läsförståelsetestet och speltid påvisas (Figur 10 höger, Appendix D).



Figur 10 Förhållande mellan spenderad speltid och poäng på efterföljande läsförståelsetest. Speltid uppdelat per spelversion (vänster). Statistiskt signifikant skillnad ($P < 0.05$) enligt t-test mellan markerade grupper är indikerad med *. Antal poäng kontra tid (höger). Linjer representerar linjära regressionspassningar för data från respektive grupp eller alla data sammantaget (VN+PW+DR).

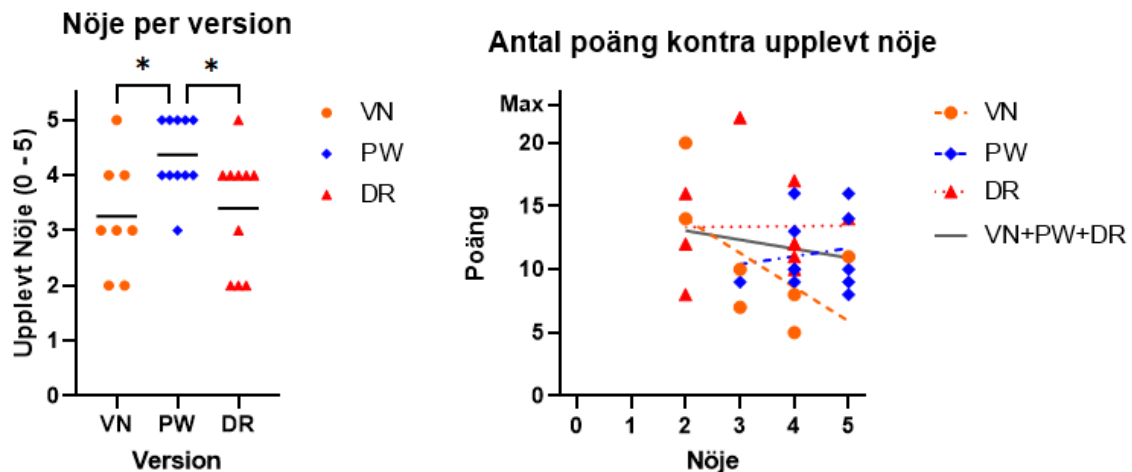
Vidare undersöktes om något samband kunde påvisas mellan antalet poäng på läsförståelsetestet och antal felval som spelare gjort under problemlösningsdelarna i PW- och DR-versionerna av spelet (Figur 11). Detta var inte relevant för version VN eftersom det inte fanns något moment av problemlösning i denna version och spelare alltså inte kunde svara fel. Varken för version PW eller DR (eller båda sammantaget) kunde ett signifikant samband mellan antal poäng på läsförståelsetestet och antal fel i spelen påvisas (Appendix D).

Antal poäng kontra antal felval i spelen



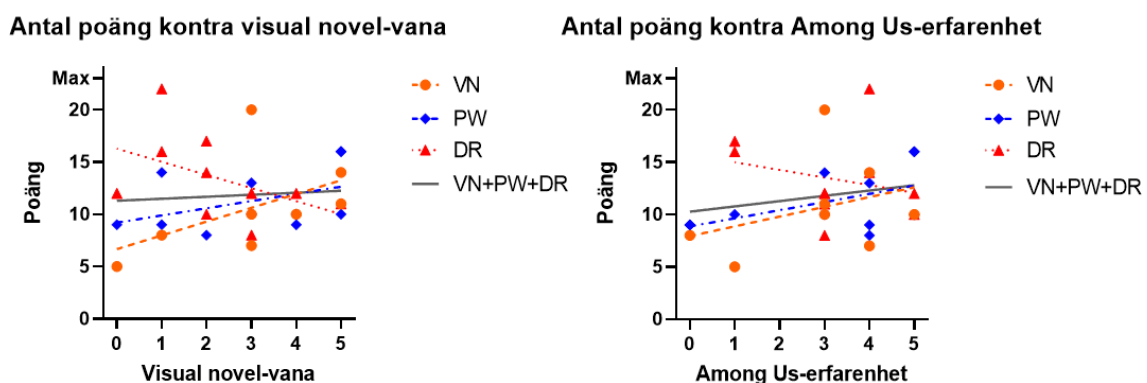
Figur 11 Förhållande mellan antal fel valda bevisobjekt (spelfel) och antal uppnådda poäng på efterföljande läsförståelsetest. Varje punkt representerar data från en enskild försöksperson. Linjer representerar linjära regressionspassningar för data från respektive grupp eller data sammantaget (PW+DR) så som indikeras i figuren.

Upplevt nöje av respektive spelversion visas i Figur 12 vänster. På en skala från 0 till 5 där 5 är bäst uttryckte försökspersonerna mest nöje (4.4 ± 0.7) med PW-versionen, vilket var signifikant högre än för versionerna VN (3.3 ± 1.0) och DR (3.4 ± 1.1), $P=0.01$ respektive $P=0.02$ (Tabell 1, Tabell 2). Figur 12 höger visar antal poäng på läsförståelsetestet kontra upplevt nöje av de tre spelversionerna. Inte för någon av spelversionerna kunde ett signifikant samband mellan antal poäng på läsförståelsetestet och upplevt nöje påvisas (Appendix D).



Figur 12 Förhållande mellan upplevt nöje och poäng på efterföljande läsförståelsetest. Upplevt nöje uppdelat per spelversion (vänster). Statistiskt signifikanta skillnader ($P < 0.05$) enligt t test är indikerade med *. Antal poäng kontra upplevt nöje (höger). Linjer representerar linjära regressionspassningar för data från respektive grupp eller alla data sammantaget (VN+PW+DR).

Slutligen undersöktes om något samband kunde visas mellan antalet poäng på läsförståelsetestet och tidigare vana av visual novel-spel (Figur 13 vänster) eller kännedom om Among Us (Figur 13 höger). Medelvärden och standardavvikelse för rapporterad visual novel-erfarenhet var för respektive version: VN 3.0 ± 1.7 , PW 3.0 ± 1.8 , DR 2.3 ± 1.5 . Motsvarande resultat gällande kännedom om Among Us var: VN 2.9 ± 1.6 , PW 3.1 ± 1.9 , DR 3.2 ± 1.4 . Inte för någon av spelversionerna kunde ett signifikant samband mellan antal poäng på läsförståelsetestet och erfarenhet av visual novel-spel eller Among Us påvisas (Appendix D).



Figur 13 Antal poäng kontra tidigare erfarenhet av visual novel-spel (vänster) respektive erfarenhet av Among Us (höger). Varje punkt representerar data från en enskild försöksperson. Linjer representerar linjära regressionspassningar för data från respektive grupp eller alla data sammantaget (VN+PW+DR) så som indikeras i figurerna.

5.3 Retrospektiv poweranalys

I avsnitt 5.2 har statistiska test utförts för att undersöka om nollhypoteser kan förkastas. Signifikansnivån (α) sattes till 0.05, vilket innebär en risk på 5% för ett falskt positivt svar, typ I fel, det vill säga att en nollhypotes felaktigt förkastas. Den motsatta situationen är ett falskt negativt svar, ett typ II-fel (sannolikheten för detta kallas β), vilket betyder att felaktigt dra slutsatsen att två grupper inte skiljer sig åt fast de i verkligheten gör det (Aczel & Sounderpandian 2008, s.260-261). Det är möjligt att minska risken för typ II-fel genom att försäkra sig om att studien i fråga har tillräcklig power (Aczel & Sounderpandian 2008, s.264). Power definieras $1-\beta$. Studier eftersträvar ofta en power större än 0.8, vilket betyder 80% chans att hitta en skillnad som finns i verkligheten.

Innan genomförandet av denna studie var det svårt att bedöma hur testresultaten skulle se ut, så som vilken totalpoäng de olika försöksgrupperna skulle få på läsförståelsetestet. Den statistiska mjukvaran G*Power (Faul et al. 2007) användes för att retrospektivt estimerar hur många försökspersoner som skulle behövas per försöksgrupp för att nå en viss power (Tabell 3) och diskuteras vidare i sektion 6.2.1. Denna poweranalys gjordes i tre grupper (A, B, C) så som indikerats i tabellen, med distinkta syften. Respektive syfte och resultat var:

- A. Syfte: Analys av gruppstorlek som krävs för att uppnå Power 0.8 i ett scenario som efterliknar data för totalpoängen på läsförståelsetestet i denna studie.

Som inputvärden användes effektstorlekar i samma storleksordning som uppmätts för totalpoängen på läsförståelsetestet. Power sattes till 0.8.

Resultat: Givet de effektstorlekar som observerats i studien visar resultatet från poweranalysen att antalet personer som behövts för att åstadkomma en Power = 0.8 är betydligt större än det antal som rekryterats till studien.

- B. Syfte: Estimering (genom iteration) av Power i ett scenario som efterliknar data i denna studie.

