LEX

Luxemburger Experimental-praktikum Journal

Zeitschrift psychologischer Forschung
Revue de recherche en psychologie

Band 17, Jahrgang 2023

Université du Luxembourg
Bachelor scientifique en psychologie

ISSN 3093-1045

Luxemburger Experimentalpraktikum Journal
Band 17, Jahrgang 2023

Inhalt

Long Covid und Chronic Fatigue: Validierung der deutschen Übersetzung des DePaul Symptom – Questionnaires-2

Laura Klein, Maryse Koedinger, Julie Krier, Sarah Niesmann, Tobias Magnus Prinz, Nena Rama und Isabell Žulič

Betreuung: Nina Buntić, M. Sc.

Etwa 250.000 Deutsche leiden an Myalgischer Enzephalomyelitis und chronischem Fati- gue Syndrom (Deutsche Gesellschaft für ME/CFS, 2022) und bis zu 59,5 % (Fischer et al., 2022) der von Covid-19 Genesenen sind von Langzeitfolgen mit ähnlichen Sympto- men betroffen. Dieses Syndrom wird in der Literatur als Post Covid beziehungsweise Long Covid bezeichnet. Um Betroffene besser behandeln zu können, wird ein schnelles und kostengünstiges Diagnoseinstrument benötigt. Der *DePaul Symptom Question- naire-2* wurde entwickelt, um das Vorhandensein und den Schweregrad der Myalgischer Enzephalomyelitis und chronischem Fatigue Syndrom festzustellen. Erstes Ziel der vor- liegenden Studie war es, die deutsche Übersetzung des *DePaul Symptom Question- naire-2* zu validieren, um ein zuverlässiges Bewertungsinstrument für die Behandlung und Erforschung von Myalgischer Enzephalomyelitis/chronischem Fatigue Syndrom und Long Covid bereitzustellen. Zweitens wurde die Wirksamkeit von Covid-19-Impfstoffen hinsichtlich des Schutzes vor Long Covid erforscht. Darüber hinaus untersuchten wir die Unterschiede in depressiven Symptomen und krankheitsbezogenen Symptomprofilen der verschiedenen Gruppen. Eine Stichprobe von 502 Erwachsenen mit Myalgischer Enzephalomyelitis/chronischem Fatigue Syndrom, Long Covid oder ohne Erkrankungen wurde gebeten, unsere deutsche Online-Umfrage zu beantworten, die aus einem Frage- bogen zu persönlichen Informationen, Gesundheits- und Covid-19-bezogenen Fragen sowie dem *DePaul Symptom Questionnaire-2, Short Form 36* und dem *Patient Health Questionnaire-9* bestand. Die statistische Analyse zeigt, dass die konvergente und dis- kriminante Validität sowie die Reliabilität tatsächlich gegeben sind. Darüber hinaus wurde beim Vergleich der Gruppen nicht geimpfter mit einmal geimpften Personen und der Gruppen nicht geimpfter mit vollständig geimpften Personen (dreimal) keine signifikante Verringerung der Entwicklung von Long Covid festgestellt. Darüber hinaus schei- nen Personen, die an Long Covid erkrankt sind, über mehr depressive Symptome zu berichten als die gesunde Bevölkerung, aber weniger depressiv zu sein als die Bevölke- rung, die an Myalgischer Enzephalomyelitis/chronischem Fatigue Syndrom leidet. Schließlich kann davon ausgegangen werden, dass Personen, die an Long Covid leiden, im Vergleich zu Personen, die an Myalgischer Enzephalomyelitis/chronischem Fatigue Syndrom erkrankt sind, keine Unterschiede in den Symptomen, sondern eher in der Symptomschwere aufweisen.

Einleitung

Im Dezember 2019 wurde der erste offi- zielle Fall des neuartigen Coronavirus 2019-nCoV (später SARS-CoV-2 ge- nannt) in China nachgewiesen. Eine In- fektion mit dem Virus kann die Atem- wegserkrankung Covid-19 (Coronavirus-Krankheit 2019) verursachen. Durch Tröpfchen- und Aerosolübertragung hat sich die Krankheit innerhalb weniger Monate weltweit verbreitet. Am 27. Ja- nuar 2020 wurde der erste Coronafall in Deutschland (Hillebrandt, 2022) und etwa einen Monat später am 29. Februar 2020 in Luxemburg nachgewiesen (Ta- geblatt.lu, 2020). Am 11. März 2020 wurde die Situation der Weltgesund- heitsorganisation (WHO) als Pandemie eingeordnet (WHO.int, 2023). Zum Schutz der Volksgesundheit wurden kurz darauf in vielen Ländern Lockdowns verhängt. Bei der Covid-19-Pan- demie handelte es sich um eine schwer- wiegende Gesundheitskrise, auch wenn diese von verschiedenen Experten und Institutionen bereits für beendet erklärt wurde, ist das Coronavirus noch

immer Bestandteil unseres alltäglichen Lebens. Diese Situation zu bewältigen, umfasst nicht nur die Infektions-Prävention und die Behandlung der aktuell Erkrankten, sondern auch die Behandlung und Erfor- schung derer, die nach Genesung wei- terhin unter Beeinträchtigungen leiden. Nach einer akuten Covid-19-Infektion leiden bis zu 60 % der Betroffenen in den folgenden sechs Monaten an min- destens einem anhaltenden Symptom (Fernández-de-las-Peñas, 2021). Der Schweregrad der Erkrankung variiert von leichter bis hin zu extremer Einschränkung im täglichen Leben. Die Symptome von Long Covid ähneln de- nen von Myalgischer Enzephalomyeli- tis/chronischem Fatigue Syndrom einer schweren chronischen Erkrankung.

Theoretischer Hintergrund

Laut der Arbeitsdefinition der US-ameri- kanischen Seuchenschutzbehörde CDC handelt es sich bei Long Covid um Symptome, die nach der Anfangsphase der Infektion, spätestens aber vier Wo- chen nach der Covid-19 Erkrankung auf- treten. Die Symptome können multisys- temisch sein, in einem remittierenden Muster auftreten, sich weiter verschlim- mern und auch Jahre später tödliche Komplikationen verursachen. Es han- dele sich um verschiedene Entitäten mit unterschiedlichen biologischen Ursa- chen und Risikofaktoren (CDC.gov, 2022). Die Weltgesundheitsorganisation (WHO) versteht unter Post Covid gesundheitliche Einschränkungen, die drei Monate nach der Infektion auftreten und mindestens zwei Monate andauern. Diese können auch wiederkehren oder in der Stärke variieren. Die Symptome von Long Covid sind sehr divers und um- fassen Kurzatmigkeit, Autoimmunreaktionen, neurologische Störungen bis hin zu kognitiven Aussetzern (WHO.int, 022). Eine komorbide Erkrankung so- wie die Ausprägung von mehr als fünf Symptomen während der Covid-19-In- fektion scheinen Risikofaktoren für Long Covid zu sein. Darüber hinaus ist die Prävalenz bei Frauen doppelt so hoch wie bei Männern. Die genaue medizini- sche Ursache von Long Covid wurde bisher nicht identifiziert, daher sind die Behandlungsmöglichkeiten begrenzt. Allerdings kann ein Teil der Symptome auf die Beschädigung von Organsyste- men, einer chronischen Entzündung oder Autoimmunreaktionen zurückge- führt werden (Raveendran et al, 2021). Neben einer symptomatischen Behand- lung einzelner Beschwerden kommen psychotherapeutische und psychosomatische Rehabilitation sowie zahlrei- che experimentelle Therapien, wie die Hyperbare Sauerstofftherapie (HBO) in Frage.

Zur Prävention einer Infektion mit dem Coronavirus SARS-COV-2 wur- den zahlreiche nichtpharmakologische Schutzmaßnahmen ergriffen. Zu diesen gesundheitspolitischen Interventionen gehörten Lockdowns und teilweise Mas- kenpflichten. Daneben wurden ei- gens zur Pandemiebekämpfung Impfstoffe gegen Covid-19 entwickelt. Bei diesen handelte es sich meist um mRNA- (messenger ribonucleic acid) oder vek- torbasierte Vakzine, die von verschiede- nen Herstellern kreiert wurden. Seit Ende Dezember 2020 werden diese in Luxemburg (gouvernement.lu, 2020) und Deutschland (Teska, 2022) verab- reicht. Aktuell sind 62,6 % der Deut- schen dreifach geimpft (Meyer, 2023).

Bei Myalgischer Enzephalomy- elitis/chronischem Fatigue Syndrom handelt es sich um eine chronische Er- krankung, die meist verwendeten Diag- nosekriterien sind die kanadischen Kon- sens-Kriterien (Canadian Consensus Criteria). Diese müssen für mindestens sechs Monate auftreten, umfassen physische und mentale Fatigue, anstren- gungsbedingtes Unwohlsein sowie Schlafprobleme. Darüber hinaus leiden Betroffene unter weiteren neurologi- schen oder kognitiven Beeinträchtigun- gen. Um die Diagnose Myalgischer En- zephalomyelitis/chronischem Fatigue Syndrom vergeben zu können, müssen zudem entweder Teile des autonomen Nervensystems, des neuroendokrinen oder Immunsystems betroffen sein (cfids-me.org). Bisher gibt es keine In- formationen über die genaue Genese von Myalgischer Enzephalomyeli- tis/chronischem Fatigue Syndrom, aller- dings wird vermutet, dass es mit Virusin- fektionen und post-viralen Herzerkran- kungen zusammenhängen könnte (Renz-Polster et al., 2022). Auch hier werden vornehmlich die behandelbaren Symptome therapiert.

Aufgrund der Aktualität der Long Covid-Problematik und der unzureichen- den Daten in Bezug auf Long Covid und Myalgischer Enzephalomyelitis/chroni- schem Fatigue Syndrom ist eine weitere Erforschung beider Erkrankun-

gen drin- gend notwendig. Um eine ideale Thera- pie zu entwickeln, sollte zunächst die Entstehung der Erkrankungen unter- sucht werden. Um passende Subjekte zu identifizieren, kann man im Engli- schen den *DePaul Symptom Questio- naire-2* verwenden. Dieser basiert auf den kanadischen Konsens-Kriterien - ein solches Instrument liegt in deutscher Sprache nicht vor. In früheren Studien wurde bereits eine englische Version des DePaul Symptom Questionnaire-2 validiert (Bedree et al., 2019). Ziel dieser Studie ist es, die deutsche Übersetzung des *DePaul Symptom Questionnaire-2* zu validieren - mit einem Teilnehmerpool von Menschen, die von Myalgischer Enzephalomyelitis/chronischem Fatigue Syndrom oder Long Covid betroffen sind, sowie gesunden Personen - um ein kos- teneffektives Diagnoseinstrument für Ärzte zu finden, mit dem sie den Schwe- regrad der individuellen Beeinträchti- gung durch die zuvor genannten Erkran- kungen nach den kanadischen Konsensus-Kriterien beurteilen können. Dar- über hinaus wollen wir untersuchen, ob Covid-19-Impfstoffe vor der Entstehung von Long Covid schützen und ob die An- zahl der erhaltenen Impfungen einen Unterschied macht. Zusätzlich wollten wir herausfinden, ob diese Krankheiten von psychischen Erkrankungen begleitet werden und ob sich unterschiedliche Symptomprofile für Myalgischer Enze phalomyelitis/chronischem Fatigue Syn- drom und Long Covid identifizieren las- sen.

Hypothesen

Um die Beziehungen zwischen unterschiedlichen Faktoren zu untersuchen, setzen sich unsere Hypothesen wie folgt zusammen: Hypothese 1: Die Impfung verringert das Risiko einer Entwicklung von Long Covid.
Hypothese 2: Ein vollständiger Impfschutz schützt besser vor Long Covid als keine Impfung.
Hypothese 3: Personen, die an Long Covid leiden, haben eher depressive Symptome als gesunde Menschen.
Hypothese 4: Personen mit Myalgischer Enzephalomyelitis/chronischem Fatigue Syndrom zeigen eine andere Symptomatik als Personen mit Long Covid.

Methoden

Im Folgenden werden die Stichprobenkonstellation sowie das Material und die Vorgehensweise erläutert.

Stichprobe

Mit Hilfe der Faustregel von fünf Ver- suchspersonen pro Item haben wir die angestrebte Stichprobengröße auf N = 500 festlegen können (Tsang, S., Royse,

C. F., & Terkawi, A. S. (2017)). Tatsäch- lich haben 502 Personen an unserer On- line-Studie teilgenommen. Davon waren 191 an Long Covid Erkrankte, 227 an chronischem Erschöpfungssyn- drom/Myalgischer Enzephalomyelitis Erkrankte und 84 gesunde Teilneh- mende. Vor Beginn der Durchführung haben wir das Mindestalter für eine Teil- nahme auf 18 Jahre begrenzt. Das Durchschnittsalter unserer Versuchs- personen entspricht ungefähr 37 Jahren (M = 36.87, SD = 13.27). Davon identifi- zierten sich 79 % mit dem weiblichen und 20.8 % mit dem männlichen Ge- schlecht. 0.2 % fühlen sich keinem die- ser beiden Geschlechter zugehörig.
Der Großteil unserer Testperso- nen lebt in Luxemburg oder Deutschland. Wichtig zu erwähnen ist vor allem, dass Deutschkenntnisse eine Voraussetzung für die Teilnahme an der Studie waren. Dies haben wir sichergestellt durch deren Erfragung mit folgendem Item: Wie gut beherrschen Sie die deutsche Spra- che? (Item acht). Nach Ausfüllen des Fragebogens hatten die Versuchsperso- nen die Möglichkeit an einer Verlosung von zehn 20€ Amazon-Gutscheinen teil- zunehmen.

Material

SoSci Survey wurde verwendet, um den Fragebogen zu erstellen. Dort wurden außerdem die Antworten der Versuchs- personen anonym erhoben und im Nachhinein sicher gespeichert. Dabei wurden die erhobenen Daten separat von den E-Mail-Adressen der Versuchs- personen aufbewahrt.

DEPAUL SYMPTOM QUESTIONNAIRE. Da der Schwerpunkt unserer Studie auf der Validierung der deutschen Übersetzung des *DePaul Symptom Questionnaire-2* (DSQ-2) liegt, haben wir ein Expertenko- mitee gegründet. Dort wurde in wö- chentlichen Treffen der englischspra- chige *DePaul Symptom Questionnaires-2* (DSQ-2) zunächst von einigen auf Deutsch übersetzt und im Nachhinein von anderen erneut in die englische Sprache rückübersetzt. Die englische Originalversion wurde daraufhin mit der rückübersetzten Version verglichen, um mögliche Diskrepanzen festzustellen.

Beim *DePaul Symptom Questi- onnaire* handelt es sich um eine Selbst- einschätzung, der von Leonard Jason und dem chronischem Fatigue Syn- drome Projekt an der DePaul University in Chicago, Illinois, USA im Jahr 2010 entwickelt wurde. Der *DePaul Symptom Questionnaire* operationalisiert die Ca- nadian Consensus Kriterien und liefert somit konkrete Richtlinien um Myalgi- sche Enzephalomyelitis/chronisches Fatigue Syndrom zu erfassen. Er be- steht insgesamt aus 54 Fragen und er- fasst die Hauptsymptome wie Fatigue, Post-exertional Malaise, Schlaf, Schmerzen, neurologische/kognitive Störungen, sowie autonome, neuroen- dokrine und auf Immunität bezogene Symptome. Die Versuchspersonen mussten hier den Schweregrad und die Häufigkeit ihrer Symptome angeben, die mit Hilfe einer fünfstufigen Likert-Skalie- rung erhoben wurden. In einem norwe- gischen Vergleich mit ärztlichen Beurteilungen erzielte der *DePaul Symptom Questionnaire* eine Sensitivität von 92 % und eine Spezifität von 75 %. Dies deu- tet darauf hin, dass der *DePaul Symp- tom Questionnaire* ein nützliches Instru- ment zur Feststellung und zum Scree- ning von Symptomen ist, jedoch ist eine ärztliche Nachuntersuchung erforderlich, um die Diagnose zu bestätigen und mögliche abgrenzbare medizinische und psychiatrische Störungen zu identi- fizieren.
Im Jahr 2019 wurde der Frage- bogen getestet und überarbeitet mit dem Ziel, seine psychometrischen Eigen- schaften zu verbessern, sowie seine di- agnostische Zuverlässigkeit zu erhöhen und die für die Falldefinitionen erforder- lichen Symptome zu erfassen. Das daraus resultierende Instrument wurde *DePaul Symptom Questionnaire-2* (DSQ-2) genannt, welchen wir für un- sere Studie verwendet haben. Dabei be- schränkten wir uns, in der für uns nützli- chen Auswahl hinsichtlich der Symp- tomerfassung auf die Items 1 bis 91. Um zusätzlich die Falldefinition-Kriterien er- füllen zu können, haben wir weitere Items ausgewählt. Insgesamt haben wir somit Item 1 bis 91, Item 92, 94, 106,
119, 127 und 128 für unsere Studie verwendet. Wir haben uns auf diese Items beschränkt, um Teilnehmende nicht zu überfordern und zu viel ihrer Zeit in An- spruch zu nehmen. Für unsere Studie wurden Items der Long Covid Sympto- matik ergänzt.
Darüber hinaus wurden einige Items von uns erstellt, um weitere sozio- demographische Daten, die nicht im *DePaul Symptom Questionnaire-2* ent- halten sind, abzufragen. Dazu gehören

z.B. Fragen zur Immunisierung, Covid- 19, Long Covid, sowie psychische Er- krankungen.

GESUNDHEITSFRAGEBOGEN SHORT *FORM-36.* In unserer Studie wurde au- ßerdem die *Short Form-36* des Gesund- heitsfragebogen abgefragt. Dieser be- steht aus einer 36-teiligen Skala zur Er- fassung des Gesundheitszustands und der Lebensqualität (Ware & Sherbourne, 1992), die in acht Gesundheitskonzepte unterteilt werden. In der Standardform werden Teilnehmende gebeten, auf Fra- gen zu antworten, wie sie sich in der ver- gangenen Woche gefühlt haben.

PATIENT HEALTH QUESTIONNAIRE-9. Zusätzlich wurden Items aus der deut- schen Version des *Patient Health Ques- tionnaire-9* (PHQ-9) verwendet, um über das Likert-Skalenniveau eine Neigung zur Depression festzustellen.
Bei dem *Patient Health Questi- onnaire-9* handelt es sich um einen Selbstauskunft-Fragebogen zur Diag- nostik des Schweregrads einer Depres- sion. In den neun Items werden die de- pressiven Symptome und deren Häufigkeit während der vergangenen zwei Wo- chen erfragt. Die Antwortmöglichkeiten umfassen eine vierstufige Likert Skala (von 0 = gar nicht, 1 = einige Tage, 2 = mehr als die Hälfte der Tage und 3 = fast jeden Tag). Die Item-Werte werden ad- diert - höhere Gesamtwerte aller Items sind Indikator für eine relevante depres- sive Symp-

tomatik.

Datenanalyse

Zusätzlich haben wir den Fragebogen auf Reliabilität und Konstruktvalidität ge- prüft sowie die interne Konsistenz be- rechnet, um diesen zu validieren. Die Berechnung der internen Konsistenz er- folgte mithilfe von Chronbachs Alpha. Die konvergente und diskriminante Vali- dität wurde anhand der Korrelation mit den Subskalen des *Short Form-36* über- prüft. Die Datenverarbeitung und die Hy- pothesenüberprüfung erfolgten über IBM SPSS Statistics 27.

Durchführung

Pilottestung

Es wurde eine Pilottestung durchgeführt, bei der das Expertenkomitee sieben Versuchspersonen getestet hat. Dies diente dem Zweck zu kontrollieren, ob der Fragebogen alle Symptome abdeckt und die Dauer einzuschätzen, die benö- tigt wird, um den Fragebogen vollständig zu beantworten. Dabei wurde außerdem die Verständlichkeit der Fragen und de- ren Inhalte geprüft. Um bei der Bearbei- tung des Fragebogens technische Feh- ler auszuschließen und um sicher zu stellen, dass der Fragebogen auf allen Geräten funktioniert, hat das Experten- komitee ihn auf verschiedenen Geräten und Browsern ausgefüllt.

Rekrutierung

Für die Rekrutierung wurden Flyer am Campus Belval, Limpertsberg und Kirch- berg der Universität Luxemburg aufge- hangen und über Social Media geteilt. Des Weiteren wurden spezifisch auf Fa- cebook Selbsthilfegruppen für an Long Covid Erkrankte, sowie für an Myalgi- scher Enzephalomyelitis/chronischem Fatigue Syndrom Erkrankte, ange- schrieben und über die Studie informiert.

Datenerhebung

Der Online-Fragebogen wurde von den Versuchspersonen von ihren eigenen Geräten aus, über die Online-Plattform SoSci Survey ausgefüllt. Die Dauer der Bearbeitung des Fragebogens variiert zwischen 30 und 60 Minuten und ist ab- hängig vom Gesundheitszustand sowie der Bearbeitungsgeschwindigkeit der je- weiligen Person. Indem es nicht möglich war, auf die nächste Seite zu gelangen, wenn nicht alle Fragen der vorherigen Seite beantwortet wurden, waren Teil- nehmende dazu gezwungen, alle Fragen zu beantworten. Jeder Versuchs- person stand offen, die Befragung jeder- zeit abzubrechen sowie die Befragung zu unterbrechen und zu einem späteren Zeitpunkt fortzuführen. Die Erhebung

der Daten erfolgte vom 31. Oktober 2022 bis einschließlich dem 20. November 2022.

Ergebnisse

Im Folgenden werden einerseits die Re- sultate der Validierung des Fragebo- gens und andererseits die Validierung der vier Hypothesen erläutert. Bei der Validierung des Tests wurde Chron- bachs Alpha zur Berechnung der Reliabilität sowie eine Analyse der Konstrukt- validität verwendet. Bei der Untersu- chung der ersten beiden Hypothesen wurde eine Odds Ratio durchgeführt, um das Quotenverhältnis zu errechnen. Auf Grund einer nicht normalverteilten Stich- probe wurden für die anderen beiden Hypothesen einen Mann-Whitney-U Test und einen Kruskal-Wallis-H Test verwendet, um Unterschiede zwischen den Gruppen nachzuweisen.

Validierung der deutschen Überset- zung des DePaul Symptom Questi- onnaire-2

Bedingt dadurch, dass der *DePaul Symptom Questionnaire-2* ein neuer Fragebogen in deutscher Sprache ist, wurde die faktorielle Validität mit Hilfe einer explorativen Faktorenanalyse untersucht. Diese Art der Faktorenanalyse wurde auch deshalb gewählt, weil sie das Standardverfahren für den *DePaul Symptom Questionnaire* darstellt. Die explorative Faktorenanalyse wurde mit der Hauptachsenmethode der Faktorex-

traktion unter Verwendung einer Promax-Rotation (Kappa = 4) durchgeführt. Faktoren mit Ladungen unter .40 wurden aus Gründen der besseren Lesbarkeit weggelassen. Alle Faktoren mit Eigenwerten über 1 wurden hingegen beibehalten. Es wurde eine Zwölf-Faktoren-Struktur gefunden mit folgenden Werten: 1.000, .724, .710, .644, .557, .637, .452
, .537, .535, .547, .468 und .370. Dies kann die Gesamtvarianz des *DePaul Symptom Questionnaire-2* erklären.

Zum Nachweis der Reliabilität wurde eine Analyse mit Cronbachs Alpha durchgeführt, welche auf eine hohe interne Konsistenz der Testitems hinweist mit Werten zwischen .63 und .97. Folgende Werte wurden errechnet: .974, .956, .852, .836, .835, .
887, .830, .834, .849, .826, .804 und .632.

Zur Ermittlung der Konstruktvalidität werden mehrere bivariate Korrelationen in *SPSS IBM Statistics 27* zwischen den Faktorwerten und den unterschiedlichen Subskalen des Fragebogens *Short Form-36* berechnet. Es ergeben sich signifikante Ergebnisse ($p < .001$), welche ein Beleg für die Konstruktvalidität sind.

Die konvergente Validität untersucht die Übereinstimmung der Faktorwerte mit Faktoren, die dasselbe Konstrukt messen. Hierzu werden die Korrelationen mit den Faktoren eins bis vier des Fragebogens *Short Form-36* zur körperlichen Gesundheit (körperliche Funktionsfähigkeit, körperliche Rolle, körperliche Schmerzen und allgemeine Gesundheit) berechnet. Es resultieren starke negative Korrelationen mit einem Bereich der Korrelationskoeffizienten zwischen $r = -.753$ und $r = -.854$. Dementsprechend wird eine hohe Übereinstimmung festgestellt.

Diskriminante Validität zeigt sich über die Übereinstimmung der Faktorwerte mit Faktoren, die ein unterschiedliches Konstrukt messen. Die Korrelationen mit den Faktoren neun und zehn des Fragebogens *Short Form-36* zur mentalen Gesundheit (emotionales Wohlbefinden und emotionale Rolle) zeigen schwache negative Korrelationen. Der Bereich der Korrelationskoeffizienten befinden sich zwischen $r = -.055$ und $r = -.258$. Man erkennt eine geringe Übereinstimmung mit Faktoren, die ein anderes Konstrukt messen.

Hypothesen

Das zweite Ziel der Studie beinhaltet die Überprüfung der Hypothesen in einer deutschsprachigen Stichprobe. Die

auf vorherigen Studien, sondern sie wurden unabhängig davon und im Interesse des Expertenkomitees formuliert.

Hypothese 1

Die Hypothese kann nicht bestätigt wer- den. Es existiert kein signifikanter Unter- schied in der Wahrscheinlichkeit, an Long Covid zu erkranken, wenn man vor einer Infektion einmal geimpft ist oder nicht, OR = .657, (95% CI: .287, 1.589). Durch eine nicht-gegebene Signifikanz ist der Interpretationsspielraum deutlich eingeschränkt, wodurch diese Hypo- these verworfen werden muss. Siehe Tabelle 1.

Wenn die Resultate signifikant wären, würde dies bedeuten, dass eine Impfung vor Long Covid schützen würde und man seltener daran erkranken würde. Das Risiko an Long Covid zu er- kranken - mit einer Impfung - wäre 0.657-mal kleiner als mit keiner.

Hypothese 2

Die Hypothese kann nicht bestätigt wer- den. Die Resultate zeigen, dass es kei- nen signifikanten Unterschied zwischen keiner oder drei Impfungen vor Infektion gibt, Long Covid zu bekommen, OR = 1.419, (95% CI: .769, 2.620). Siehe Tabelle 1.

Wenn die Ergebnisse signifikant wären, würde dies bedeuten, dass die Wahrscheinlichkeit, Long Covid zu be- kommen, nach drei Impfungen vor der Infektion 1.4-mal höher wäre, als wenn man keine Impfung zuvor erhielt.

Tabelle 1: Impfungen vor Covid-19 Infektion

	Gesunde	Long Covid
Nicht geimpft	18	10
1x geimpft	104	39
2x geimpft	94	37
3x geimpft	77	31
Total	293	117

Hypothese 3

Die Hypothese kann bestätigt werden. Bei der Analyse ergibt sich ein Mittelwert von M = 9.43 bei einer Gesamtpopula- tion von N = 502. Laut des *Patient Health Questionnaire-2* gilt eine Person als moderat depressiv, wenn diese ei- nen Wert zwischen 10 bis 14 erreicht.

Im nächsten Schritt wird der Kruskal-Wallis-H Test eingesetzt zum Vergleich der Gruppen hinsichtlich de- pressiver Symptome. Der Test zeigt ei- nen signifikanten Unterschied zwischen den drei Gruppen, $H(4) = 115.83$, $p < .001$. Individuen mit Long Covid (Mitt- ler Rang = 271.9) erzielen einen höhe- ren Rang als Gesunde (Mittler Rang = 169.76), aber niedriger als Individuen mit Myalgischer Enzephalomyeli- tis/chronischem Fatigue Syndrom (Mitt- ler Rang = 315.38).

Im Folgenden wird der Mann- Whitney-U Test angewendet zur Ana- lyse, welche Gruppen sich unterschei- den. Siehe Abbildung 1. Der Mann-Whit- ney-U Test zeigt, dass Individuen mit chronischem Erschöpfungssyndrom/Myalgischer Encephalitis (Mittler Rang = 252.36, Rangsumme = 46938.5) signifikant höher depressiv sind als Ge- sunde (Mittler Rang = 140, Rangsumme = 28139.5), U (n_1 = 186, n_2 = 201) = 7838.5, $p < .001$. Ebenfalls wird eine starke Effektstärke von r = .502 erzielt.