Som inputvärden användes effektstorlekar i samma storleksordning som uppmätts för totalpoängen på läsförståelsetestet. Power itererades tills output blev 10 personer per försöksgrupp.

Resultat: Power estimerades till värden betydligt lägre än 0.8, vilket indikerar hög sannolikhet för typ II-fel.

- C. Syfte: Analys av gruppstorlek som krävs för att uppnå Power 0.8 i ett scenario med samma skillnad mellan medelvärden som i denna studie. Standardavvikelsen sattes arbiträrt till ett lågt värde för att simulera att testet hade utförts på en mer homogen population försökspersoner. Detta scenario kan till exempel jämföras med att samtliga försökspersoner hade haft en mycket likartad kunskapsnivå i engelska.

Som inputvärden användes skillnader i medelvärden tagna från denna studie. Standardavvikelsen sattes till 1 och power till 0.8.

Resultat: Den gruppstorlek som krävts estimerades till 34 per grupp för skillnader mellan medelvärden som motsvarar vad som uppmättes mellan versionerna VN och PW. Motsvarande gruppstorlekar mellan versionerna VN och DR estimerades till 4. Så få deltagare som 4 försökspersoner per grupp låter orimligt i en verklig studie, men exemplet illustrerar den stora betydelse som storleken på standardavvikelsen har för att identifiera signifikanta skillnader mellan jämförda grupper.

Tabell 3 Poweranalys. Olika värden på M1-M2, SD och Power användes som input för beräkning av antalet försökspersoner per grupp som krävs för att uppnå indikerad Power.

Definitioner: M1-M2, Medelvärde för grupp 1 minus medelvärde för grupp 2; SD, standardavvikelse; Effektstorlek (Effect size) beräknas som $(M1-M2)/SD$.

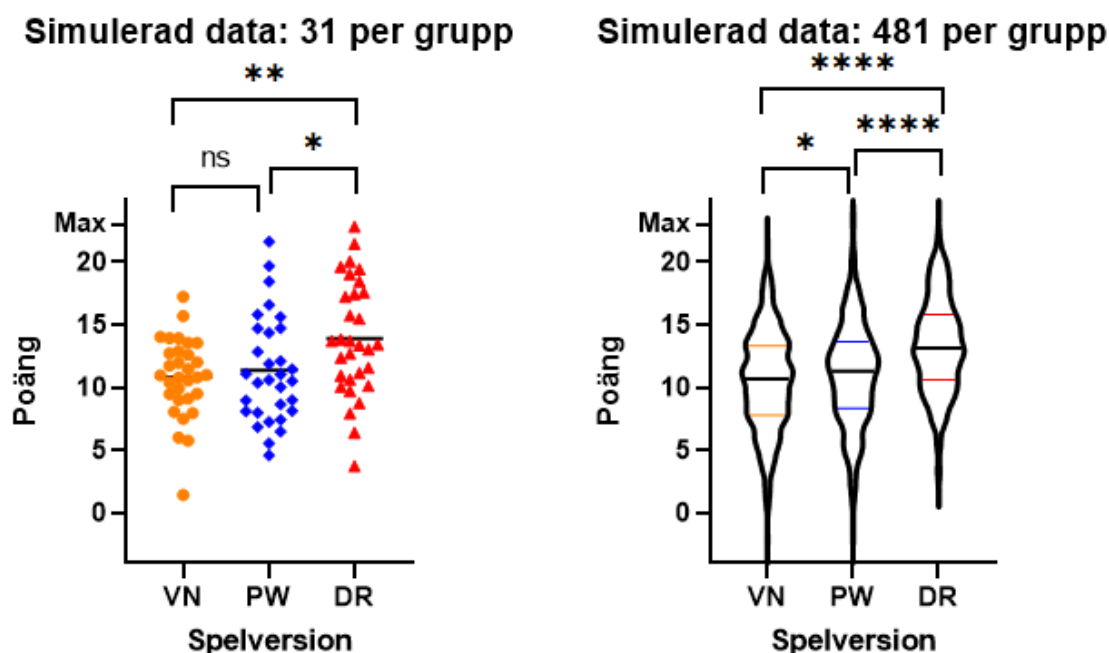
^a uppmätt skillnad i totalpoäng på läsförståelsetestet mellan version PW och VN.

^b uppmätt skillnad i totalpoäng på läsförståelsetestet mellan version DR och VN.

^c genomsnittlig SD för totalpoängen för de tre spelversionerna.

<i>Syfte</i>	<i>M1 - M2</i>	<i>SD</i>	<i>Effektstorlek</i>	<i>Power</i>	<i>Personer per grupp</i>
A	0.7 ^a	3.87 ^c	0.18	0.80	481
	2.8 ^b	3.87 ^c	0.72	0.80	31
B	0.7 ^a	3.87 ^c	0.18	0.065	10
	2.8 ^b	3.87 ^c	0.72	0.32	10
C	0.7 ^a	1	0.70	0.80	34
	2.8 ^b	1	2.80	0.80	4

Utifrån det antal försökspersoner som bedömdes krävas för att uppnå power 0.8 gjordes simuleringarna i Figur 14. Grupper på 31 simulerade datapunkter per grupp användes för att efterlikna en studie med power 0.8 för jämförelsen mellan DR och VN (Figur 14 vänster). 481 simulerade datapunkter användes för motsvarande jämförelse mellan PW och VN (Figur 14 höger).



Figur 14 Simulerade data för totalpoäng baserad på gruppstorlekar med 31 (vänster) respektive 481 (höger) datapunkter per grupp. Horisontella svarta streck representerar medelvärdet för respektive grupp. I violinplottarna (höger) representerar färgade streck 25-percentiler. Signifikansnivåer (enligt student t-test): * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, ns icke signifikant.

Datapunkter simulerades (Random scatter, Gauss-distribuerad) med Prism utifrån medelvärdena 10.6 (VN), 11.3 (PW) och 13.4 (DR). Standardavvikelse 3.87 användes i samtliga fall. Skillnader mellan olika grupper analyserades parvis med student t-test. I det simulerade exemplet med 31 datapunkter per grupp uppnåddes tvåstjärnig signifikans mellan grupperna DR och VN ($P = 0.0031$), enstjärnig mellan DR och PW ($P = 0.027$). Ingen signifikant skillnad uppnåddes mellan PW och VN ($P = 0.57$). Observera att detta endast är ett exempel och att resultaten från separata simuleringar inte blir identiska. I exemplet med 481 simulerade datapunkter per grupp uppnåddes fyrstjärnig signifikans i jämförelserna mellan DR och VN ($P < 0.0001$) samt mellan DR och PW ($P < 0.0001$). Jämförelsen mellan PW och VN ($P = 0.047$) uppnådde enstjärnig signifikans.

5.4 Slutsatser

Hypotesen att de som aktivt engagerat sig i en text via logikpussel (versionerna PW och DR) kommer att prestera bättre på efterföljande läsförståelsetest än de som läst texten i ett mera bokliknande format kunde inte bekräftas. En tendens till att de som spelat version DR presterade bättre på läsförståelsetestet anas men var inte statistiskt signifikant.

Retrospektiv analys visade att studien var “underpowered”. Det finns en stor risk för typ II-fel, det vill säga att faktiska skillnader mellan olika försöksgrupper inte påvisats i studien på grund av för få försökspersoner i förhållande till standardavvikelser och skillnader i medelvärden mellan jämförda grupper.

Trots stor risk för typ II-fel kunde några samband påvisas:

- Det tog längre tid att spela version DR, som innehöll problemlösningselement, än läsversionen VN som saknade dem (och en trend mot att PW också tog längre tid även om det inte kunde påvisas statistiskt).
- Upplevt nöje var högre för PW än för de andra versionerna.

Ytterligare potentiella samband undersöktes, men inget av dessa kunde påvisas statistiskt:

Antal poäng på läsförståelsetestet kontra tidsåtgång, antal spelfel, upplevt nöje, tidigare erfarenhet av visual novel-spel och kännedom om Among Us.

6 Avslutande diskussion

6.1 Sammanfattning

Denna studie syftade till att kvantitativt jämföra hur spelmekaniska skillnader påverkar spelares förmåga att ta till sig och förstå innehållet i en text. Hypotesen som formulerades var att de som aktivt engagerat sig i en text via logikpussel kommer att prestera bättre på ett efterföljande läsförståelsetest med textens innehåll som ämne än de som fått texten presenterad i ett mer bokliknande format.

Bakgrunden till denna hypotes är att inlevelse och problemlösning länge använts som drivkraft för lärande (Johnson & Swain 1997). Förhoppningen är att dataspel anpassade för undervisning ska kunna höja studieengagemang och i förlängningen förbättra studieresultat, vilket också flera tidigare studier indikerar (Rosas et al. 2003; Faizal 2016; Gros 2007; Jere-Folotiya 2014; Liu & Chu 2010).

För att studera spelmekanikers påverkan på försökspersoner att ta till sig en text skapades tre versioner (VN, PW och DR) av ett visual novel-spel som alla användes till att presentera en och samma text, men på olika sätt. Texten som presenterades var en mordgåta adapterad från det sociala spelet Among Us (2018). Version VN liknar läsning i en bok där spelaren klickar sig framåt i berättelsen utan att kunna kontrollera mycket annat. Version PW (2001) skiljer sig från version VN i särskilda pusselsektioner. I dessa sektioner kan spelaren fritt bläddra fram och tillbaka bland dialogreplikor, men är tvungna att klicka på motbevisbara påståenden med rätt bevisobjekt (valt ur en lista) för att komma vidare i spelet. Version DR (2014) liknar version PW, förutom att det är (motoriskt) svårare att klicka på påståendena under pusselsektionerna, då texten rör sig över skärmen och dialogen spelar upp sig själv efter bestämda tidsintervaller.

För att testa om versionerna hade olika effekt på förståelsen av texten fick försökspersoner göra ett läsförståelsetest med innehållet i texten som ämne. Läsförståelsetestet var inspirerat av och efterliknade ett QRI-test (Leslie & Caldwell 2021). QRI-test används vanligtvis för att bedöma skolelevers läsförståelse för att kunna anpassa vilken svårighetsgrad på texter eleven i fråga bör få i framtida undervisning och om extra stöd eventuellt behövs. Testet bestod av en sammanfattning där försökspersonerna med egna ord skulle beskriva händelseförloppet i berättelsen (kunde ge max 9 poäng), och en del med specifika enpoängsfrågor (max 14 poäng).

Försökspersonerna utgjordes av ett bekvämlighetsurval och söktes främst bland studenter inom Spelutvecklingsprogrammet på Högskolan i Skövde och bland deras vänner. I storleksordningen 90 försökspersoner (30 per grupp) eftersträvades (Faizal 2016; Somekh & Levin 2005, s.218). I slutändan rekryterades totalt 29 deltagare (8, 11 respektive 10 slumpade till grupperna VN, PW och DR). Maxpoängen på läsförståelsetestet var 23. Medelvärde och standardavvikelse för de olika grupperna var: VN 10.6 ± 4.7 , PW 11.3 ± 2.9 , DR 13.4 ± 4.0 . Skillnaderna i medelvärde var inte statistiskt signifikanta (tvåsidigt student t-test) vid parvis jämförelse mellan grupperna. Hypotesen att de som aktivt engagerat sig i en text via logikpussel (versionerna PW och DR) kommer att prestera bättre på efterföljande läsförståelsetest än de som läst texten i ett mera bokliknande format (version VN) kunde därmed inte bekräftas. En tendens till att de som spelat version DR presterade bättre på läsförståelsetestet anas men var inte statistiskt signifikant.

En retrospektiv analys av data gjordes och visade att studien var “underpowered”. Det finns alltså en stor risk för typ II-fel, och betyder att faktiska skillnader mellan olika försöksgrupper inte påvisats i studien på grund av för litet antal försökspersoner i förhållande till standardavvikelser och skillnader i medelvärden mellan jämförda grupper. Mer omfattande studier med ett större antal försökspersoner behövs om hypotesen ska kunna bekräftas eller förkastas.

Trots detta kunde några samband påvisas:

- Version DR, som innehöll problemlösningselement, tog längre tid att spela än version VN som saknade dem (och en trend mot att PW också tog längre tid, men det kunde inte påvisas statistiskt).
- Upplevt nöje var högre för version PW än för de andra versionerna.

Ytterligare potentiella samband så som antal poäng på läsförståelsetestet kontra tidsåtgång, antal spelfel, upplevt nöje, tidigare erfarenhet av visual novel-spel och kännedom om Among Us undersöktes. Inte i något av dessa fall kunde ett samband påvisas statistiskt.

6.2 Diskussion

6.2.1 Trovärdighet och koppling till antal försökspersoner

Denna studie kunde inte fånga upp några signifikanta skillnader i prestation på efterföljande läsförståelsetest som kan härledas till att försökspersoner spelat de olika versionerna av visual novel-spelet. Möjligen hade en signifikant skillnad kunnat uppnås om ett större antal försökspersoner hade deltagit. Stor möda lades ner på att försöka rekrytera försökspersoner. Ambitionen då studien designades var att engagera ungefär lika många försökspersoner som Faizal (2016) vars studie hade 27 försökspersoner per grupp. Den retrospektiva poweranalysen, vars slutsatser dock inte var klara innan datainsamlingen avslutats, visar tydligt att studien var “underpowered”. Den indikerar att cirka 30 personer per grupp skulle behövas för att fånga upp en verklig skillnad mellan versionerna VN och DR givet den effektstorlek som observerades i studien.

Ju mindre skillnad mellan medelvärden man vill kunna mäta mellan grupper, desto fler försökspersoner krävs givet att standardavvikelsen är densamma (Faul et al. 2007). En studies storlek behöver balanseras mellan tidsåtgång, kostnad och hur stor/liten effekt som är relevant att mäta. Det är svårt att avgöra hur stor/liten prestationsförbättrande effekt, i detta fall resultat på läsförståelsetestet, som är meningsfull att kunna mäta. Jag spekulerar att cirka 3 poängs skillnad (vilket motsvarar den som observerats mellan VN- och DR-grupperna) är en rimlig nivå. Detta stödjer också att cirka 30 försökspersoner per grupp hade varit ett lämpligt minsta antal.

Meddelanden för att få tag på försökspersoner riktades framför allt till studenter på Spelutvecklingsprogrammet vid Högskolan i Skövde och lades upp i flera olika grupper på sociala medier. Rekryteringen av försökspersoner gick trögt, även efter upprepade förfrågningar i dessa grupper. Därför utvidgades sökandet till att inkludera mina vidare sociala kretsar. Det infördes även utlottning av trisslotter som incitament för personer att vara med som försökspersoner. Efter utvidgandet till bredare sociala kretsar och införandet av belöning ökade antalet försökspersoner med cirka 50%. För att möjliggöra dataanalys inom tidsramen för detta examensarbete stoppades datainsamlingen efter att totalt 29 personer

deltagit i studien. Potentiellt hade antalet försökspersoner kunnat utökas till exempel genom att kontakta andra högskolor eller gymnasieskolor med utbildningar inom spelutveckling. Detta var dock i praktiken svårt då datainsamlingen skedde under den tidsperiod då skolorna var stängda för sommarlov. Möjligen var det ett överambitiöst mål att inom ramen för ett examensarbete rekrytera så många som 90 försökspersoner.

Varje försöksperson slumpades att spela en av de tre olika spelversionerna baserat på födelsedatum. Personer födda 1-10 i varje månad fick spela version VN, 11-20 version PW och 21-31 version DR. Detta system för att dela in försökspersoner ledde till likartat men inte identiskt antal försökspersoner i varje grupp. I undersökningar som denna, med relativt få deltagare, kan slumpen göra så att den procentuella skillnaden mellan grupperna blir rätt stor. Om ett större antal försökspersoner hade deltagit i studien förväntar jag mig att antalet personer per grupp hade jämnat ut sig. Så här i efterhand är det tydligt att ett slumpsystem som genererat samma antal försökspersoner per grupp hade varit att föredra. Detta hade kunnat åstadkommas till exempel genom ett system som alltid placerat nästkommande försöksperson i en/den grupp som för tillfället hade minst antal deltagare.

6.2.2 Val av försökspersoner och språk

Initialt söktes en grupp av försökspersoner med god spelvana och kännedom om spelet Among Us. En anledning var att minimera tröskeln för personer att vilja delta som försökspersoner i studien. En annan anledning var strävan att ha liten variation inom respektive grupp vad gäller spelvana och kännedom om Among Us för att minimera dessa variablers påverkan på spridningarna bland resultaten på läsförståelsetestet, med syftet att lättare identifiera skillnader som beror på vilken spelversion som använts. Detta kan jämföras med experimentella kvantitativa försöksupplägg i kontrollerade system (till exempel inom kemi eller fysik) där man strävar efter att hålla alla parametrar så lika som möjligt mellan de grupper som ska jämföras, förutom den parameter som ska testas (Somekh & Lewin 2005, s. 215-225).

En invändning mot att söka en population med god spelvana och kännedom om Among Us är att den inte nödvändigtvis är representativ för befolkningen i stort. Som tidigare kommenterats i metodavsnittet (sektion 3.1.2) så är inte strävan i denna studie att undersöka läsförståelsen hos befolkningen som helhet utan att jämföra olika spelversioners inverkan på hur försökspersoner kan ta till sig en text. Även om försökspersonerna inte nödvändigtvis representerar befolkningen i stort så är det rimligt att anta att den genomsnittliga läsförståelsen är likartad i de tre olika försöksgrupperna eftersom försökspersonerna slumpmässigt fördelats mellan grupperna. Ett alternativt upplägg som inte rymdes inom ramen för detta projekt hade varit att försökspersonerna innan indelningen i grupper gjort ett normalt QRI-test vars testresultat använts för att göra grupper med samma genomsnittliga läsförståelse.