Es existiert ein signifikanter Un- terschied zwischen Gesunden (Mittler Rang = 116.27, Rangsumme = 23370.5) und Individuen mit Long Covid (Mittler Rang = 171.66, Rangsumme = 9269.5), U (n_1 = 201, n_2 = 54) = 3069.5, $p < .001$. Individuen mit Long Covid sind signifi- kant stärker depressiv als Gesunde. Es liegt eine mittlere Effektstärke von r = .308 vor.

Es existiert ein signifikanter Un- terschied zwischen Individuen mit Myalgischer Enzephalomyelitis/chronischem Fatigue Syndrom (Mittler Rang = 125.85, Rangsumme = 23408) und Individuen mit Long Covid (Mittler Rang = 102.07, Rangsumme = 5512), U (n_1 = 186, n_2 = 54) = 4027, $p < .05$. Individuen mit Myalgischer Enzephalomyelitis/chronischem Fatigue Syndrom sind leicht depressiver als Individuen mit Long Covid. Eine schwache Effektstärke von r =.143 wird erreicht.

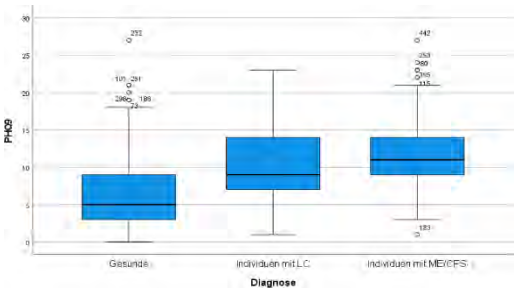


Abbildung 1: Unterschiede der 3 Gruppen im Hinblick auf depressive Symptome an- hand des *Patient Health Questionnaires-9*

Hypothese 4

Die Hypothese kann nicht bestätigt wer- den. Der Mann-Whitney-U Test zeigte, dass an Chronischem Fatigue Syndrom Erkrankte (Med = 50.0 bis 75.0, N = 186) signifikant höhere Werte bei Sympto- men erzielten als Long Covid Erkrankte (M = 12.5 bis 50.0, N = 54) (U = 2496.0 bis 4174.5, z = -5.804 bis -1.305, $p < .05$).

Gemessene Symptome, die sich als signifikant herausstellten, sind: Fati- gue, bleiernes Gefühl, Schmerzen, geis- tig müde, körperlich erschöpft, körper- lich erschöpft nach Aktivität, Muskel- schmerzen, Kopfschmerzen, Muskelschwäche, Aufmerksamkeit, Konzentra- tion und Herzrasen.

Daraufhin wurde die Effektstärke einzelner Symptome errechnet, die von schwach (r = .13) bis mittel (r = .37) reichte. Durch den Vergleich der Mittel- werte beider Gruppen zeigte sich, dass alle genannten Symptome signifikant stärker bei Individuen mit chronischem Fatigue

Syndrom ausgeprägt sind.

Allerdings kann die Hypothese nicht bestätigt werden, da die Symptom- profile der beiden Krankheiten gleich sind und sich lediglich in Intensität und Häufigkeit unterscheiden.

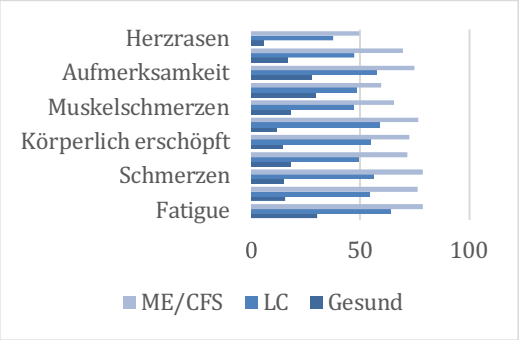


Abbildung 2: Darstellung der Unterschiede der 3 Gruppen im Hinblick auf relevante Symptome

Diskussion

Hypothesen

Die erste Hypothese befasst sich damit, dass eine Impfung vor einer Covid-19- Infektion das Risiko an Long Covid zu erkranken reduziert. Die Ergebnisse führten dazu, dass die Hypothese ver- worfen werden musste, da keine signifi- kanten Unterschiede bei der Entstehung von Long Covid zwischen einer oder kei- ner Impfung gefunden wurden. Dies könnte auf die Tatsache zurückzuführen sein, dass ein einziger Impfstoff nicht ausreicht, um Schutz vor Long Covid zu gewährleisten. Menschen galten dies- bezüglich in den meisten Fällen erst als geimpft, wenn sie zwei Impfungen erhal- ten haben. Ausschließlich Personen, die eine Johnson & Johnson Impfung erhal- ten hatten, galten bereits nach einer Impfung als geschützt. Darüber hinaus reichen mittlerweile ausschließlich zwei Impfdosen nicht länger aus, um als voll- ständig geimpft zu gelten.

Die zweite Hypothese, dass eine vollständige Immunisierung vor der Co- vid-19 Infektion besser vor einer Entste- hung von Long Covid schützt als keine Impfung, musste verworfen werden. Dies könnte auf die Ungewissheit über Veränderungen des Gesundheitszu- standes der Teilnehmenden zwischen den Impfungen zurückzuführen sein. Hierbei muss beachtet werden, dass zwischen den Impfungen sehr viel pas- siert sein kann und dies sehr schwer zu erheben ist. In der Zwischenzeit könnten die Versuchspersonen Risikopatienten geworden sein sowie möglicherweise mehrere Erkrankungen durchlebt haben, die neben den Impfungen einen Einfluss auf unsere Hypothese haben könnten.

Die dritte Hypothese konnte be- stätigt werden. Es konnte festgestellt werden, dass Personen, die an Myalgi- scher Enzephalomyelitis/chronischem Fatigue Syndrom und Long Covid leiden, depressiver sind als gesunde Menschen. Dies könnte auf die von den Erkrankun- gen verursachten Einschränkungen und Symptome zurückzuführen sein.

Zusätzlich konnte herausgefun- den werden, dass Personen mit Myalgi- scher Enzephalomyelitis/chronischem Fatigue Syndrom etwas höhere Werte auf der Depressionsskala aufweisen als Menschen mit Long Covid. Dies könnte durch die Schwere und die Häufigkeit ihrer Symptome verursacht werden. Dar- über hinaus muss beachtet werden, dass es sich bei Long Covid um eine noch nicht besonders lange existierende Erkrankung handelt. Es besteht diesbe- züglich die Möglichkeit, dass die Dauer der Erkrankung von Menschen, die un- ter Myalgischer Enzephalomyelitis/chro- nischem Fatigue Syndrom leiden, über die der Long Covid Erkrankten hinaus- geht.

Die vierte Hypothese musste ebenfalls verworfen werden. Die von uns erwarteten unterschiedlichen Symp- tomprofile in Bezug auf Menschen mit Myalgischer Enzephalomyelitis/chroni- schem Fatigue Syndrom und Long Co- vid sind nicht aufgetreten. Obwohl die Symptome von jeweils beiden Gruppen erlebt werden, berichten Menschen mit Myalgischer Enzephalomyelitis/chroni- schem Fatigue Syndrom, dass sie diese häufiger und intensiver erleben. Dieser Befund könnte mit den bereits genann- ten Einschränkungen einhergehen. Auch andere Studien haben bereits ge- zeigt, dass die Symptomprofile beider Erkrankungen vergleichbar sind, aber sich in der Dauer unterscheiden. (Po- enaru et al., 2021)

Validierung der deutschen Überset- zung des Symptom Questionnaire- 2

Bezogen auf die interne Konsistenz deu- ten Cronbachs Alpha Werte auf sehr hohe Korrela- tion hin. Diese Werte zei- gen somit, dass die Items beinahe das Gleiche messen. Es ist wichtig klarzu- stellen, dass dies für das Ziel unserer Studie jedoch vorteilhaft ist.

Einschränkungen

Das Ausfüllen des Fragebogens nimmt viel Zeit der Versuchspersonen in An- spruch. Für Menschen, die unter Myalgi- scher Enzepha- lomyelitis/chronischem Fatigue Syndrom oder Long Covid lei- den, stellt dies eine Herausforderung dar. Dies könnte sie dazu veranlassen, den Fragebogen unaufmerksam und übereilt auszufüllen. Dadurch besteht das Risiko, dass falsche Angaben getä- tigt werden.

Darüber hinaus beruhen die er- hobenen Daten auf Selbstauskünften. Sogar die Angabe einer ärztlich bestä- tigten Diagnose ist auf Selbstauskunft zurückzuführen. De- mentsprechend wur- den keine Diagnosen durch ärztliches Fachpersonal bestätigt.

Da die Befragten ihre Symptome der letzten sechs Monate angeben mussten, könnte dies zu einer Verzer- rung führen. Dies beruht auf der An- nahme, dass die Versuch- spersonen sich nicht genau an die Intensität und Häufigkeit der Symptome der letzten sechs Monate erinnern können und ihre Anga- ben somit verfälscht sein könnten.

Ausblick

Der Befund, dass Impfungen nicht signi- fikant vor Long Covid schützen, könnte darauf zurückzuführen sein, dass sich auf Selbstberichte verlassen wurde. In unserer Studie wurden die Angaben der Teilneh- menden nicht ärztlich bestätigt. Aus diesem Grund kann nicht sicher da- von ausgegangen werden, dass die Teil- nehmenden tatsächlich an Long Covid erkrankt sind.

Wichtig zu erwähnen ist, dass, um eine Diagnose zu erlangen, eine be- stimmte Prozedur an Untersuchungen erfolgen muss. Wenn anhaltende Symp- tome bestehen bleiben, sollte eine ärzt- liche Untersuchung vier Wochen nach der Infektion erfolgen. Hierbei geht es darum, Organfunktionsstörungen und andere Ursachen von Fatigue auszu- schließen. Bei anhaltenden Beschwer- den sollten gezielte symptomorientierte Untersuchungen unter Ein- beziehung verschiedener Fachärzte erfolgen. Hier- für gibt es eine Vielzahl diagnostischer Möglichkeiten in den verschiedenen me- dizinischen Fachbereichen. Es wird eine pneu- mologische Diagnostik zur Unter- suchung der Atemwege und der Lungen durchgeführt, intern- istische Diagnostik zum Beispiel Röntgen werden gemacht, um eventuelle Schäden des Nervensys- tems festzustellen, werden neurolo- gi- sche Untersuchungen gemacht, es wird eine Labordiagnostik also eine Blutana- lyse ge- macht und die Diagnostik des Fatigue- Syndroms wird erfasst. (ME- DIAN Unternehmensgruppe B.V. & Co. KG, 2021).

Dies könnte ein Anreiz für wei- tere Studien darstellen. In diesen könnte es verpflichtend sein, eine ärztliche Be- stätigung der Diagnose vorzulegen. Dies ist vor allem aus dem Grund inte- ressant, dass einige Studien bereits ge- zeigt haben, dass Impfungen vor Long Covid schützen können. Forschende aus Israel haben anhand von rund 3500 Teilnehmern untersucht, welchen Ein- fluss zwei oder mehr Dosen des BioN- Tech/Pfizer- Vakzins auf das spätere Long Covid-Risiko ha- ben. Der Befund zeigt, dass es trotz Immun- isierung zu Covid-19 kommen kann, jedoch die Imp- fungen vor Long Covid schützen (Kuodi, P. et al.).

Literatur

Bedree, H., Sunnquist, M. & Jason, L. A. (2019). The DePaul Symp- tom Questionnaire-2: a valida- tion study. *Fatigue: Biomedi- cine, Health & Behavior*, *7*(3), 166–179. https://doi.org/10.1080/2164184 6.2019.1653471

Brown, A. A. & Jason, L. A. (2014). Val- idating a measure of myalgic encephalomyeli- tis/chronic fa- tigue syndrome sympto- matol- ogy. *Fatigue: Biomedicine, Health & Behavior*, *2*(3), 132–152.

https://doi.org/10.1080/21641846.2014.928014

Business Insider Deutschland. (2022, 27. Januar). *Studie aus Israel: Impfung gegen Corona schützt vor Long Covid*. Business In- sider. https://www.busi- nessinsider.de/wissenschaft/ge- sundheit/studie-aus-israel- impfen-schuetzt-vor-long-covid/

Coronavirus disease (COVID-19) pan- demic. (2023, 10. Januar). https://www.who.int/europe/emergencies/situa- tions/covid-19

Coronavirus: Welche Langzeitfolgen gibt es? Wer ist betroffen | Al- tea. (o. D.). https://www.altea- network.com/long-covid/was-ist- long-covid/

COVID.gov. (o. D.). *COVID.gov/long- covid - Virus that causes COVID-19 can experience long- term effects from their infection*. https://www.covid.gov/long- covid/definitions

Deutsche Gesellschaft für ME/CFS. (2022, 7. Mai). *Daten & Fakten*. https://www.mecfs.de/daten-fak- ten/

Documents for Petition Drive for Cana- dian Consensus document for ME/CFS. (o. D.). https://www.cfids-me.org/con- sensusdefine.html

Erste Personen in Luxemburg gegen COVID-19 geimpft. (2020, 28. Dezember). gouvernement.lu. https://gouvernement.lu/de/actualites/toutes_actualites/communiques/2020/12-decembre/28- vaccin-covid.html

Erstmals mehrere Formen von Long COVID identifiziert. (2022, 13. September). Monitor Versor- gungsforschung. https://www.monitor-versor- gungsforschung.de/news/erst- mals-mehrere-formen-von-long- covid-identifiziert/?cookie-state- change=1673606329123

Fernández-de-las-Peñas, C. (2021). Long COVID: current definition. *Infection*, *50*(1), 285–286. https://doi.org/10.1007/s15010-021-01696-5

Fischer, A., Zhang, L., Elbéji, A., Wil- mes, P., Oustric, P., Staub, T., Nazarov, P. V., Ollert, M. & Fagherazzi, G. (2022). Long COVID Symptomatology After 12 Months and Its Impact on Quality of Life According to Ini- tial Coronavirus Disease 2019 Disease Severity. *Open Forum Infectious Diseases*, *9*(8). https://doi.org/10.1093/ofid/ofac 397

Froehlich, L., Hattesohl, D. B. R., Ja- son, L. A., Scheibenbogen, C., Behrends, U. & Thoma, M. (2021). Medical Care Situation of People with Myalgic Enceph- alomyelitis/Chronic Fatigue Syn- drome in Germany. *Medicina*, *57*(7), 646. https://doi.org/10.3390/medicina57070646

Hillebrandt, T. (2022, 27. Januar). *Vor zwei Jahren: Erster Covid-19- Fall in Deutschland bestätigt*. swr.online. https://www.swr.de/wissen/zwei- terjahrestag-corona-in-deutsch- land-100.html

Jesús García de Yébenes Prous, M., Salvanés, F. R. & Ortells, L. C. (2009). Validation of question- naires. *Reumatología Clínica (English Edition)*, *5*(4), 171–177. https://doi.org/10.1016/s2173-5743(09)70115-7

Kroenke, K., Spitzer, R. L. & Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Med- icine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kuodi, P., Gorelik, Y., Zayyad, H., Wert- heim, O., Wiegler, K. B., Abu Jabal, K., Dror, A. A., Nazzal, S., Glikman, D. & Edelstein, M. (2022). Association between BNT162b2 vaccination and re- ported incidence of post-COVID-19 symptoms: cross- sectional study 2020-21, Israel. *npj Vaccines*, *7*(1). https://doi.org/10.1038/s41541- 022-00526-5

Leonard A, J. (2015). Factor Analysis of the DePaul Symptom Question- naire: Identifying Core Domains. *Journal of Neurology and Neu- robiology*, *1*(4). https://doi.org/10.16966/2379- 7150.114

MEDIAN Kliniken. (o. D.). *MEDIAN Kli- niken*. https://www.median-klini- ken.de/. https://www.median-kli- niken.de/long-covid/de/

Meyer, Z. R. M. (2023, 12. Januar). *Wie viele bisher gegen Corona ge- impft wurden*. Corona-Impfun- gen: So viele wurden bisher ge- impft - ZDFheute. https://www.zdf.de/nachrichten/politik/corona-impfung-da- ten-100.html

Morfeld, M., Kirchberger, I., & Bullinger, M. (2011). SF-36. Fragebogen zum Gesundheitszustand. Ho- grefe.

no-author@tageblatt.lu. (2020, 7. Au- gust). *Chronologie So verlief die Corona-Pandemie bisher in Lu xemburg - Tageblatt.lu*. Tage- blatt. https://www.tage- blatt.lu/headlines/so-verlief-die- corona-pandemie-bisher-in-luxemburg/

Poenaru, S., Abdallah, S. J., Corrales- Medina, V. & Cowan, J. (2021). COVID-19 and post-infectious myalgic encephalomyelitis/chronic fatigue syndrome: a narrative review. *Therapeutic Advances in Infectious Disease*, *8*, 204993612110093. https://doi.org/10.1177/2049936 1211009385

Post COVID-19 condition (Long COVID). (2022, 7. Dezember). https://www.who.int/europe/news-room/fact-sheets/i- tem/post-covid-19-condition

Post-COVID Conditions. (2022, 16. Dezember). Centers for Disease Control and Prevention. https://www.cdc.gov/corona- virus/2019-ncov/long-term-effects/index.html

Raveendran, A., Jayadevan, R. & Sashidharan, S. (2021). Long COVID: An overview. *Diabetes & Metabolic Syndrome: Clinical Research & Re- views*, *15*(3), 869–875. https://doi.org/10.1016/j.dsx.2021.04.007

Renz-Polster, H., Tremblay, M. E., Bienzle, D. & Fischer, J. E. (2022). The Pathobiology of My- algic Encephalomyelitis/Chronic Fatigue Syndrome: The Case for Neuroglial Failure. *Frontiers in Cellular Neuroscience*, *16*. https://doi.org/10.3389/fncel.2022.888232

SF-36 - Fragebogen zum Gesundheits- zustand | Testzentrale. (o. D.). https://www.testzent- rale.de/shop/fragebogen-zum- gesundheits-zustand.html

Teska, A. (2021, 27. Dezember). *Ein Jahr nach der ersten Impfung - eine Corona-Chronologie*. Nach- richten - WDR. https://www1.wdr.de/nachrichten/themen/coronavirus/ein-jahr- impfungen-nrw-chronologie-100.html

Therapie von Long Covid - «Kann Long Covid auch bei Geimpften auf- treten?». (2022, 21. November). Schweizer Radio und Fernse- hen (SRF). https://www.srf.ch/wissen/wis- senschats/therapie-von-long- covid-kann-long-covid-auch-bei- geimpften-auftreten

Tsang, S., Royse, C. F., & Terkawi, A.S. (2017). Guidelines for develo- ping, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi jour- nal of anaesthesia*, *11*(Suppl 1), S80–S89. https://doi.org/10.4103/sja.SJA_ 203_17

Gaming for research: Happy or sad? An experimental study on emotions & video games

Tascha Fairon, Kevin Hargiono, Lisa Maksimov, Elie Paul, Laura Peters; Saskia Schablowski

Supervisor: Elisabeth Holl, Gary Wagener

Video games are not only a tool of entertainment and pastime, but also a tool for therapeutic approaches and psychological utilities such as mood management and eudaimonic purposes. For this study, Gris, a video game based on Elizabeth Kübler-Ross's model of the five Stages of Grief, has been chosen to identify possible mood alterations and relaxation effects. Furthermore, the study explores the significance of prior depressive symptoms and their gravity on this alteration. During a laboratory study, self-reported data was gathered before and after the gameplay and depressive symptoms were priorly measured with the DASS-21 (Nilges & Essau, 2021). As a psychophysiological recording, the heart rate variability (HRV) was measured during three different intervals. Results demonstrate no effect on the relation between mood affect and alteration of depressive symptoms as participants (N=64) did not report any subjective change in their emotional well-being. Results of the physiological measurements, however, did show a significant change in the HRV. The consistent increase of the RMSSD and the decrease of the LF/HF ratio during gameplay confirmed the (physiological) relaxation effect of Gris. Results are discussed in terms of the additional purpose of video games in the clinical and therapeutical domain for mood management and overall well-being.

Introduction

Video game research has experienced great popularity in the last few years, as video games have become a big part of people's everyday entertainment choices. About half of Germany's population plays video games regularly (*Gaming & digitale Inhalte im Alltag | evz.de*). In addition to this, the gender distribution is nearly equal between men and women and the average age of all players is about 37 years (*Computerspieler - Anzahl in Deutschland bis 2020*, 2020).

As already shown by Reinecke (2017) and Rieger et al. (2014) video games can be used for mood management and eudaimonic purposes, as in feelings of meaningfulness and self-actualization or -reflection. Eudaimonic video games can evoke mixed emotions and meaningful affects (Oliver et al., 2012). Furthermore, video games can be seen as a tool for recovering from negative emotional states and for interacting with players' emotions and mood (Villani et al., 2018). Video games in therapeutical treatments have already shown great success in reducing symptoms of depression and anxiety (Kowal et al., 2021) as gameplay can lead to change in cognitive behavior and promote enjoyment, flow states, and motivation in patients (Colder Carras et al.,2018; Pine et al., 2020). In order to contribute new knowledge to this field of research, this study concentrates on the relationship between playing (meaningful) video games and mood, and how this is moderated by depressive symptoms.

Theoretical background

Video games are mostly seen as tools to maximize fun and pass time, reducing them to hedonistic uses. However, a considerable number of studies have shown that video games can be used in many other contexts, for ex-

ample in mood management and eudaimonic purposes (Reinecke, 2017; Rieger et al., 2014). Mood management theory states that video game consumers choose the type of game according to their current affective state to optimize their mood.
Regarding the anhedonia experienced by people with depression, recent evidence has shown the usefulness of video games in arousing positive emotions such as joy and happiness, appreciation, competence and connections between individuals. (Kowal et al., 2021).
Additionally, research on role-playing video games has shown that these games can be valid therapeutic tools. Through self-identification with the played character, patients could challenge and change ingrained patterns of thinking and find positive alternatives. When it comes to cognitive behavioral outcomes, researchers have shown that games can promote goal achievement and re-evaluate and facilitate effective emotional flexibility and regulation as a coping tool (Kowal et al., 2021).
Research on media-induced recovery, for example, suggests that in addition to mood repair, media exposure may also facilitate the restoration of other recourses, such as general vitality and cognitive performance after phases of stress (Reinecke, 2017).
Additionally, prior research on mood repair via media found out that computer games might be able to lead to a more active, vital state, especially under the condition of a negative state such as depression (Rieger et al., 2015), which leads us to believe that media exposure, in this case the meaningful video game GRIS, might be able to mitigate and/ or alleviate depressive symptoms or improve an already pleasant mood.

As for the game GRIS itself, it depicts Elizabeth Kübler-Ross' 5-stage model of grief represented in a self-healing journey through a role-playing game story. (Sandra & Mutiaz, 2022).
Thus, the game covering a serious topic offers non-hedonic gratification, such as feelings of meaningfulness and/or contemplation.

Playing video games can have a positive effect on gamers lives and lead to learning processes that can be divided into 3 categories. Manipulation and control experience. Identity formation and identification with the character. Feedback and reward through one's own actions (see Villani et al., 2018). Individuals who regularly play video games, and thus may be more proficient in the above, may experience a greater mood alteration through playing, than individuals with less gaming experience, as proficient gamers usually already have positive connotations with video games and their possible benefits on one's well-being. However, some findings indicated that the number of hours spent playing games is not related to well-being (Vuorre et al., o. J., 2022). On the other hand, some findings claim that no video game use at all can have poor well-being as an outcome, moderate gaming a positive effect, and excessive gaming a negative one (Jones et al., 2014).
Due to the mixed findings, the influence of hours spent playing video games on well-being needs further research.

In consideration of the findings of previous research, this study was designed to test if the video game (Gris[1]) is capable of mood alteration in consideration of prior depressive symptoms and how they influence mood alteration. In addition, it is tested if playing Gris has a relaxation effect on its players. In detail, the following Hypothesis were tested:

H1: Playing Gris improves positive mood and decreases negative mood from T1 (before playing) to T2 (after playing).

H1a: Moderation - depressive symptoms:
The increase in positive mood and decrease in negative mood are more pronounced in individuals with elevated DASS-21.

H2: Playing Gris has a relaxation effect:

[1] developed by Nomada Studios and published by Devolver Digital. Trailer: https://youtu.be/BRiKQIVo7ao

The slope of heart rate variability from T1 (baseline: before playing) to T2 (during playing) and to T3 (after playing).

H3a: Subjects that spend on average more hours playing video games per week experience greater alteration in positive mood affects as a result of playing Gris than people who spend less hours playing video games in a week.
H3b: Subjects that spend on average more hours playing video games per week experience greater alteration in negative mood affects as a result of playing Gris, than people who spend less hours playing video games.

These hypotheses were tested in a lab experiment in which participants provided information about depressive symptoms and played the meaningful game Gris for 20 minutes. Before and after gameplay, their positive and negative affect were assessed as well as their heart rate variability.

Methods

Sample

An a priori power analysis was conducted using G*Power (Version 3.1.9.7) (Faul et al., 2007) for sample size estimation. Results indicated a minimum sample size of N = 43 to achieve 95% power with a significance criterium of α = .05. Recruitment took place via flyers in the buildings and the area around campus of the University of Luxemburg. Sixty-four participants, aged between 17 and 33 (N = 64, M = 22.56, SD = 3.35) took part in the study. Fifteen participants were male (23.40%), forty-eight were female (75%) and one participant was non-binary (1.60%). Furthermore, data was collected about the gaming experience of our participants, by asking them about their usual gaming hours per week. There was a range from zero to fifty-six hours per week (M = 5.17; SD = 9.70). About a third of the participants stated that they do not play video games at all (21 participants, 32.8%).

Procedure

The experiment took place at the University of Luxemburg and began with participants being provided with informed consent. Afterwards, three reusable electrodes were placed on the participant's upper body according to the Eindhoven placements. Next, participants filled out the first part of the questionnaire, which contained the Positive and Negative Affect Schedule (PANAS) (Breyer & Bluemke, 2016), the Difficulties in Emotion Regulation Scale (DERS) (Kaufman et al., 2016), the Depression, Anxiety and Stress Scale – 21 Items (DASS21) (Nilges & Essau, 2021), and the Motivation to Play Scale (MOPS) (Holl et al., 2022). After completing the questionnaire, participants received the gaming instructions and game controls and then played *Gris* on the *Nintendo Switch* for 20 minutes. During the game participants were left alone in the room with dimmed lights and medium-level volume. This was done in order to create a suitable environment for participants to have the most immersive gaming experience possible. After 20 minutes of gameplay, participants filled out the second part of the questionnaire, which contained the PANAS (Breyer & Bluemke, 2016) again. In addittion, a self-constructed scale to measure how the gaming experience was perceived and if there have been any difficulties while playing, consisting of 19 items, was used. Afterwards, participants were allowed to take off the electrodes and were debriefed and renumerated for their participation with 10€ and study participation credits.

Material

Depressive symptoms -DASS

The Depression-Anxiety-Stress-Scale (DASS-21) (Nilges & Essau, 2021) is an instrument to

measure the different symptoms of depression, anxiety and stress while trying to separate them as accurately as possible. By creating items that are clearly attributable to depression but not to anxiety and vice versa and those where the classification was not clear. The questions refer to the week prior to taking the test and can be answered on a likert scale from 0-3 with the following response possibilities:
From 0 = Did not apply to me at all to
3 = Applied to me very much or most of the time.
A possible Item for the subscale 'depression' is:
Item 3: "I couldn´t experience positive feelings".
The internal consistency of the scales is Cronbach's α = .88 for the depression scale, α = .76 for the anxiety scale, and α = .86 for stress.

Emotion regulation - DERS

The DERS (Difficulties in Emotion Regulation Scale) (Kaufman et al., 2016) is a self-report instrument to measure emotion regulation problems, which consists of 36 items and six subscales: Nonacceptance, difficulties engaging in goal-directed behavior, impulse control difficulties, lack of emotional awareness, limited access to emotion regulation strategies and lack of emotional clarity. A possible item is: Item 2: "I pay attention to how I feel". The questions are rated on a 5-point Likert-scale ranging from 1 "almost always", to 5 "almost never". The internal reliability for the current sample was .95. The DERS (Kaufman et al., 2016) was used a control measure.

Positive and negative affect – PANAS

The PANAS is a 20-item-mood scale, measuring positive and negative affect with 10 items each (Breyer & Bluemke, 2016). Participants are asked to respond to each item on a scale from one to five, with 1 = very slightly and 5 = very much. The answers refer to how often the person has felt the mood matching item. In this study it is used to assess the mood of participants before and after playing the game. Afterwards both measurements will be compared, to check if mood alters after due to gaming. The positive affect subscale consists of 10 items, one of them is called "inspired". The negative affect subscale also consists of 10 items, one example for negative affect would be: "distressed". Pretest negative affect Cronbachs α = .83 and posttest negative affect α = .80. Cronbachs alpha for positive affect equals: for pretest α = .81 and for posttest α = .85.