En faktor som kan ha försvårat möjligheten att förkasta nollhypotesen är att spelen och läsförståelsetestet gjordes på engelska. Engelska användes eftersom den berättelse som adapterats till manuset ursprungligen är på engelska (Nutrin 2020). De flesta försökspersonerna har svenska som modersmål och engelska som andraspråk. Jag spekulerar att variationen i kunskapsnivå mellan individer inom varje försöksgrupp skulle vara mindre om modersmålet används istället, vilket leder vidare till hypotesen att om spelen och läsförståelsetestet varit på svenska så hade variationen i testresultat inom respektive försöksgrupp varit mindre. Två olika sätt att hantera detta hade varit att antingen komplettera

frågeformuläret med en bedömning av kunskapsnivån i engelska (till exempel med ett gymnasiebetyg i engelska) eller att skriva spelet och läsförståelsetestet på svenska.

6.2.3 Tekniska förbättringsmöjligheter

Ett antal svårigheter/begränsningar orsakade av designmässiga eller tekniska brister kom fram under datainsamlingen. Dessa var:

- Vissa spelare misstolkade instruktionerna och trodde att det behövdes ett motbevis per markerat påstående i varje debatt, när det egentligen endast behövdes ett motbevis totalt per debatt. Detta resulterade i att ett antal spelare fastnade längre tid än nödvändigt i debatt nummer två där detta för första gången var relevant (gäller PW- och DR-versionerna). Tydligare instruktioner bör implementeras i framtida versioner.
- Instruktionerna verkar inte heller varit helt tydliga angående vad som förväntades i sammanfattningsdelen av läsförståelsetestet då svaren var av vitt skilda karaktärer. En beskrivning av manusets mordfall önskades. Vissa försökspersoner valde att fokusera sammanfattningarna mer på dialogen / spelflödet. Det senare fallet är inte fel, men tenderade att ge färre poäng då information om mordgåtan utelämnades. I framtida versioner bör en tydligare instruktion skrivas att sammanfattningen ska fokusera på mordgåtan.
- Det hade inte implementerats något sätt att räkna antalet personer som börjat spela någon av spelversionerna men inte lämnat in några svar på läsförståelsetestet. Det är rimligt att anta att åtminstone någon eller några försökspersoner fallit ifrån under antingen spelet eller läsförståelsetestet. Det saknas information om någon av spelversionerna är associerad med ett större avhopp av försökspersoner. En funktion som registrerar att en spelversion startats bör implementeras i framtida versioner.
- Några deltagare rapporterade problem med att ladda ner och starta spelen på grund av antivirusprogram. I en del fall kunde detta lösas genom att försökspersoner spelade direkt på testledarens dator. Det är inte känt hur många som totalt försökt vara med i studien men misslyckats på grund av detta problem.

6.2.4 Forskningsetiska aspekter och genusaspekter

Studien har följt individskyddskravet som beskrivs i Vetenskapsrådets forskningsetiska principer (2002). I praktiken innebär det ett antal punkter:

- Det har varit frivilligt att delta i studien och att avbryta deltagandet när som helst ifall en försöksperson velat göra det. Försökspersonerna har informerats om syftet med studien innan de deltagit.
- Ingen information insamlad i denna studie har använts till eller kommer att användas för några andra syften än detta examensarbete.
- Inga personuppgifter har samlats in. Försökspersoner har dock frivilligt kunnat lämna en valfri typ av kontaktuppgift (med möjlighet att vara anonym) så som ett discord-användarnamn eller en mejladress för delta i utlottning av trisslotter. Utlottning av trisslotter infördes som ett sätt att försöka locka fler deltagare att vara med som försökspersoner. Detta är en vanlig metod som till exempel opinionsinstitut använder för att få deltagare till sina studier (Yougov 2000). Efter att datainsamlingen slutförts och trisslotter delats ut raderades alla kontaktuppgifter. Inga enskilda testresultat är kopplade till kontaktuppgifter.

Endast begränsad information samlades in om försökspersonernas profiler av respekt för individskyddskravet. Den information som trots allt samlades in var tidigare erfarenhet av Visual Novel-spel och kännedom om Among Us. Information om till exempel försökspersonernas kön, ålder och kunskapsnivå i engelska hade också varit relevanta att samla in för att kunna göra fördjupade analyser, och hade kunnat göras inom ramarna för individskyddskravet. Jag spekulerar att det statistiska värdet av sådan information hade varit begränsad i denna studie på grund av det låga antalet deltagare. I en större studie skulle värdet kunna bli större och samband eventuellt kunna påvisas och undergrupper identifieras.

Vad gäller genusaspekter så har inga data gällande försökspersoners genus eller kön samlats in. Denna studie har därför inte adresserat om genus/kön påverkar testresultat eller hur roande spelen uppfattades.

Könsaspekter existerar bland karaktärerna som är med i spelets manus. Karaktärerna har en mycket enkel grafisk utformning. Även om karaktärerna uppträder som människor så ser de inte ut så. Könsskillnader tydliggjordes genom stereotypa traditionella attribut. Detta eftersom det finns frågor i det efterföljande läsförståelsetestet där kön hos karaktärer skulle pekats ut. Exempelvis så heter två av de kvinnliga karaktärerna Lila och Rosa, färger typiskt associerade till kvinnlighet (Cunningham & Macrae 2011; Shakin, Shakin, & Sternglanz 1985), och de har klänning på sig respektive blomma i håret.

En reflektion är att huvudpersonen som leder utredningen av mordgåtan är man och mordoffret kvinna. Detta scenario är taget från en existerande Among Us-berättelse (Nutrin 2020) och inte modifierat. Det kunde mycket väl ha gjorts tvärtom eller utan att kommentera kön.

6.2.5 Samhällelig nytta

Som beskrivet i inledningen är motivation och engagemang viktiga för inläring (Johnson & Swain 1997). Eftersom dataspel är konstruerade för att skapa engagemang (Jennett et al. 2008) så är förhoppningen att dataspel användas i undervisning ökar studieengagemang och inläring jämfört med traditionell undervisning, vilket flera tidigare studier också tyder på (Rosas et al. 2003; Faizal 2016; Gros 2007; Jere-Folotiya 2014; Liu & Chu 2010).

Denna studie syftade till att undersöka hur olika spelmekaniker kan påverka hur väl spelare tar till sig textinnehåll. Denna begränsade studie kunde inte statistiskt visa att en text presenterad i spel med problemlösningsmoment var bättre för inläringen jämfört med när texten visades på ett sätt som mera liknar läsning i en bok. Det finns dock ett stort värde att undersöka vidare, till exempel i liknande men större studier än denna, om och hur dataspel kan användas i skolundervisning och privat för att förbättra möjligheten till inläring.

6.3 Framtida arbete

6.3.1 Kortsiktiga möjligheter

Som poweranalysen visade så var antalet försökspersoner som deltog i denna studie för litet för att kunna dra säkra slutsatser. Flera försökspersoner gav också värdefull återkoppling på möjliga förbättringar av spelversionerna och associerat läsförståelsetest.

Baserat på detta kan denna studie betraktas som en pilotstudie för en potentiell framtida studie med betydligt fler försökspersoner. Betydelsen av antalet försökspersoner har

presenterats och diskuterats utförligt i avsnitt 5.2 och 6.2.1. Om en liknande framtida studie görs så anser jag att minst 30 försökspersoner per grupp bör rekryteras. Några mindre justeringar av spelen och läsförståelsetestet bör också implementeras. Några exempel är att vissa spelinstruktioner behöver förtydligas och i läsförståelsetestet behövs ett förtydligande av vad som förväntas ingå i sammanfattningsdelen. Ett alternativ är att sammanfattningsdelen ersätts med ett större antal enpoängs-/flervalssfrågor eftersom försökspersoner hade olika tolkningar om vad som borde ingå i sammanfattningsdelen och därmed gjorde den mer svårriktad. Testet hade då liknat ett QRI-test mindre än nuvarande version.

Spelen i denna studie gjordes inom begränsade tidsramar och av en enskild person med kompetens framför allt inom programmering. En spelupplevelse av högre kvalitet kunde ha uppnåtts om personer med specialistkompetenser inom estetiska områden deltagit i projektet. En spekulering är att en mer komplett spelupplevelse med större inlevelse hade gett större effekt på inläring och därigenom högre testresultat generellt. Med en lite större arbetsinsats och på lite längre sikt än att bara repetera studien beskriven i denna rapport, skulle det vara intressant att se en liknande studie gjord på ett mera genomarbetat spel.

En annan möjlighet för en framtida studie på medellång sikt är att försöka konvertera existerande QRI-texter och test (förutom den muntliga delen) till digitalt format. Det gjordes inte i denna studie eftersom typiska QRI-texter inte är helt enkla att anpassa till spelmekanikerna (problemlösning genom att motbevisa påståenden) som utvecklades i denna studie.