Motivation to play scale – Mops

The MOPS (Holl et al., 2022) is a 58-item scale, measuring gaming motivation with 10 subscales. These subscales are creativity/exploration, escapism, competition, prestige, enjoyment, achievement, socializing, boredom, aggression, and skill. Participants are presented with the stem "I play video games…" and have to rate items (e.g., …to explore imaginary worlds) on a 5-point scale from 1 = strongly disagree to 5 = strongly agree." Cronbach´s alpha for all items is α = .96. There have not been any Hypothesis regarding the MOPS (Holl et al., 2022), but it is used as an addition scale.

Gaming Stimulus – Gris

Gris is a single player Jump-´n´-run adventure game developed by Nomada Studios, published in 2018. It illustrates the story of a female protagonist on her journey through grief.
The participants played the first passage of the game in which the protagonist experiences the first stage of grief. In this passage the player had to do simple jump n´ run tasks and gained one of the many special abilities. The first passage was chosen, as the player gets an introduction to the game, it's theme and aesthetic. The cutscene of the first passage gives the player context of the background history and therefore promotes the immersion into the game. In addition, the beginning of the game can be seen as less demanding than the more advanced stages of the game in terms of needed skill, which leads to less frustration for inexperienced players. The game Gris was chosen for this study, as it has been designed based on Elizabeth Kübler-Ross's model of the 5 Stages of Grief: denial, anger, bargaining, depression and acceptance. This model is represented heavily by illustrative narration in GRIS (Sandra Mutiaz, 2021). Thus, the game, covering a serious topic,

does offer non-hedonic gratification, such as feelings of meaningfulness and/or contemplation and offering possible self-identification with people suffering from depressive symptoms.

Heart rate variability

Heart rate variability (HRV) is a measure of the variation in time between each heartbeat. This variation is controlled by the autonomic nervous system, which is subdivided into the sympathetic and parasympathetic nervous system. The heart is innervated by the autonomic nervous system such that a relative increase in sympathetic activity is associated with a decrease in heart rate variability and a relative increase in parasympathetic activity is associated with an increase in heart rate variability. Thus, a relative increase in sympathetic activity shortens the time between heartbeats, and a relative increase in parasympathetic activity lengthens the beat-to-beat interval (Lane et al., 2009).
In the past, HRV measures have been successfully used to identify the physiological state in mediated contexts, for example, while playing a video game (Hou et al., 2021).
A longer time interval between two heartbeats thus meant an activation of the parasympathetic nervous system and a higher degree of relaxation.
Heart rate variability was measured during three intervals in the experiment by placing three disposable electrodes on the participant's body (to the left and right sides of the collarbone just below the trapezoid muscles to address movement artifact), after cleaning and gently abrading the skin surface, and then connecting these electrodes to the BIOPAC MP36. First, a baseline measure was taken before playing the game, while participants were filling out the first questionnaire. Then during the game play again and lastly a post-game measurement was taken, while the participant filled out the second part of the questionnaire. The software "Acqknowledge" (Version 5.0.6) was used to view the ECG data in real time on a computer. Afterwards the data was polished using the software "Physio Data Toolbox" (Version 0.6.3) (Sjak-Shie, E. E. ,2022).

Results

Change in mood

A within-subjects ANOVA was calculated to see if playing Gris increased positive affect and decreased negative affect from before to after gameplay. There was no significant change for positive affect (pre = 3.10 and post = 3.06), $F = .12$, $p = .73$, $\eta p^2 = .002$ and for the negative scales (*pre:* 1.40 *and post:* 1.40*). F* = .001, $p = .98$, $\eta p^2 = .00$.
. Mean in mood is displayed in figure 1.
A Pearson correlation was conducted to establish the linear relationship between the variance in mood (PANAS) (Breyer & Bluemke, 2016) and depressive symptoms (DASS-21) (Nilges & Essau, 2021). Results showed no significant correlation between the two variables (positive affect: $p = .76$, $r = -.04$; negative affect: $p = .86$, $r = .02$)
So, the hypothesis H1 had to be refuted.

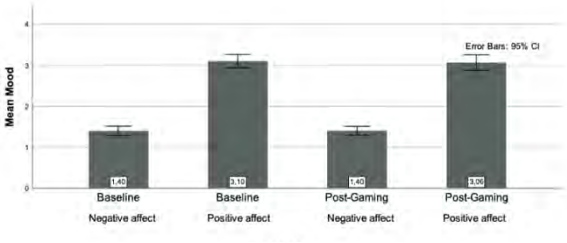


Figure 1: Mean Mood Results

Heart rate variability

Whether 'Gris' has a relaxation effect on the subjects (H2) is illustrated by two graphs, each of which measures the mean of the participants.

First, the root mean square of successive differences between normal heartbeats in millisecond (RMSSD) was analysed. Results of RMSSD are displayed in figure 2. A within-subjects ANOVA indicates that the RMSDD is

lowest before the game (T1) (36.29 ms) and increases during the game (43.44 ms) and after the game (45.10 ms). F(2,122) = 7.17, p = .001 η^2 = .11

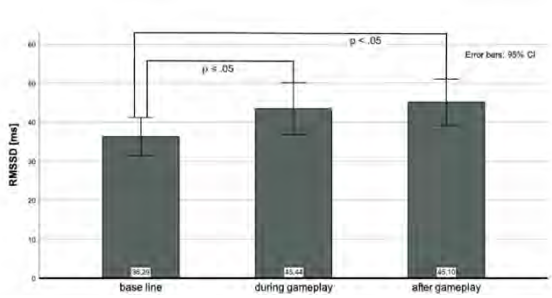


Figure 2: RMSSD results

In figure 3 the LF/HF ratio is displayed. The results also support the hypothesis that the video game 'Gris' has a relaxation effect. The LF/HF ratio was analysed during T1, T2, and T3. Results of a within-subjects ANOVA showed a high LF/HF ratio indicates a dominant activity of the sympathetic nervous system. (p <. 05) F(1.75,106.44) = 3.45 p = .04, η^2 = .06
It can be seen on the graph that the sympathetic nervous system is higher before (*T1* = 2.72) and after the game (*T3* = 2.77) than during the game (*T2* = 2.11)

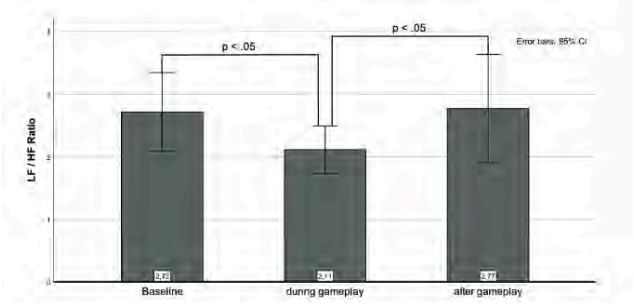


Figure 3: LF/HF ratio results

The results on the third hypothesis show that there are no significant results between gaming-hours and change in positive affect.
The null hypothesis is confirmed, so H3a is refuted.

The results of the ANOVA and linear regression for the hours of gaming per week and change in negative affects are not significant, F = 2.88, p = .10.
However, a significant result on the Pearson correlation (p = .05) has be found. Time spent playing and change in negative affects correlate (r = -.21, p < .001), which means the more participants played video games in their free time, the smaller their change in negative affects were.
The null hypothesis is confirmed, so H3b is also refuted

Discussion

The results on mood indicate that participants subjectively felt almost no mood-alteration after playing ‘GRIS’.

However, results based of the heart rate variability indicate that the parasympathetic nervous system becomes increasingly active and remains so after having played the game. (Lane et al., 2009). Since the increasing activity of the parasympathetic nervous system is related to the slowing of the heart rate, this graph speaks for the relaxation effect of the game on the subject

Calculated values in the third figure correspond to a relaxation effect during the game, as the LF/HF ratio decreases during the gameplay. The increase in sympathetic activity T3 is explained by the fact that the subjects are pulled out of the game after 20 minutes and are again given a questionnaire to answer.

The aims of this study were to identify meaningful video games’ possible capability of mood alteration in regards of prior depressive symptoms and its relaxation effect.
The results obtained from the PANAS (Breyer & Bluemke, 2016) questionnaire did not support the hypothesis that playing Gris has an impact on mood alteration from T1 to T2. Subjectively, there was no effect on the well-being of our participants.
However, the physiological data collected by measuring the heart rate variability supports the hypothesis that there is a relaxation effect

after playing the game. The phenomenon that physiological well-being and subjective mood do not align has already been shown in past studies (Sommerfeldt et al., 2019).
The fact that participants did not indicate a change in mood subjectively (PANAS) (Breyer & Bluemke, 2016), but physiologically a relaxation effect was observed, is relevant for future studies, as the number of stress-related diseases rises and there is a high demand for low cost and effective interventions, such as meaningful video games. (Twenge et al., 2019)
The present findings indicate, that even though subjects did not perceive a change in mood subjectively, physiologically, participants felt more relaxed after playing the meaningful video game. This adds to prior findings that video games could have positive therapeutic effects on people.

Limitations

A major limitation of our study is the lack of variance in our participants.
More healthy individuals generally enrolled, so the rate of people with depressive and anxious symptoms was very low. This limitation affected the study because it is difficult to determine whether the game Gris specifically has a positive effect on people with depressive symptoms. Consequently, our study suffered from a so-called 'sampling or convenience bias', as most of the subjects were university students.
There are also limitations caused by the game itself, as 'Gris' is a unique game and therefore does not represent all meaningful video games. There is also the possibility that the male and diverse participants have difficulty identifying with the female protagonist.
Furthermore, it could be argued that game experience affected the 20 minutes of playtime. Even though all participants played the same passage of Gris and played for exactly 20 minutes, progress varied greatly among participants. Future studies should, therefore, perhaps include game progress as a control variable.
Important to recognize is also that the physiological relaxation might not be the result of the video game *Gris* in itself, but instead stems from the participant being at ease in the environment, playing.

Conclusions

What emerges from this study is that meaningful video games such as *Gris* do influence mood alteration and present a relaxation effect, as the activity of both the sympathetic and parasympathetic regions increased and a psychophysiological change has been noticed. This observation can be used by future researchers and members in the therapeutical field with the focus on video games' contribution to the achievement of an overall state of well-being and in treatments regarding mental health problems, like depressive symptoms. Even though the relaxation effect has been proven on a physiological level, no subjective change in participants' mood were indicated during the study and no mood alteration was observed in individuals with elevated depressive symptoms. Furthermore, participants with higher hours of playing video games in their free time, represented less mood alterations on the negative scale. Future studies could, therefore, elaborate on the different findings on the subjective and psychophysiological level. A follow-up study regarding the cause of relaxation might bring more precise insights on the effects of meaningful video games as an additional control group would not be playing *Gris* but participate in a different method of relaxation (e.g., playing a puzzle).

References

Breyer, B., & Bluemke, M. (2016). Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. https://doi.org/10.6102/ZIS242

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 scienc-

es. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Holl, E., Sischka, P., & Melzer, A. (2022). *Development and Validation of the Motivation to Play Scale (MOPS)*. 1–7.

Hou, C.-J., Chen, Y.-T., Capilayan, M., Lin, Y.-S., Huang, M.-W., & Huang, J.-J. (2021). Analysis of Heart Rate Variability in Response to Serious Games in Elderly People. *Sensors*, *21*(19), Art. 19. https://doi.org/10.3390/s21196549

Jones, C., Scholes, L., Johnson, D., Katsikitis, M., & Carras, M. (2014). Gaming well: Links between videogames and flourishing mental health. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00260

Kaufman, E. A., Xia, M., Fosco, G., Yaptangco, M., Skidmore, C. R., & Crowell, S. E. (2016). The difficulties in emotion regulation scale short form (DERS-SF): Validation and replication in adolescent and adult samples. *Journal of Psychopathology and Behavioral Assessment*, *38*(3), 443–455. https://doi.org/10.1007/s10862-015-9529-3

Kowal, M., Conroy, E., Ramsbottom, N., Smithies, T., Toth, A., & Campbell, M. (2021). Gaming your mental health: A narrative review on mitigating symptoms of depression and anxiety using commercial video games. *JMIR Serious Games*, *9*(2), e26575. https://doi.org/10.2196/26575

Lane, R., Mcrae, K., Reiman, E., Chen, K., Ahern, G., & Thayer, J. (2009). Neural correlates of heart rate variability during emotion. *NeuroImage*, *44*(1), 213–222. https://doi.org/10.1016/j.neuroimage.2008.07.056

Nilges, P., & Essau, C. (2021). *DASS. Depressions-Angst-Stress-Skalen—Deutschsprachige Kurzfassung*. https://doi.org/10.23668/PSYCHARCHIVES.4579

Reinecke, L. (2017). Mood management theory. In P. Rössler (Hrsg.), *The international encyclopedia of media effects* (S. 1–13). Wiley-Blackwell.

Rieger, D., Frischlich, L., Wulf, T., Bente, G., & Kneer, J. (2015). Eating ghosts: The underlying mechanisms of mood repair via interactive and noninteractive media. *Psychology of Popular Media Culture*, *4*(2), 138–154. https://doi.org/10.1037/ppm0000018

Rieger, D., Wulf, T., Kneer, J., Frischlich, L., & Bente, G. (2014). The winner takes it all: The effect of in-game success and need satisfaction on mood repair and enjoyment. *Computers in Human Behavior*, *39*, 281–286. https://doi.org/10.1016/j.chb.2014.07.037

Sandra, N. P. A. A., & Mutiaz, I. R. (2022). *The Embodiment of Kübler-Ross Model through Game Elements in GRIS and Its Impact on Player Experience:* ICON ARCCADE 2021: The 2nd International Conference on Art, Craft, Culture and Design (ICON-ARCCADE 2021), Bandung, Indonesia. https://doi.org/10.2991/assehr.k.211228.033

Sommerfeldt, S. L., Schaefer, S. M., Brauer, M., Ryff, C. D., & Davidson, R. J. (2019). Individual Differences in the Association Between Subjective Stress and Heart Rate Are Related to Psychological and Physical Well-Being. *Psychological Science*, *30*(7), 1016–1029. https://doi.org/10.1177/0956797619849555

Twenge, J. M., Cooper, A. B., Joiner, T. E., Duffy, M. E., & Binau, S. G. (2019). Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005–2017. *Journal of Abnormal Psychology*, *128*, 185–199. https://doi.org/10.1037/abn0000410

Villani, D., Carissoli, C., Triberti, S., Marchetti, A., Gilli, G., & Riva, G. (2018). Videogames for Emotion Regulation: A Systematic Review. *Games for Health Journal*, *7*(2), 85–99. https://doi.org/10.1089/g4h.2017.0108

Vuorre, M., Johannes, N., Magnusson, K., & Przybylski, A. K. (o. J.) (2022). Time spent playing video games is unlikely to impact well-being. *Royal Society Open Science*, *9*(7), 220411. https://doi.org/10.1098/rsos.2204

Too Anxious to Think

Thalia AZMEH, Leonie BACKES, Liz EIDEN, Sophie SCHÄDLER, Michaela SCHERSCHEL, Lily TANDEL and Lara WEYDERT

Supervision: Dr. Mila MARINOVA

The aim of this study is to investigate to what extend language profile (German /French) and Cognitive Test Anxiety (CTA) affect test performance on a general knowledge test for Luxembourgish students. Therefore, participants were tested on two different ses-sions (once in German and once in French). For the collection of the data, they com-pleted self-assessment questionnaires, about the demographic background, language level-information, State and Trait Anxiety level and the cognitive test anxiety. Further-more, they had to complete a computerized task, to test their general performance in an exam-like situation.

Participants (N = 53) were aged between 19 and 34 years old, among these mostly women (71.7%). They all followed the Luxembourgish school curriculum and thus all spoke Luxembourgish, German, French and English.

The results show significant differences between language conditions for test accuracy (performance), confidence in the test accuracy and solution time of the General Knowledge test tasks. Furthermore, significant correlations were found between trait anxiety and CTA in both language conditions. In addition, a negative correlation between CTA and test performance was found in the French language condition. In the German language condition, there was found a positive correlation between trait anxiety and test performance. The language preference of the participants is presumably the reason for the differences between the two language conditions (regarding test accuracy, confidence, and solution time). In this study, the German language has been predominantly preferred, which can be explained by various factors. A negative correlation between CTA and test performance (French condition) can be attributed to the fact that anxiety becomes maladaptive above a certain level and performance is thus negatively affected.

Introduction

After compulsory education until the age of 16, many pupils continue their schooling and pursue further studies at universities. The educational context is linked to evaluation: to measure the students' abilities and knowledge, their performance in tests is evaluated (Angelidis et al., 2019). Due to this evaluation, there are many inter-individual and intra-individual comparisons. On the one hand students compare their performance to that of their peers and on the other hand they compare their performance in the different subjects. The questions that arise are what causes these inter- and intra-individual differences and whether performance is a true representation of the student's ability. Research has shown that this is not always the case as performance depends on several variables, such as the anxiety one feels for a particular subject or test, but also the language of instruction, for example (Howie, 2003; Mendez et al., 2019; Horwitz, 2001; Teimouri et al., 2019).

Research has shown that anxiety can have a negative impact on performance. Moreover, foreign language anxiety negatively influences language achievement (Horwitz, 2001; meta-analysis by Teimouri et al., 2019). The meta-analytical study by Teimouri and colleagues (2019) clarified the relationship between language anxiety and second language learning and tried to assess the possible sources of anxiety, such as teaching practices, for example. The negative influence of language anxie-

ty on achievement was not only found for students but also on different instructional levels and even for pre-service teachers. Furthermore, they showed that students with higher levels of anxiety expected and received lower grades than the less anxious colleagues. However, it is important to distinguish between the role of anxiety in language learning and its role in language performance. Unfortunately, the study found contradictory information. While some of the studies they examined showed a negative correlation between anxiety and language achievement, others found no correlation or even a positive correlation. In addition, it is difficult to determine if anxiety has interfered with the learning progress, thus influencing achievement levels or if anxious learners have difficulties displaying the competence they have attained. Finally, there is no mention of controlling for confounding variables.

In their educational career, people will not simply have to perform but will also be evaluated in many situations. As top performance is expected in such test situations, the evaluation of their performance can have significant consequences for the rest of their lives (Angelidis et al., 2019). Angelidis and colleagues (2019) examined how cognitive performance anxiety (CPA) influences the performance in tasks and especially the performance of working memory. It was found that, in general, people who experienced high levels of stress, had a significantly slower working memory performance. They explained this through the impairment of stress on the attentional control of the participants. They also found differences in the performance, depending on the level of CPA. On the one hand, the study demonstrated more reality than other studies, as the stress levels were high for a longer period of time as the participants had to perform multiple cognitive tasks, but on the other hand, not all realistic factors were taken into consideration. For example, a detailed preparation could reduce anxiety for some people. In their study, they only included women and excluded men, so this raises the question of how representative the study is for the population as a whole.

Research shows that in evaluative situations, Cognitive Performance Anxiety (CPA) can arise and reveal obstacles to academic achievement (Thomas et al., 2018). Language anxiety can occur while performing in foreign language tasks. Consequently, such anxieties could be linked to situations in which people are being asked to perform and are evaluated simultaneously (Angelidis et al., 2019; Cassady & Johnson, 2002).

Similarly, Zheng and Cheng (2018) wanted to investigate the nature and effect of anxiety. In their study, they used a context of high stakes testing for Chinese University students learning English as a foreign language. Therefore, the study examined students' perceptions of their anxiety levels including foreign language classroom anxiety and test anxiety and the relation to their test performance. They found that Cognitive Test Anxiety (CTA) is a significant predictor of language achievement. However, the interviews about the perceived anxiety of the students were conducted a few days after the testing, which might distort their results about anxiety levels.
Overall, research on CPA/CTA demonstrates a negative influence on language performance (Zheng & Cheng, 2018), working memory (Angelidis et al., 2019) and general test performance (Cassady & Johnson, 2001).

In the study by Kleemans and Segers (2020), it was examined whether basic linguistic skills determine growth in advanced mathematics. To assess this, they examined the difference between first language and second language learners in advanced math growth to investigate to what level the basic and advanced linguistic skills of first/second language learners directly/indirectly affect mathematical growth. They found lower scores for second language learners on advanced mathematics and linguistic skills. Advanced linguistic skills directly predicted the growth in geometry and fraction skills whereas basic linguistic skills predicted growth in advanced mathematics through arithmetic. They observed no difference in gender and socioeconomic status. Unfortunately, they didn't take any other variables into consideration, such as reading comprehen-

sion or other extraneous variables, for example. Furthermore, they only drew conclusions from measurements of fifth graders. Therefore, this sample is not representative of the whole population. We know from studies on mathematical achievement that those whose mother tongue is different from the language of instruction perform particularly poorly in math tasks (Greisen et al., 2021).
According to the study conducted by Greisen and colleagues (2021), there is a relation between mathematical performance and language. To study this relation, the researchers analysed existing data of Luxembourgish 3rd graders in two consecutive years, separating them in groups according to their native language (Luxembourgish, French, Portuguese, south Slavic language). They had to solve mathematical problems, of which the instructions were written in German. The Luxembourgish group performed better in both mathematics achievement and German reading comprehension. This leads to the conclusion that lower reading comprehension of mathematics instruction is indeed responsible for lower mathematics achievement in speakers of different mother tongues. However, one cannot fully depend on those findings as the only reference group were German people. Moreover, the use of existing data means that there is a lack of control over collected data. Finally, the study lacks the aspect that the motivation to learn German or Math could be a potential driving factor resulting in improved or impaired math performance.

There is a large body of research conveying that the language of instruction plays a critical role in a student ‘s performance. People whose first language is not the one in which content is taught and tasks are given perform comparatively worse than people whose mother tongue is the language of instruction. (i.e., Howie, 2003; Mendez et al., 2019).
This research is of particular importance for multicultural and multilingual educational institutions that try to integrate students with diverse language profiles (and often immigration backgrounds). Such students come from homes where the language spoken is different from the language in which the educational institution executes its lectures (Greisen et al., 2021). In this context, Luxembourg stands out as a trilingual country with a multilingual education system (Ministère de l'Education nationale et de la Formation professionnelle, n.d.). Concretely, in primary school, the languages Luxembourgish, German and French are learned, with the content of non-linguistic subjects being taught in German. In secondary school, another language (English) is added and, in addition, a change in the language of instruction takes place. From the fourth year on, most of the content, except the language lessons, is taught in French for the people following the classical secondary education. As for the people following the general secondary education, the language of instruction remains German with a few isolated exceptions depending on the chosen specialization.

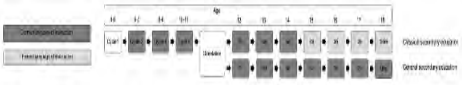


Figure 1: Luxembourgish school curriculum

Through the influence of the non-native language of instruction on academic performances, this linguistic diversity can be challenging for students (Ministère de l'Education nationale et de la Formation professionnelle, n.d.). Especially the results of the study by Greisen and colleagues (2021) and a lot of other research that has been conducted in the educational context, presented earlier, point to this problem. Still, a gap remains concerning studies about the relationship between cognitive test anxiety and test performance which take the language of instruction into account.
Therefore, this study examines the relationship between language profile, CTA and test performance for the special case of Luxembourg.
This within subject designed experiment uses a general questionnaire in English to assess the participant’s data. Furthermore, the State Trait Anxiety Inventory (STAI) by Spielberger (1983) is conducted to evaluate the participant’s trait and state anxiety and the second edition of the Cognitive Test Anxiety Scale (CTAS-2) by Thomas et al (2018) offers insight about the anxiousness level after the tests. Finally, the test accuracy of the cognitive

computer test is utilized to measure performance.

The hypotheses of the study were based on the following:
When one feels anxiety, the body gets aroused, and it is subsequently assumed that this arousal interferes with the test performance. Therefore, in our first hypothesis, we assume that the higher the CTA for the German language of instruction, the lower the performance in the tasks will be. Similarly, the participants with a higher CTA level for the French language of instruction would also perform worse. In other words: If one is not anxious of German/French instructional language on a test, performance will not suffer. Since German is not only the first language taught in Luxembourgish schools, but also the most common language of instruction, it is more present than French during the academic career.
For this reason, it is assumed that individuals will perform better in the German condition than in the French condition. Moreover, it is presupposed that participants perform better in the test in the preferred language of instruction. As a result, the participants who prefer French would achieve higher accuracy levels in the tests where the language of instruction is also French. In contrast, the participants with a preference for German would achieve higher accuracy levels in the tests with the German instructional language.

H1. There is a relationship between cognitive test anxiety and performance.
H1a. The higher the CTA in the German language of instruction, the lower the test accuracy.
H1b. The higher the CTA in the French language of instruction, the lower the test accuracy.
H2. The language of instruction influences the test accuracy.

Methods

Sample Description

Altogether $N = 57$ participants were tested in the present study, but only the ones with a whole data set were included. This gave us a sample of $N = 53$. The recruitment of participants took place via the university's Moodle forum, social media and WhatsApp groups. In addition, some participants were not recruited online but in person.
Participants were aged between 19 and 34 years old, the average age was $M = 21.96$ ($SD = 3.36$) years. As far as sex distribution is concerned, 71.7% were female and 26.4% were male. One participant gave no information about their sex.

The following inclusion criteria had to be fulfilled for participation: participants must have normal to corrected-to-normal vision and must have followed the Luxembourgish school curriculum as well as speaking German, French and English. 62.3% claimed to be fully proficient in and 18.9% to be native speakers of German. For French, 56.6% claimed to be advanced, 15.1% to be fully proficient and 17% to be native. As for English, 47.2% claimed to be advanced and 35.8% to be fully proficient. Participants with learning difficulties (e.g., dyslexia or dyscalculia), mood disorders (e.g., anxiety or depression) or attention disorders (e.g., ADD or ADHD) were excluded. In our study, there was no person who met these criteria, so no one had to be excluded because of this.

As a compensation, participants received two 10€ vouchers and two course credit hours. The ethical approval was obtained for the University's Ethical Committee. Informed consent was sought from all participants prior to the study.

Sample characteristics are summarized in Table 1.

Table 1: Average Age, Language Proficiency Levels and Preferred Exam Language

	Age	Proficiency German	Proficiency French	Percent
Mean	21.96	3.98	3.37	
Standard Deviation	2.36	0.66	0.90	
Preferred Exam Language				
German				67.92%
French				15.09%

Measures and Instruments

The participants completed self-assessment questionnaires and a computerised task, described in the sections below.

General Questionnaire

Demographic information and language level information were collected using a general questionnaire created by the researchers.

Information such as age, languages spoken at home ('Which language(s) do you speak with your mother?' and 'Which language(s)do you speak with your father?'), additional languages ('If you speak any additional language, please specify below') and 'Preferred language' ('In which language(s) do you prefer to speak?') was recorded.

In addition, the participants were asked to self-assess their language level on a scale from 1 'beginner' to 4 'native speaker'. The questionnaire also included questions, checking whether the participants met the inclusion or exclusion criteria mentioned above.

State-Trait Anxiety Inventory (STAI)

In order to control whether cognitive test anxiety was associated with general anxiety or trait anxiety, we also administered the State-Trait Anxiety Inventory (STAI) by Spielberger et al. (1983) to measure participants' state (S) and trait (T) anxiety. Moreover, we wanted to ensure that the state anxiety does not differ between the two testing sessions.

The State Anxiety measures the anxiety level at the moment and the Trait Anxiety measures the general anxiety level of a person. The psychological Inventory consists of 40 self-reported items on two scales, 20 each of State-Anxiety and Trait Anxiety. The participants had to complete the items on a 4-point Likert-scale from 1 ('not at all' for S-anxiety and 'almost never' for T-anxiety) to 4 ('very much so' for S-anxiety and 'almost always' for T-anxiety) (Spielberger et al., 1983). The STAI was completed by the participants in English, the original language. To score the STAI, the average total score was used, so the higher the scores on the scale level, the higher the levels of anxiety (Spielberger et al., 1983).

General Knowledge Test

In order to stimulate an exam-like situation in a laboratory setting and to ensure a higher ecological validity (i.e., a stimulated real-life situation), we administered a timed General Knowledge Test via the programme PsychoPy (Peirce, 2007).

The participants' general performance was measured by a General Knowledge Test created by the researchers. It consists of 12 general knowledge items of varying degrees of difficulty (e.g., 'When did the Second World War take place?', 'what is the longest river in the world?', etc.). For each item, the participants had to select one of four possible answers. The General Knowledge Test was divided into two parts, so that 6 items per session were presented one after the other to the participants (either in French or German, depending on the session). The items differed in both language sessions, so they were not used twice.