6.3.2 Långsiktiga möjligheter

I dagsläget är vanligt förekommande läsförståelsetest som GORT, PIAT, WJPC och QRI (kort summerade av Keenan & Meenan 2014) pappersbaserade/analog. Vad gäller läsförståelsetexter och test så finns det möjlighet att digitalisera sådana på ett sätt som liknar spelen i denna studie, vilket underlättar distribution och rättning jämfört med traditionella pappersbaserade test. Vad som dock är svårt att härma med digitala test är den muntliga dialogen mellan testledare (lärare) och försöksperson (elev), vilken är en viktig del speciellt för yngre barn.

Den typ av test som är lättast att digitalisera innefattar ämnen med konkreta lätt definierbara korrekta svar som inte kräver långa eller subjektiva resonemang. Exempel på sådana möjligheter är geografi, glosor och matematik. Webb sidan elevspel.se (2014) är exempel på en gratis resurs med spelbaserade övningar som även kan fungera som enkla test i olika skolämnen. MyStudyWeb (2012) är en digital resurs som har tagit ett steg längre. Vad gäller läsförståelse så har de utvecklat digitala självriktade prov i svenska och engelska läsförståelse där elevernas resultat jämförs mot den nationella normen för respektive årskurs. (MyStudyWeb 2012). Detta är en början, men det finns stor potential för framtida utveckling av ytterligare sofistikerade digitala läsförståelsetest och test anpassade för andra skolämnen.

Med utgångspunkten att engagerande och inlevelseskapande spel har en positiv effekt på lärande så finns i det närmaste oändliga möjligheter att använda spel i undervisningssyfte. Detta används redan på mycket olika nivåer och många exempel kan ges. Exempel på ett företag i närområdet (Göteborg) som arbetar med detta är eEducation Albert AB (2015) som utvecklar spel för lärande av matematik, svenska, geografi och programmering riktade till barn mellan 3 och 16 år. Ett annat exempel från Göteborg är Edvirt AB (2013) som utvecklar simulatorer där gruvarbetare på ett säkert sätt kan öva sig i att spruta betong i gruvgångar för

att säkra tunnlarna. I framtiden kommer allt fler och mer sofistikerade digitala spel, simulatorer och test att utvecklas för lärande och utvärdering av kunskap.

Referenser

Aczel, A., Sounderpandian, J. (2008). *Complete Business Statistics, Seventh Edition*. Boston, MA: McGraw-Hill Higher Education, s. 740-767.

Among Us (2018). InnerSloth [Dataspel].

Bateman, C. (Ed.) (2009). *Beyond game design: nine steps towards creating better videogames*. Boston, MA: Charles River Media/Cengage Technology, s. 21.

Buser, T., & Peter, N. (2012). Multitasking. *Exp Econ*, 15, s. 641–655. doi: 10.1007/s10683-012-9318-8

Cavallaro, D. (2010). *Anime and the visual novel: narrative structure, design and play at the crossroads of animation and computer games*. Jefferson, NC: McFarland & Company, s. 8–9.

Cunningham, S. J., & Macrae, C. N. (2011). The Colour of Gender Stereotyping. *British Journal of Psychology*, 102 (3), s. 598–614. doi: 10.1111/j.2044-8295.2011.02023.x

Danganronpa: Trigger Happy Havoc (2014). NIS America [Dataspel].

Din, F. S., & Caleo, J., (2000). Playing Computer Games Versus Better Learning. *Proceedings of 70th Annual Conference of the Eastern Educational Research Association*. Clearwater, FL 16-19 Februari 2000.

Edvirt (2013) *About Edvirt*. [Webbsida] <https://sv.edvirt.com/about-edvirt> (Hämtad 2021-07-26).

eEducation Albert AB (2015) *Home* [Webbsida] <https://hejalbert.se/> (Hämtad 2021-07-26).

Ejvegård, R. (2009) *Vetenskaplig metod*. Lund, Sverige: Studentlitteratur AB, 4(8), s. 38-40.

Elevspel (2014) *Om Elevspel*. [Webbsida] <https://www.elevspel.se/om/> (Hämtad 2021-07-26).

Faizal, A (2016) The Effects of Conversation-Gambits Visual-Novel Game on Students' English Achievement and Motivation. I *Proceedings of the 2016 International Electronics Symposium (IES)*. Denpasar, Indonesia. 29-30 september 2016, s. 481-486.

Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 39(2), 175-191. doi: 10.3758/bf03193146.

Google Inc. (2006) *Google Forms* [Webbtjänst] Tillgänglig: <https://docs.google.com/forms> (hämtad 04/12/2021).

GraphPad (1994). *Prism* (Version 9.1.2) [Programvara].

Gros, B. (2007). Digital Games in Education, *Journal of Research on Technology in Education*, 40(1), s. 23-38. doi: 10.1080/15391523.2007.10782494

Guitar Hero (2005). RedOctane [Dataspel].

- Jennett, C., Cox, A., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66(9), s. 641-661. doi: 10.1016/j.ijhcs.2008.04.004
- Jere-Folotiya, J., Chansa-Kabali, T., Munachaka, J., C., Sampa, F., Yalukanda, C., Westerholm, J., & Lyytinen, H. (2014). The effect of using a mobile literacy game to improve literacy levels of grade one students in Zambian schools. *Educational Technology Research and Development*, 62(4), s. 417-436. doi: 10.1007/s11423-014-9342-9
- Juul, J. (2003). The Game, the Player, the World: Looking for a Heart of Gameness. I Copier, M. & Raessens, J. (eds) *Proceedings of Level Up: Digital Games Research Conference*. Utrecht Universiteit, Utrecht, Nederländerna. 4-6 november 2003, s. 30-45.
- Johnson, R. K., & Swain, M. (1997). *Immersion Education: International Perspectives*. Cambridge, UK: Cambridge University Press, s. 1-16.
- Keenan, J., Betjemann, R., & Olson, R. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3), s. 281-300. doi: 10.1080/10888430802132279
- Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities*, 47(2), s. 125-135. doi: 10.1177/0022219412439326
- Lantz, B. (2020). *Grundläggande statistisk analys*. Lund, Sverige: Studentlitteratur AB.
- Leslie, L. & Caldwell, J. (2021). *Qualitative Reading Inventory*. 7. uppl., London, UK: Pearson.
- Liu, T. Y., & Chu, Y. L. (2010). Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Computers & Education*, 55(2), s. 630-643. doi: 10.1016/j.compedu.2010.02.023.
- Lopez-Morteo, G., & López, G. (2007). Computer support for learning mathematics: A learning environment based on recreational learning objects. *Computers & Education*, 48(4), s. 618-641. doi: 10.1016/j.compedu.2005.04.014
- MyStudyWeb (2012) *Svenskaproven Omfattar Följande Områden* [Webbsida] <https://mystudyweb.com/corporate/kartlaggaren/svenska/> (Hämtad 2021-07-26).
- Myst* (1993). Brøderbund [Dataspel].
- Nutrin. (2020). *Among Us except it's Danganronpa and on the level of a PowerPoint presentation*. [Onlinevideolista]. Första videon uppladdad september 2020. Tillgänglig: https://www.youtube.com/playlist?list=PLwDpTS9B1Jz4qDnTjuaMDO-XX6ob_5Jy1 (Hämtad: 12/04/2021).
- Oakhill, J., Cain, K., & Elbro, C (2015). *Understanding and teaching reading comprehension: A handbook*. Abingdon-on-Thames, UK: Routledge.
- Phoenix Wright: Ace Attorney* (2001). Capcom [Dataspel].

Rosas, R., et al. (2003). Beyond Nintendo: design and assessment of educational video games for first and second grade students. *Computers & Education*, 40(1), s. 71-94. doi: 10.1016/S0360-1315(02)00099-4.

Schell, J., (2008) *The Art of Game Design: A Book of Lenses*. 1. uppl., Boca Raton, FL: CRC Press.

Shakin, M., Shakin, D., & Sternglanz, S. H. (1985). Infant clothing: Sex labeling for strangers. *Sex Roles: A Journal of Research*, 12 (9-10), s. 955–964. doi: 10.1007/BF00288097.

National Research Council (1998). Predictors of Success and Failure in Reading. I Snow, C. E., Burns, S. M., & Griffin, P (red.) *Preventing Reading Difficulties in Young Children*. Washington, DC: The National Academies Press. s. 100-134. doi: 10.17226/6023.

Somekh, B. & Lewin, C. (red.) (2005). *Research Methods in the Social Sciences*. Thousand Oaks, CA: SAGE Publishing. s.266-235.

Unity Technologies (2005). *Unity 3D* (Version 2019.3.7f1) [Programvara].

Vetenskapsrådet (2002). *Forskningsetiska principer inom humanistisk-samhällsvetenskaplig forskning*. Stockholm: Vetenskapsrådet. Tillgänglig: <https://www.vr.se/analys/rapporter/vara-rapporter/2002-01-08-forskningsetiska-principer-inom-humanistisk-samhällsvetenskaplig-forskning.html> (Hämtad 2021-05-08).

Yougov (2000) *Lotteri*. [Webbsida] <https://yougov.se/lotteri/> (Hämtad 2021-07-26).

Appendix A - Manus

Har adapterats från de första tre videor följande videolista (Nutrin 2020):
https://www.youtube.com/playlist?list=PLwDpTS9B1Jz4qDnTjuaMDO-XX6ob_5Jy1

EVIDENCE

A. **Map of the Ship**

I went and got this map from Admin on my way to the meeting. The added red lines show how the vents are connected between the rooms.

B. **Sabotage**

While we were working on our tasks, the power suddenly went out and all the lights went off. The ship's doors are all powered by electricity and cannot be forced closed if the power is out.

C. **My Account**

I had tasks to do in Storage. While I was working, Yellow came from the north and passed through Storage and went west. The blackout happened quite a while afterwards.

D. **Pink's Account**

Pink says she was in Lower Engine during the blackout. She noted hearing someone pass through but couldn't see who in the darkness.

Key:

(xxx) = Red's internal monologue / thinking

[xxx] = Evidence may be presented

DISCUSSION

Red: (It's hard to believe we're in this situation... One of our crewmembers is a killer. What a nightmare scenario. But this isn't the time to grumble. We have to figure who they are!)

Pink: So, uh... Where do we even start?

Red: Green found the body, so asking him first could be a start. How did you come across the body, Green?

Green: Well, I was in Reactor when the lights went out, so I went down the hallway through Lower Engine to try and fix the issue in Electrical. When I got there the door was closed, and the button didn't work because of the blackout. So I had to break it open to get inside. Someone had trashed the insides of the main breaker. But I did what I could to fix it and luckily the lights turned back on again. When I could see properly, I took a quick glance around the room, and that's how I saw the body... I didn't see anyone else on my way down, though. It was too dark.

Red: Alright then. I guess from here we should-

Purple: Self report.

Blue: Eh? What do you mean, Purple?

Purple: What I'm saying is that Green killed yellow and then reported it to make himself look less suspicious.

Green: What? No! I- I'd never do something like that!

Black: Then perhaps you have a way of confirming your innocence, Green?

Green: I-, Well-, I mean-

Purple: Case closed, suckers. Allow me to explain, in detail, how Green did it.

Green: But I didn't!

Red: (Green is struggling to put up a defence, but something isn't adding up here... I need to help him out. If there is a [contradiction] in Purple's argument, I'll need to prove it with [Evidence].)

DEBATE 1 - START

Purple: Let's start from the beginning. After our gathering in the cafeteria we all went in different directions. Green did go towards Reactor then, but he didn't actually stay there for long... Since there was [no one to witness] him leave, he could go straight to Electrical instead. There, he caused the blackout and waited for Yellow to come and fix it... And WHACK! He killed her in cold blood! After doing all that, he reported the body to take the heat off himself. But he couldn't fool me...

Proof: Pink's Account -> No Witness. *Green was witnessed walking south when he said

DEBATE 1 - END

Red: Actually, Purple, not only could someone witness. There's someone who did.

Green: There is?

Blue: Why are you surprised by that?

Purple: You'd better not be messing with me, Red.

Red: Pink. You said you heard something during the blackout, right?

Pink: Yeah, that's right. I was in Lower Engine, trying my best to work in the dark, when I heard footsteps somewhere behind me. I didn't catch who it was, though. Whoever they were, they left as fast as they came.

Red: So someone was indeed walking from Reactor through Lower Engine shortly after the blackout occurred, just like Green said. Unless someone else wants to come forward about being there at that time, it must've been Green passing through.

Purple: Hold up! If Green WAS seen by Pink, why didn't he see her, then? He said he saw no one on his way down!

White: I m- mean... P- Pink didn't actually see G- Green either, right? B- Blackout and all that.

Orange: I suppose ya don't make much noise if yer just standin' in place and fiddlin' with the screens. But Purple's idea is still valid.

Green: N- No it isn't! How could I sabotage the breaker in Electrical if I was in Reactor when the power went out!?

Orange: Yer the one sayin' the sabotage was made from Electrical. But if you're actually guilty, what yer sayin' don't mean squat. Messin' with the reactor could definitely take out the power for a lil' bit. Certainly enough time for getting' to Electrical at least. So causin' the blackout from Reactor before doin' the rest of what Purple said is very possible.

White: S- So, Green is still a suspect, huh...

Green: ...

Red: (What Orange is saying is possible... But it sounds like a long shot. I'm pretty sure there's a way to disprove his theory. And if there is, I need to find it.)

DEBATE 2 - START

Orange: Red's probably right about Pink hearin' Green in the corridor. But that only proves [Green was goin' towards Electrical] in the dark, only to stumble upon Yellow's body. The whole ordeal seems pretty sus to me.

Purple: Yeah, totally! Everything I said still fits! [Green shook up the reactor], causing some power issues, and went to Electrical under the cover of darkness. There, he [waited for Yellow] to come and try fixing the lights. And then... WHACK! He killed her in cold blood!

Orange: 'Bout a minute later, [the reactor stabilizes itself] and the lights come back. He reports the body and lies about fixin' the issues. It wouldn't be a bad plan.

Proof: My Account -> Yellow Fixed. *Yellow entered Electrical before the blackout

DEBATE 2 - END

Red: No, Green is definitively innocent here.

Orange: Interestin'... How'd ya figure that?

Red: I've seen Yellow. I was working in Storage when she passed through and went down the west corridor.

Pink: West from Storage, huh? If she did that, then she must've stopped in Electrical. I was in Lower Engine, and I never spotted her. Electrical is the only stop on the way, so... Yeah.

Blue: So Yellow went to Electrical by herself. That's good info, but why would that prove anything new? Couldn't Green have killed her still?

Red: I'm not so sure about that... Because Yellow went there long before the blackout and must've still been inside when it occurred. But she wasn't killed near the breaker. Her body was in the back.

Purple: Now that you mention it, that's pretty weird. If she was in Electrical when the power went out, she would totally try to fix the breaker herself.

White: That makes s- sense. But... What does it m- mean?

Red: It means that the blackout happened after Yellow was killed.

Green: But I only left Reactor at all because of the blackout. I had to break into Electrical in the dark... ...Which proves I'm not the culprit! You're a genius, Red!

Orange: Well I'll be damned... I suppose I owe ya an apology, Green.

Green: ... It's understandable. I guess...

Black: The culprit must have been in Electrical at the same time as Yellow. We should question whoever else was close to the room at the time. Pink?

Pink: I- I didn't do it! I was just completing my tasks at the engine!

Black: I believe you. If you were in fact the culprit, Pink, having Green not see you in that corridor is too unlikely, despite the lighting conditions. You on the other hand, Red, were supposedly alone in Storage, correct? I think you need a bit stronger of an alibi.

Green: No, no, no! I'm not believing this! If Red was the culprit, he could've easily thrown me under the bus just now!

Black: A compelling argument, but not technically evidence. I need something a little more definitive.

Red: (Of course someone would accuse me at some point. I was close to the scene. But I know I'm innocent. It's just a matter of proving it...)

DEBATE 3 - START

Green: Red could not have done it!

Black: Red could absolutely have done it. Since he was [in Storage by himself], he had ample opportunity. When he saw Yellow go into Electrical, he could easily [follow her inside]. And after committing the murder, he could just [walk out the door] before green arrived at the scene.

Green: But if he is the killer, why would he stand up for me? He could've thrown me under the bus and gotten away with it.

Black: We can discuss potential reasons after my claim has been proven valid. But as it stands, the fact remains that [only Red had a clear opportunity] to commit this murder.

Proof: Major Sabotage -> walk out the door. *I can't walk through walls

DEBATE 3 - End

Red: You're forgetting a critical detail about the sabotage, Black.

Black: Oh? Do tell.

Red: It's not just the lights that run on electricity on this ship. All of the doors do as well. Green. When you got to Electrical, the door was closed, right?

Green: Yeah that's right. Just making a gap to barely squeeze through was really tough. It was a big hassle overall. Wait a minute...

Red: That's right. When Green came to Electrical, the door was stuck shut. That means it must've been closed before the sabotage happened. No one could have opened the door before Green did, which means the killer closed themselves inside the room.

Black: And as such, Red being close to the scene proves nothing as long as he was outside when Green discovered the corpse. A well made argument. Consider my accusation revoked.

Green: I knew it couldn't be you, Red!

Pink: But if the killer couldn't use the door, how did they actually leave Electrical? Green, you opened the door and called everyone to the scene. No one else was inside, right?

Green: Uh- No... It was just me and the body.

Orange: That is strange. Mighty strange indeed... What does it all mean?

Blue: It means it's about time we went over some alibis. Personally, I was using the Security cameras the whole time. Right up until the blackout.

Red: Alright. As I've said already, I was fixing the wires in Storage.

White: I was in O2, and I saw B- Black in Weapons! He was shooting asteroids until j- just before the b- blackout.

Black: And I can confirm White was still in O2 when I was finished with the barrage. His reaction to the blackout was... Rather loud.