In addition, we also collected information about the subjective retrospective confidence of the participants. Concretely, after each question, the participants rated how confident they were about the accuracy of their answer ('certainly wrong', 'probably wrong', 'maybe wrong', 'maybe correct', 'probably correct' or 'certainly correct'). The participants had a total of 25 minutes to complete the tasks but should do it as fast as possible and could have breaks in-between. Hence, their performance was not only measured based on the accuracy of their answers but also on their solution time.

Cognitive Test Anxiety Scale – 2nd Edition (CTAS-2)

To measure the participants' cognitive test anxiety, the 2nd edition of the Cognitive Test Anxiety Scale (CTAS-2) by Thomas et al. (2018) was used. The scale consists of 24 items and was developed to measure cognitive indicators of test anxiety at all stages of preparation and performance in the learning test cycle (Thomas et al., 2018). Participants had to complete the items using a four-point scale ranging from 1 ('not at all typical of me') to 4 ('very typical of me'). The CTAS-2 was translated into German and French by the researchers.

To score the CTAS-2, the average score was used, so higher scores on the scale indicate higher levels of cognitive test anxiety. Additionally, the order

of the items was randomised across both languages (German and French).

Study Design and Procedure

All measurements above were administered in a within-subject-design with language condition (French vs. German) as the independent variable. As dependent variables we used the average score on the anxiety scale as well as the mean accuracy and solution time on the general knowledge test.

Participants were tested in two different sessions (once in German and once in French, preferably with 3-7 days in between) and the order of the language was counterbalanced across participants (half of them first started with the German condition and half with the French condition).
The actual testing began with the participants filling out a General Questionnaire, gathering general socio-demographic and language information (in English). It was also possible to complete the questionnaire at the recruitment stage (before coming to the lab). Afterwards, the researchers administered the State-Trait Anxiety Inventory (STAI in English) to the participants.
The next step was the Computer Testing. The researchers handed the participants a laptop with the General Knowledge Test for which they had a maximum of 25 minutes to complete (in French or German, depending on the session) but they were asked to get it done as fast as possible. After each item, the participants had to rate on how confident they were about the accuracy of their answer. After the computer task was, the participants completed the CTAS-2 (in French or German, depending on the session).
In the second session, the participants only had to fill out the State Anxiety Scale (in English), complete the General Knowledge Test (in French or German, depending on which language they had used in the prior session) and finish with the CTAS-2 (in the language of the session).
In general, all questionnaires were self-administered, and the general knowledge test was self-paced as well. The duration of each session was approximately 45 minutes, and the participants received their reward (10-euro voucher) after each session and could get course credit hours if needed. Debriefing was done only at the end of the second session.
From the greeting to the computer testing, the researchers spoke English with the participants to avoid language context effects, since research has shown that language context can provide cognitive cues and facilitate or hinder further performance depending on whether it was done in a congruent or incongruent language (Van Rinsveld et al., 2017). From the start of the computer testing, the test leaders could speak English or the language of the session (German in the German session, French in the French session) to the participants. For the debriefing the language was chosen freely.

Results

Table 2: Descriptive of Measurements

Measurements	N	Mean	Standard Deviation
State Anxiety Session 1	53	2.27	0.20
State Anxiety Session 2	53	2.23	0.22
Trait Anxiety	53	2.36	0.24
CTAS German Session	53	1.97	0.49
CTAS French Session	53	1.98	0.54
Test Accuracy German Session	53	0.63	0.19
Test Accuracy French Session	53	0.41	0.21
Solution Time German Session	53	11.55	4.82
Solution Time French Session	53	14.29	5.14
Confidence in Accuracy German Session	53	3.72	0.76
Confidence in Accuracy French Session	53	3.42	0.77

Control Tests

First, as a form of control, we conducted a dependent, two-tailed *t*-test to check for differences concerning the state anxiety of the participants between both sessions. We found that rated state anxiety in the first session (M = 2.27, SD = 0.20) does **not significantly** differ from the rated state anxiety in the second session (M = 2.23, SD = 0.22), $t(52) = 1.61$, $p = 0.113$, Cohen's $d = 0.22$. This means that any potential differences in the CTA between French and German are unlikely to be due to the fact, that participants were more anxious that day.
We also conducted a correlation analysis between Trait Anxiety and State Anxiety for the complete sample and found that there is a **significant** positive relationship between the state anxiety scale from the first session and the trait anxiety scale, $r = .5$, $p < .001$, Cohen's $d = 0.22$. The higher the anxiety on the trait scale (M = 2.36, SD = .24), the

higher the anxiety on state scale (M = 2.27, SD = .20).

Since both scales correlate positively with each other, the Trait Anxiety Scale was used as the variable for "anxiety" for all further analyses.

General Knowledge Test Performance

Test Accuracy

Results of a dependent, two-tailed t-test across the whole sample on the accuracy of the General Knowledge Test Tasks showed that the performance in the German condition differs **significantly** from the performance in the French condition, $t(52) = 7.96$, $p < .001$, Cohen's $d = 1.09$. As a conclusion the first hypothesis that participants perform better on tasks in their preferred exam language is considered as confirmed. Participants performed better on the General Knowledge Tasks in the German condition (M = .63, SD = .19) than in the French one (M = .41, SD = .21).

Confidence in Accuracy in both language conditions

To check the above findings further, we conducted a dependent, two-tailed t-test across the whole sample on the confidence of the participants in the accuracy of their answers. Results showed that the confidence regarding the answering of the General Knowledge Test Tasks in the German condition **significantly** differs from the confidence regarding the answering of the General Knowledge Test Tasks in the French condition, $t(52) = 3.180$, $p < .01$, Cohen's d = .69. This means that participants are more confident when answering General Knowledge Test Tasks in the German condition (M = 3.72, SD = .76) than when answering General Knowledge Test Tasks in the French condition (M = 3.42, SD = .77).

However, the results of a further dependent, two-tailed t-test show that the solution time when indicating the confidence regarding the answering of the General Knowledge Test Tasks in the German condition (M = 2.69, SD = .74) does **not significantly** differ from the solution time when indicating the confidence regarding the answering of the General Knowledge Test Tasks in the French condition (M = 2.64, SD = 1.04), $t(52) = .28$, p = .778, Cohen's d = .04. So the confident judgements were done equally fast in both language conditions.

Solution Time in both language conditions

We conducted a dependent, two-tailed t-test across the whole sample on the solution time for the General Knowledge Test Tasks. Results showed that the solution time in the German condition **significantly** differs from the one in the French condition, $t(52) = -5.04$, $p < .001$, Cohen's $d = -0.69$. When answering General Knowledge Test Tasks in the German condition, participants were faster (M = 11.55, SD = 4.82) than when answering the tasks in the French condition (M = 14.29, SD = 5.14).

Cognitive Test Anxiety and Task performance

First, we conducted a dependent, two-tailed t-test to check for differences concerning the cognitive test anxiety of the participants between the French and German session. We found that rated cognitive test anxiety in the German condition (M = 1.97, SD = .49) does **not significantly** differ from the rated cognitive test anxiety in the French condition (M = 1.98, SD = .54), t (52) = -.32, p = 0.751, Cohen's d = -.04.

Table 3: Correlations between CTAS (DE/FR), Trait Anxiety and Test Accuracy

Variable		Trait Anxiety	CTAS DE	CTAS FR
CTAS DE	Pearson's r p-value	.37 .006		.88 <.001
CTAS FR	Pearson's r p-value	.37 .006		
Test Accuracy DE	Pearson's r p-value		.16 .245	
Test Accuracy FR	Pearson's r p-value			-.27 <.05
Test Accuracy DE, German preferred Language	Pearson's r p-value		.14 .411	
Test Accuracy FR, German preferred Language	Pearson's r p-value			-.23 .182

Results of a correlation analysis showed that there is a **significant** positive correlation between trait anxiety (M = 2.36, SD = .24) and cognitive test anxiety (M = 2.00, SD = .49) in the German condition as well as in French condition (M = .41, SD = .21) (r = .37, p < .01). These results were expected as trait anxiety is a

more general mechanism than cognitive test anxiety.
Furthermore, we wanted to find out whether there is a relationship between cognitive test anxiety and performance. Results of a correlation analysis on the whole sample (N = 53) showed that in the German condition there is **no significant** correlation between the cognitive test anxiety (M = 2.00, SD = .49) and the performance (M = .63, SD = .19), r = .16, p = .245. However, results of a further correlation analysis showed a significant positive correlation between trait anxiety (M = 2.36, SD = .24) and performance (M = .63, SD = .19) in the German condition (r = .35, p < .01).

In addition, results of a correlation analysis showed that in the French condition, there is a **significant** negative correlation between the performance and the cognitive test anxiety, r = -.27, p < .05. For the French condition, the lower the participant's cognitive test anxiety (M = 1.98, SD = .54), the better their performance (M = .41, SD = .21).

When controlling for trait anxiety in an additional correlation analysis, the obtained results remain the same: there is still **no significant** correlation between cognitive test anxiety (M = 2.00, SD = .49) and performance (M = .63, SD = .19) in the German condition (r = .04, p = .79) and there remains a **significant** negative correlation between cognitive test anxiety (M = 1.98, SD = .54) and performance (M = .41, SD = .21) in the French condition (r = -.27, p < .05).

COGNITIVE TEST ANXIETY AND PERFORMANCE WITH GERMAN AS PREFERRED EXAM LANGUAGE

As outlined in the analysis of our sample, most of the participants (N = 36) stated that they preferred taking exams in German language of instruction. As already mentioned in the results before (see section concerning "General Knowledge Test Performance: Test Accuracy"), participants also performed better on the General Knowledge Tasks in the German condition than in the French condition. This raised the question of whether the language of preference might play a bigger role than the language of instruction when it comes to the relationship between cognitive test anxiety and performance. Therefore, participants who preferred taking exams in German language were filtered out to check for them whether there is a relationship between the performance and the cognitive test anxiety in the German and French condition.
Results of correlation analysis showed that for those who indicated German as their preferred exam language, there is **no significant** relationship between the performance (M = .61, SD = .18) and the cognitive test anxiety (M =1.95, SD = .48) in the German condition, r = .14, p = .411. With regards to the French condition, the results also showed there is **no significant** relationship between the performance (M = .39, SD = .22) and the cognitive test anxiety (M = 1.96, SD = .56) for those who indicated German as their preferred exam language, r = -.23, p = .182.

Summary of the results

With specific reference to our hypotheses, the most important results are summarised below.
In general, there were no differences in the general anxiety between the two sessions.
Differences between language conditions were observed for test accuracy (performance), confidence in the test accuracy and solution time of the General Knowledge test tasks.
No differences between language conditions have been found for cognitive test anxiety and solution time when indicating the confidence in the accuracy of the answers on the General Knowledge Test.
Overall, significant correlations were found between trait anxiety and cognitive test anxiety in both language conditions. Furthermore, in the French language condition there was a negative correlation between cognitive test anxiety and test performance and in the German language condition there was a positive correlation between trait anxiety and test performance. However, in the German language condition, no correlation was found between cognitive test anxiety and test performance. The correlations found remained even when trait anxiety was controlled for.
In the analysis of participants who indicated German as their preferred exam language, no correlations were found between CTA and test performance in either language condition.

Discussion

Since previous studies have examined the influence of either anxiety or language on performance, this study focuses on the integration of these two factors. Thus, the aim of the study is to investigate the influence of Cognitive Test Anxiety (CTA) and language of instruction on performance.

Consequently, two test sessions, either with French or German as language of instruction, were conducted with 53 participants who went through the Luxembourgish school system. The aims of conducting this study were the following:

The first goal was to examine whether there is a relationship between Cognitive Test Anxiety and performance (H1). This hypothesis has been divided into two sub-hypotheses: the higher the CTA in German language of instruction, the lower the test accuracy (H1a). The higher the CTA in French language of instruction, the lower the test accuracy (H1b). The second hypothesis (H2) stated that the language of instruction influences the test accuracy.

To test these hypotheses, we chose a within-subject-design and the following methods: First, a general questionnaire was used to obtain socio-demographic information and to classify the language levels of the participants. Second, a state and trait anxiety questionnaire was applied to measure the general and current anxiety. Third, the participants were asked to complete a general knowledge test on the computer - once in German and once in French. Finally, a validated CTA questionnaire was administered to measure the Cognitive Test Anxiety.

The results concerning the measurement of trait and state anxiety show that participants do not show any differences regarding the state anxiety of the first and the second session. This finding implies that there are no differences in both sessions in terms of the current anxiety before testing. Consequently, differences between the French and German session, if present, cannot be explained by the fact that participants just had a bad day. Likewise, the analysis reveals a positive correlation between trait anxiety and state anxiety. This can be explained by the fact that people who generally have a high trait-anxiety are also more likely to have higher state-anxiety in a given situation (Leal et al., 2017). Previous literature has also shown that there is a difference in performance between people with higher and lower trait anxiety because it also has a detrimental effect on the working memory (Angelidis et al., 2019).

The first sub hypothesis, regarding the German language, turned out to be non-significant and therefore could not be confirmed. The sub-hypothesis regarding the French language was confirmed since results showed a significant negative correlation. The higher the CTA, the worse the test accuracy. A possible explanation for that could be the fact that anxiety becomes maladaptive above a certain level and hence impact the performance in a negative way. These results align with previous findings: students with higher levels of anxiety expected lower grades and furthermore there is a negative correlation between anxiety and achievement (Horwitz, 2001; Teimouri et al., 2019). The fact that there were significant results only in the French part also strengthened the hypotheses regarding the influence of the preferred language. In this case, the majority of the participants chose German as their preferred exam language, so more anxiety and poorer performance in the French part had been expected.

Regarding our second hypothesis, we can conclude that there is an influence of language on test accuracy. In general, the results showed that the participants in the German part of the Computer Tasks performed better than in the French part. This can be explained by the fact that most of the participants chose German as their preferred exam language. Consequently, participants find it easier to answer the tasks in the language they favour. Literature showed that foreign language anxiety is responsible for students' negative emotional reaction to learning a language and therefore influences their performance when taking a test in that language (Horwitz et al., 2001). Further evidence is provided by the solution time. The German part was answered

faster than the French part. Participants were also more confident in the German part. The preference for German language can probably be attributed to the fact that the study was conducted on a Luxembourgish sample. In the Luxembourgish primary school system, all subjects, apart from French, are taught in German for six years. (Hu & Wagner, 2020). In addition, the Luxembourgish language originated from Germanic, which is why there are many similarities between German and Luxembourgish (Scheer, 2017). Consequently, this could explain why the German language is preferred, especially in terms of general knowledge. In addition, a relationship between performance and confidence is observed – the higher the confidence, the better the performance. There is a positive significant correlation for both languages. This result is an interesting finding, as it strengthens the two confirmed hypotheses. In the German part, confidence was higher, and the performance also proved to be better. In the French part, confidence and performance were lower. This could be connected to reasons of anxiety, the higher the anxiety the lower the performance and the lower the performance, the lower the confidence.
Additionally, the results showed that there is a positive correlation between trait-anxiety and the test performance in the German condition. A possible explanation for these findings could be the fact, that for people that are more anxious on a general level (trait anxiety), cognitive test anxiety plays a smaller role when the language of instruction is German (since it's the predominantly preferred language in this study). Concerning French language of instruction, the trait anxiety has less impact and cognitive test anxiety is more dominant (this might be because it is not the mainly preferred language).
Considering these results, we can conclude that Cognitive Test Anxiety has a negative influence on the performance, especially when the instructions are in a non-preferred language of instruction.