Pink: As I said before, I was in Lower Engine. I went there with Blue, Green, and Orange in tow, but they stopped in other rooms on that side as we passed them on the way.

Purple: I was hanging out with Yellow in the Cafeteria before she had to go and calibrate the distributor. I felt like stretching my legs a bit, so I went down the western hallway to Upper Engine.

Orange: Can confirm. She was there with me at Upper Engine right 'round the blackout happenin'.

White: D- Does this mean we're out of l- leads?

Pink: Huh? What's that supposed to mean, White?

White: We're the only l- living beings on this sh- ship, but we all have s-solid alibis. So who could've killed Yellow then!? We still don't k- know how she died!

Red: (White's right, one of us must be the killer. The question is how they pulled off the murder...)

DEBATE 4 - START

White: We are th- the [only people on this ship], so one of us k- killed Yellow!

Pink: But we all have alibis, right? Does that mean someone faked theirs somehow?

Black: Hiding places inside Electrical are very limited, so my guess is that there is some means of entering Electrical without using the door.

Blue: What means could there even be? Other than cables, there's [nothing connecting Electrical] to any other room.

Purple: I mean, Green got through without power. Maybe the killer left a gap and closed it behind them?

Green: [Definitely impossible!] It was a pain to pry those doors apart, but there was at least some leverage to do that. For pushing together, not so much.

White: Then we're s- stuck! Th- This is an impossible murder!

Proof: Map of the Ship -> Nothing else connected. *The vent is a path that leads into Electrical.

DEBATE 4 - END

Red: That's it! There is another pathway!

Blue: Nonsense. What would that pathway be?

Red: The air ducts! The killer could've come out of the vent in the back.

Orange: That... Could actually work. And it would explain a lot of things. Sure, why not? The killer was crawlin' through the vents.

Black: ...Feasible enough of a theory, I suppose...

White: B- But if they left through the vent, d- does that mean anyone who passed a vent is a s- suspect again!?

Red: Thankfully not. Not all vents are connected to each other. I got this blueprint from Admin, and it shows only two rooms that are connected to the same vent as Electrical. The killer left through the vent leading into one of those rooms. Those rooms being Medbay...and Security.

Pink: But if those are the only two rooms, that means our prime suspect is...

Blue: ... Uh-oh. Um... Uh... I didn't do it? ...

Orange: That ain't enough of an argument, Blue.

Red: I think we've heard enough...

GAME END

Appendix B - Läsförståelsetest

OVERVIEW

Everyone was gathered in the cafeteria. Then split up to do their various tasks:

- Yellow (victim) & Purple stayed in Cafeteria.
- Red went south to Storage.
- Black, & White went east to Weapons & Navigation respectively.
- Orange, Blue (killer), Green, & Pink went west to Upper Engine, Security, Reactor, & Lower Engine respectively.

Yellow (victim) goes to Electrical through storage. Purple goes west to Upper Engine, passing the closed Medbay.

Blue (killer) crawls through the vent from Security, kills Yellow (victim) in Electrical. He closes the door and causes a blackout before returning to Security through the same vent.

Green goes to Electrical to fix the lights, passing Pink in Storage. Green finds the body and reports it.

SUMMARY

(Försöksperson får 1 poäng per sak punkt av de nedan som nämns i deras sammanfattning)

- Yellow was killed.
- The murder happened in Electrical.
- Green found the body when he came to fix the lights.
- The door to Electrical was closed when Green arrived, so he broke in.
- The door could only be/was shut while power was still on.
- Blue was the killer.
- The killer left his post in Security.
- The killer got into Electrical through the vent and left the same way.
- Before leaving the scene, the killer sabotaged the breaker in Electrical.

QUESTIONS

(Försöksperson får 1 poäng per korrekt givet svar per fråga. Alla svar med samma mening är ok.)

1: "Which crew members were girls?"

Answer [Explicit]: Pink, Purple, & Yellow.

Options: One for each crewmember.

2: "In which room was everyone gathered before splitting up to do their tasks?"

Answer [Implicit]: Cafeteria

Options: One for each room on the map of the ship.

3: "Which two crew members were in the same room together when the crew split up to do their tasks?"

Answer [Explicit]: Purple & Yellow

Options: One for each crewmember.

4: "Why did the victim go to Electrical?"

Answer [Explicit]: She had to complete a task (calibrate the distributor).

Written answer.

5: "Which crew member last saw the victim alive (aside from the killer)?"

Answer [Inferred]: Red

Options: One for each crewmember.

6: "Which rooms do we know Yellow has been in?"

Answer [Inferred]: Cafeteria, Electrical, Storage.

Options: One for each room on the map of the ship.

7: "From which rooms could one get to Electrical through a vent?"

Answer [Explicit]: Security, Medbay

Options: One for each room on the map of the ship.

8: "Where was Black when the murder occurred?"

Answer [Explicit]: Weapons.

Options: One for each room on the map of the ship.

9: "In which rooms are there means to cause a blackout?"

Answer [Inferred]: Electrical, Reactor

Options: One for each room on the map of the ship.

10: "How did White react to the blackout?"

Answer [Inferred]: He didn't like it/He yelled (in fear).

Written answer.

11: "Which two crew members were in the same room together when the blackout happened?"

Answer [Explicit]: Orange & Purple.

Options: One for each crewmember.

12: "Who did Green walk past during the blackout?"

Answer [Explicit]: Pink

Options: One for each crewmember.

13: "How do we know that the killer did not exit Electrical via the door?"

Answer [Explicit]: The doors cannot be closed during a blackout.

Written answer.

14: "Why was Red in Admin?"

Answer [Inferred]: To pick up a map/blueprint for the discussion.

Written answer.

SCREENSHOTS

Rikard Larsson - Bachelor Thesis

This study only concerns the answers you give and some statistics regarding a session of gameplay.
Your personal information will not be gathered, stored or used by anyone associated with this study.
You can quit any part of the test at any time before submitting. If you do so, no data about your session will be stored.

Namnet och fotot som är kopplat till ditt Google-konto registreras när du laddar upp filer och skickar det här formuläret

Är du inte rikard.nt.larsson@hotmail.com? [Byt konto](#)

***Obligatorisk**

Thanks for participating!

In this study you will be instructed to download and play a short game with a lot of dialogue. After finishing the game, you'll be given a reading comprehension test regarding the dialogue from the game.

To confirm that you've finished a playthrough, you'll have to take a screenshot of the game's end screen and upload the picture to this form before accessing the test.

To begin, you'll be assigned a random test group by the question below:

What day of the month is your birthday? *

☒ 1-10

☐ 11-20

☐ 21-31

Nästa

Rikard Larsson - Bachelor Thesis

Namnet och fotot som är kopplat till ditt Google-konto registreras när du laddar upp filer och skickar det här formuläret

Är du inte rikard.nt.larsson@hotmail.com? [Byt konto](#)


***Obligatorisk**

Test Group - VN

Please download the game found at the following link:
<https://drive.google.com/file/d/1r5PEB-YGWWYQxHo1u0d7FEzhzTF7iefP4/view?usp=sharing>

Play it to completion and take a screenshot of the end screen.
Estimated playtime: No longer than an hour.
Upload that screenshot here to continue the study.

Slutgiltigt spelresultat (VN) *

 Combat Options... ✕

Bakåt Nästa

Rikard Larsson - Bachelor Thesis

Namnet och fotot som är kopplat till ditt Google-konto registreras när du laddar upp filer och skickar det här formuläret

Är du inte rikard.nt.larsson@hotmail.com? [Byt konto](#)

***Obligatorisk**

Summary

Describe the incident discussed by the in game characters. (2000 letters max) *

Yellow was killed.
The murder happened in Electrical.
Green found the body when he came to fix the lights.
The door to Electrical was closed when Green arrived, so he broke in.
The door could only be/was shut while power was still on.
Blue was the killer
The killer left his post in Security.
The killer got into Electrical through the vent and left the same way.
Before leaving the scene, the killer sabotaged the breaker in Electrical.

Bakåt Nästa

Rikard Larsson - Bachelor Thesis

Namnet och fotot som är kopplat till ditt Google-konto registreras när du laddar upp filer och skickar det här formuläret

Är du inte rikard.nt.larsson@hotmail.com? [Byt konto](#)

***Obligatorisk**

Questions

Which crewmembers were girls?

☐ Black

☐ Blue

☐ Green

☐ Orange

☐ Red

☐ Pink

☐ Purple

☐ White

☐ Yellow

In which room was everyone gathered before splitting up to do their tasks?

☐ Admin

☐ Cafeteria

☐ Communication

☐ Upper Engine

☐ Lower Engine

☐ Marthau

Questions

Which crewmembers were girls?

- ☐ Black
- ☐ Blue
- ☐ Green
- ☐ Orange
- ☐ Red
- ☒ Pink
- ☒ Purple
- ☐ White
- ☒ Yellow

In which room was everyone gathered before splitting up to do their tasks?

- ☐ Admin
- ☒ Cafeteria
- ☐ Communication
- ☐ Upper Engine
- ☐ Lower Engine
- ☐ Medbay
- ☐ Navigation
- ☐ O2
- ☐ Reactor
- ☐ Shields
- ☐ Security
- ☐ Storage
- ☐ Weapons

Rensa marking

Which two crew members were in the same room together when the crew split up to do their tasks?

- ☐ Black
- ☐ Blue
- ☐ Green
- ☐ Orange
- ☐ Red
- ☐ Pink
- ☒ Purple
- ☐ White
- ☒ Yellow

Why did the victim go to Electrical?

She has to calibrate the distributor

Which crewmember last saw the victim alive (aside from the murderer)?

- ☐ Black
- ☐ Blue
- ☐ Green
- ☐ Orange
- ☒ Red
- ☐ Pink
- ☐ Purple
- ☐ White
- ☐ Yellow

Rensa marking

Which rooms do we know Yellow has been in?

- ☐ Admin
- ☒ Cafeteria
- ☐ Communication
- ☒ Electrical
- ☐ Upper Engine
- ☐ Lower Engine
- ☐ Medbay
- ☐ Navigation
- ☐ O2
- ☐ Reactor
- ☐ Shields
- ☐ Security
- ☒ Storage
- ☐ Weapons

From which rooms could one get to Electrical through a vent?

- ☐ Admin
- ☐ Cafeteria
- ☐ Communication
- ☐ Electrical
- ☐ Upper Engine
- ☐ Lower Engine
- ☒ Medbay
- ☐ Navigation
- ☐ O2
- ☐ Reactor
- ☐ Shields
- ☒ Security
- ☐ Storage
- ☐ Weapons

Where was Black when the murder occurred?

- ☐ Admin
- ☐ Cafeteria
- ☐ Communication
- ☐ Electrical
- ☐ Upper Engine
- ☐ Lower Engine
- ☐ Medbay
- ☐ Navigation
- ☐ O2
- ☐ Reactor
- ☐ Shields
- ☐ Security
- ☐ Storage
- ☒ Weapons

Rensa marking

In which rooms are there means to cause a blackout?

- ☐ Admin
- ☐ Cafeteria
- ☐ Communication
- ☐ Electrical
- ☐ Upper Engine
- ☐ Lower Engine
- ☐ Medbay
- ☐ Navigation
- ☐ O2
- ☒ Reactor
- ☐ Shields
- ☐ Security
- ☐ Storage
- ☐ Weapons

Rensa marking

How did White react to the blackout?

It startled him|

Which two crew members were in the same room together when the blackout happened?

- ☐ Black
- ☐ Blue
- ☐ Green
- ☒ Orange
- ☐ Red
- ☐ Pink
- ☒ Purple
- ☐ White
- ☐ Yellow

Who did Green walk past during the blackout?

- ☐ Black
- ☐ Blue
- ☐ Green
- ☐ Orange
- ☐ Red
- ☒ Pink
- ☐ Purple
- ☐ White
- ☐ Yellow

Rensa marking

How do we know that the killer did not leave Electrical via the door?

ithout power, but this door was closed anyway

Why was Red in Admin?

To pick up a blueprint for the discussion|

Bakåt

Nästa

Your Experience

Just a few questions about your experience.
This is not a test in any way, just give an honest review :)

Did you enjoy the game experience?

0 1 2 3 4 5
No! ☐ ☐ ☐ ☐ ☐ ☐ Yes, very much!

Have you expienced any game similar to this one before?

0 1 2 3 4 5
No! ☐ ☐ ☐ ☐ ☐ ☐ Yes, I've played several!

Are you familiar with Among Us?

0 1 2 3 4 5
No! ☐ ☐ ☐ ☐ ☐ ☐ Yes, I play it all the time!

Anything you'd like to add?

Ditt svar

Bakåt

Skicka

Appendix C - Amelia Earhart QRI-test

Amelia Earhart

Amelia Earhart was an adventurer and a pioneer in the field of flying. She did things no other woman had ever done before.

During World War I, Earhart worked as a nurse. She cared for pilots who had been hurt in the war. Earhart listened to what they said about flying. She watched planes take off and land. She knew that she, too, must fly.

In 1928, Earhart became the first woman to cross the Atlantic in a plane. But someone else flew the plane. Earhart wanted to be more than just a passenger. She wanted to fly a plane across the ocean herself. For four years, Earhart trained to be a pilot. Then, in 1932, she flew alone across the Atlantic to Ireland. The trip took over fourteen hours.

Flying may seem easy today. However, Earhart faced many dangers. Airplanes had just been invented. They were much smaller than our planes today. Mechanical problems happened quite often. There were also no computers to help her. Flying across the ocean was as frightening as sailing across it had been years before. Earhart knew the dangers she faced. However, she said, "I want to do it because I want to do it. Women must try to do things as men have tried. When they fail, their failure must be a challenge to others."

Earhart planned to fly around the world. She flew more than twenty thousand miles. Then, her plane disappeared somewhere over the huge Pacific Ocean. People searched for a long time and many gave up. Earhart and her plane were never found, but people are still looking.

Adapted from *Scott Foresman Social Studies, Grade 4: Regions of Our Country and Our World*, copyright © 1983 by Scott Foresman and Co. Adapted by permission of Pearson Education, Inc.

Setting/Background

- _____ Amelia Earhart was an adventurer.
- _____ During World War I she was a nurse who cared for pilots who had been hurt.
- _____ Earhart watched planes take off and land.

Goal

- _____ She knew that she must fly.
- _____ Earhart was the first woman to cross the Atlantic in a plane.
- _____ Someone else flew the plane.
- _____ She wanted to fly a plane across the ocean.

Events

- _____ Earhart trained to be a pilot.
- _____ In 1932 she flew alone across the Atlantic to Ireland.
- _____ Earhart faced dangers.
- _____ Airplanes were smaller.
- _____ There were no computers.
- _____ Earhart said women must try to do things as men have tried.
- _____ Earhart planned to fly around the world.

Resolution

- _____ Her plane disappeared over the Pacific Ocean.
- _____ People searched for a long time, and many gave up.
- _____ Earhart and her plane were never found.
- _____ People are still looking.

1. What was Amelia Earhart's main goal?

Implicit: to fly; or to do things that were challenging

2. What was Amelia Earhart doing in a plane when she first crossed the Atlantic?

Explicit: she was a passenger

3. How long did it take Amelia Earhart when she flew alone across the Atlantic?

Explicit: over fourteen hours

4. Why would flying *alone* across the Atlantic be an especially dangerous thing to do?

Implicit: it was a long trip; there was no one to help with problems; or there was no one to help her stay awake or give her a break

5. What was one of the dangers of flying in those early days?

Explicit: small planes; mechanical problems; or no computers

6. How do we know Amelia Earhart believed in equal rights for women?

Implicit: she said women should try to do things just as men have tried

7. What was Amelia Earhart trying to do when her plane disappeared?

Explicit: fly around the world

8. Why do you think her plane was never found?

Implicit: probably sank in the ocean; ocean was so big; or plane was very small

Appendix D - F-testtabell för hypotesprövning av linjers lutningar

	VN			PW			DR			Totalt		
	F	DFn, DFd	P	F	DFn, DFd	P	F	DFn, DFd	P	F	DFn, DFd	P
<i>Poäng - Tid</i>	4.40	1, 6	0.08	0.63	1, 9	0.45	0.28	1, 8	0.61	6.1×10^{-3}	1, 28	0.94
<i>Poäng - Antal Felval</i>	NA	NA	NA	3.72	1, 5	0.11	0.12	1, 7	0.74	1.50	1, 14	0.24
<i>Poäng - Nöje</i>	3.38	1, 6	0.12	0.20	1, 9	0.67	8.4×10^{-4}	1, 8	0.98	1.01	1, 27	0.32
<i>Poäng - Visual Novelvana</i>	2.02	1, 6	0.21	1.90	1, 9	0.20	2.20	1, 8	0.18	0.18	1, 27	0.67
<i>Poäng - Among Us Erfarenhet</i>	0.73	1, 6	0.43	3.15	1, 9	0.11	0.54	1, 8	0.48	1.26	1, 27	0.27

Linjär regression enligt minsta kvadratmetoden gjordes för att undersöka samband mellan variablerna poäng, antal felval i spelen, nöje, visual novel-vana och Among Us-erfarenhet. Genom F-test med Prism (GraphPad 1994) undersöktes om variansen hos datapunkterna var tillräckligt liten för att stödja att linjernas lutningar signifikant skiljde sig från 0. Nollhypotesen är generellt att lutningen är 0. Om en skillnad från lutningen 0 inte kan visas kan inte nollhypotesen förkastas och heller inget samband mellan variablerna påvisas (Somekh & Lewin 2005, s.266-235).

F, F-ratio; DFn, frihetsgrader i täljaren; DFd, frihetsgrader i nämnaren; P, sannolikhet.

I inget av fallen uppnåddes en signifikansnivå <0.05 . Inga samband mellan undersökta variabler kunde därför påvisas.