Since Cronbach's alpha of the study was high, we can assume a high reliability. Regarding the ecological validity and reasonability, it can affirm that it was not too time-consuming for the study participants. After the completion of the session, each participant received a compensation voucher. Due to the division into two sessions, within the break lasting maximum one hour, effects of fatigue or stress can be ruled out. Furthermore, it can be stated on the part of the investigators that the creation of the study, as well as the execution, was very quick and easy. In addition, the objective execution and the precise scheduling made it possible to divide the testing between many different investigators. Since it was resolved after the study what the content was and what the intentions of the individual measurements were, a high level of transparency can also be established. The usefulness is explained in more detail in the outlook.

The testing of several participants at the same time could have a further influence on anxiety. In the test situations where more participants were tested at the same time, they gave the feedback that it led to a direct competition between them, which leads to more pressure in general. Also trying to be faster than the other test taker or to keep up with them, could in this case have an influence on the accuracy, as it can lead to tasks or questionnaire items being read less thoroughly.
In general, one needs to consider that anxiety may also be triggered by several other factors, which could not be included in this study.

Limitations

To sufficiently test the effect of cognitive test anxiety, it is very important to create a stressful environment for participants. Due to a limited number of space and testers, we only tested one to three people at the same time, which already put mild pressure on the participant. The presence of other people during a test can increase anxiety levels because it creates an exam-like atmosphere. While completing a task in presence of other participants, participants' anxiety levels can get triggered by thinking about how others will perform in the task, instead of focusing on the test. In reality when you are taking an exam,

you are nearly never alone in the room, so for our study it would also have been very interesting to create an even more exam-like atmosphere, and to test many more participants at a time to trigger the full amount of anxiety.

Not only could the presence of other participants during testing have had an influence on the anxiety, but also the presence of the experimenter. But since in this study, many participants were known to the experimenters (as they were friends or family members), anxiety could have decreased. Even though every experimenter tried to be as objective as possible, interpersonal interactions and emotions, which may have created a more comfortable and anxiety-reducing atmosphere, could not completely be avoided.

Another major point is that we could not use the exact same task set in both languages, so there was a slightly different set of tasks in the German part compared to the French part. It is very complicated to match the level of difficulty of two different tasks perfectly. The fact that the German computer task has a better result than the French, can be due to the fact that the German tasks were slightly easier than the French ones. Furthermore, the fact, that Luxembourgish students are generally very fluent in French and in German, needs to be considered regarding the results. For future studies in this field, it would be very important to further examine the effect of language on anxiety. It is also necessary to have a bigger variance in participants' language proficiency, because in our study most of the people claimed German to be their preferred language, which could also have had an influence on the better outcome in the German part.
For future studies of this type, it would be very interesting to test anxiety levels of participants by measuring cortisol.

Educational implications

There are many different contexts in our everyday life in which we are exposed to testing situations and where a good performance is expected: for example, at school, university or later in working life. Exams/tests are used to find out how well a person has mastered a certain subject or whether he or she is qualified enough to do a certain job. Therefore, performance has an influence on the further path of life or course of one's career and thus can be very decisive (Zheng & Cheng, 2018). However, the shown performance during a test does not always reflect the actual or exact level of knowledge/qualifications of a person. Performance can also be influenced by other factors, as the study results show. In particular, our data implies that the language of instruction has an influence on performance. The results of the tasks, which were processed in the preferred language of the respective participants, proved to be better in comparison to the tasks with the non-preferred language. Regarding anxiety, there were significant results in the French part. A negative correlation between the participant´s performance and anxiety is shown: High anxiety correlates with poorer performance. These results imply that anxiety has a negative impact on performance. Confidence also correlates positively with test performance: a high level of confidence correlates with a better performance. These findings suggest that other factors, not only the level of knowledge, have an influence on performance. Therefore, it is not always fair to conclude from the performance of a test only on the level of knowledge or qualifications of a person.

Since every person (especially young people) is repeatedly confronted with exam situations in everyday life, it is important to draw attention to these other factors that can have an impact on the performance. Both sides should be made aware of this: the examiner and the person being examined. The person tested may get a better ability to handle a poorer performance by being aware of these additional influences on performance, since it is clear that the test does not necessarily reflect the actual level of knowledge. On the examiners side, work can be done to minimize the influences of these factors on performance. For example, the language of examination could be chosen by the person being tested (unless

the exam is intended to measure language proficiency). Especially in multilingual countries such as Luxembourg, this could be useful. The Luxembourgish school system regularly is based on two languages: German and French. Even though most Luxembourgish people are capable of speaking both languages fluently, they usually have one preferred language. A test in the chosen and preferred language, could lead to a better performance. In order to reduce anxiety, relaxation exercises could be carried out before the test. Generally, an atmosphere, in which the person being tested feels comfortable and safe, could be created. Thus, a high level of anxiety during the test could be avoided, and this could lead to a better performance.

To summarize, we can affirm that the quality of performance in testing situations depends on several factors and does not necessarily reflect the actual knowledge of a person. The results of this study support this: A significant negative correlation was found between cognitive test anxiety and performance. The higher the cognitive test anxiety, the worse the performance (only significant for the French part). In addition, the study participants performed better in the German part. Since the predominantly preferred language of the participants was German, this finding supports that the instructional language also plays a role. Being tested in one's preferred language may have a positive effect on performance. These findings should be considered and included in all settings where testing performance occurs. It would be desirable to take actions in order to reduce influencing factors such as anxiety and the language of instruction, so that at some point, performance will be more in line with the actual knowledge of a person.

References

Angelidis, A., Solis, E., Lautenbach, F., van der Does, W., & Putman, P. (2019). I'm going to fail! Acute cognitive performance anxiety increases threat-interference and impairs WM performance. *PloS one, 14*(2), e0210824. https://doi.org/10.1371/journal.pone.0210824

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary educational psychology, 27*(2), 270-295. https://doi.org/10.1006/ceps.2001.1094

Greisen, M., Georges, C., Hornung, C., Sonnleitner, P., & Schiltz, C. (2021). Learning mathematics with shackles: How lower reading comprehension in the language of mathematics instruction accounts for lower mathematics achievement in speakers of different home languages. *Acta Psychologica, 221*, 103456. https://doi.org/10.1016/j.actpsy.2021.103456

Horwitz, E. (2001). Language anxiety and achievement. *Annual review of applied linguistics, 21*, 112-126. https://doi.org/10.1017/S0267190501000071

Howie, S. J. (2003). Language and other background factors affecting secondary pupils' performance in mathematics in South Africa. *African Journal of Research in Mathematics, Science and Technology Education, 7*(1), 1–20. https://doi.org/10.1080/10288457.2003.10740545

Kleemans, T., & Segers, E. (2020). Linguistic precursors of advanced math growth in first-language and second-language learners. *Research in Developmental Disabilities, 103*, 103661. https://doi.org/10.1016/j.ridd.2020.103661

Hu, A., & Wagner, J. M. (2020). Zwischen Tradition und Globalisierung. In I. Gogolin, A. Hansen, S. McMonagle & D. Rauch (Eds.), *Handbuch Mehrsprachigkeit und Bildung* (pp. 353-357). Springer VS.

Leal, P. C., Goes, T. C., Da Silva, L. C. F., & Teixeira-Silva, F. (2017). Trait vs. state anxiety in different threatening situations. *Trends in psychiatry and psychotherapy*, *39*, 147-157. https://doi.org/10.1590/2237-6089-2016-0044

M´endez, L. I., Hammer, C. S., Lopez, L. M., & Blair, C. (2019). Examining language and early numeracy skills in young latino dual language learners. *Early Childhood Research Quarterly, 46*, 252–261. https://doi.org/10.1016/j.ecresq.2018.02.004

Ministère de l'Education nationale et de la Formation professionnelle, n.d., Minist`ere de l'Education nationale et de la Formation professionnelle (n.d.). The key figures of the national education (No. 978-99959-1-129–4). Retrieved from https://men.public.lu/dam-assets/catalogue-publications/statistiques-etudes/themes-transversaux/16-17-chiffres-cles-fr.pdf

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, *162*(1-2), 8-13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Scheer, F. (2017). *Deutsch in Luxemburg: Positionen, Funktionen und Bewertungen der deutschen Sprache.* Narr Francke Attempto Verlag.

Spielberger, C. D. (1983). State-trait anxiety inventory for adults. *APA PsycTests.* https://doi.org/10.1037/t06496-000

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition, 41*(2), 363-387. https://doi.org/10.1017/S0272263118000311

Thomas, C. L., Cassady, J. C., & Finch, W. H. (2018). Identifying severity standards on the cognitive test anxiety scale: Cut score determination using latent class and cluster analysis. *Journal of Psychoeducational Assessment, 36*(5), 492-508. https://doi.org/10.1177/0734282916686004

Van Rinsveld, A., Dricot, L., Guillaume, M., Rossion, B., & Schiltz, C. (2017). Mental arithmetic in the bilingual brain: Language matters. *Neuropsychologia*, *101*, 17-29. http://dx.doi.org/10.1016/j.neuropsychologia.2017.05.009

Zheng, Y., & Cheng, L. (2018). How does anxiety influence language performance? From the perspectives of foreign language classroom anxiety and cognitive test anxiety. *Language Testing in Asia*, *8*(1), 1-19. https://doi.org/10.1186/s40468-018-0065-4

Too Anxious to Think: The Effect of Instructional Language and Math Anxiety on Math Performance

Leonie Kuester, Emina Rastoder, Mélissa Reding and Laura Thill

Supervisor: Dr. Mila Marinova

In educational context students' abilities are measured by their performance. However, their performance might not always be a true representation of their abilities as it is influenced by multiple factors, including language of instruction, which seems to influence the math performance through reading comprehension. Additionally, worse understanding of instructional language might also increase Math Anxiety, which in turn might influence math performance.
This study aims to take a deeper look into the relation between language, math, and anxiety from the perspective of a multilingual country such as Luxembourg, where students are taught math first in German and then in French.
In order to obtain the data, 53 participants aged between 19 and 34 (M = 22, SD = 3.4) filled in the State-Trait Anxiety Inventory (STAI) and an originally composed Math Anxiety Questionnaire. They had to solve basic arithmetic problems (TTR) and solve problems from a language-based math performance test. Furthermore, this study was divided into two sessions for which language was counterbalanced.
We hypothesized that Math Anxiety would have an influence on math performance (Hypothesis 1) and that participants who report lower proficiency in the instructional language would also report higher Math Anxiety (Hypothesis 2) and would achieve lower results in the language-based math performance test (Hypothesis 3).
While we did observe evidence for our first hypothesis, we did not find any support for the effect of instructional language on performance or on Math Anxiety. When focusing on participants who prefer French as instructional language, we did however find that they showed less Math Anxiety when solving French math problems. Our results also hint towards the specificity of Math Anxiety, and they show a speed-accuracy trade-off pattern.

Introduction

Many people were at some point in their life in school, university, or some other educational setting. But how is it that some finished this part of life better than others? In educational context students' abilities are mostly measured by their performance. However, their performance might not always be a true representation of their abilities, as research has shown that there are multiple factors which determine whether a person will perform good or not. Such factors might be cognitive (e.g., language) as well as affective (e.g., anxiety).

Starting with the language factor, research has shown that there is a relationship between language of instruction and performance. A study investigating the relationship between English proficiency – with English being the language of instruction – and students´ academic performance in secondary schools in Tanzania, found a weak positive relationship between English language proficiency and academic achievement (Wilson et al., 2012) thus showing that the better the understanding of the instructional language, the better the students perform.

Regarding the anxiety factor, Cognitive Performance Anxiety (CPA) also plays an important role in performance results. Under stress, participants performed significantly

slower during working memory tests, and those with high CPA, compared to those with low CPA, performed even slower (Angelidis et al., 2019), proving that CPA has a negative impact on the working memory while performing. The mechanism of CPA might be explained through attentional control theory. This theory assumes that the goal-directed attentional system – which is governed by expectations, knowledge, and current goals – is impaired by anxiety and that anxiety increases the extent to which processing is influenced by the stimulus-driven attentional system – which is sensitive to salient stimuli. Thus, anxiety not only decreases attentional control but also increases attention to threat-related stimuli (Eysenck et al., 2007).

Moreover, performance can also be influenced by both language and anxiety, as language can induce language anxiety which in turn can influence performance. Multiple studies have shown that there is a significant negative correlation between language anxiety and performance (Horwitz, 2001; Teimouri et al., 2019). This means that students with higher levels of anxiety have lower levels of achievement and students with lower levels of anxiety have higher levels of achievement. Both Horwitz (2001) and Teimouri et al. (2019) studied this phenomenon by measuring the language anxiety of second language learners and concluded that the effects of anxiety on student´s performance are stronger in the context of second language learners.

Furthermore, language and anxiety do not only have an effect on general performance but more specifically also on math performance.

To study the relation between language and math performance, existing data of Luxembourgish 3rd graders has been analyzed, where they had to solve mathematical problems for which the instructions were written in German. The analysis showed that native Luxembourgish speakers performed better in both math achievement and German reading comprehension. Thus, showing that there is a significant negative correlation between reading comprehension in the instruction language and math achievement, which means that lower reading comprehension of math instruction is responsible for lower math achievement (Greisen et al., 2021). Additionally, research shows that second language learners have lower scores in advanced mathematics and linguistic skills and that there is a correlation between linguistic skills and math growth (Kleemans et al., 2020). The study shows that growth in geometry and fraction skills is directly predicted by advanced linguistic skills, and basic linguistic skills indirectly predict growth in advanced math through arithmetic.

Moreover, there is not only a relation between language of instruction and math performance, but also between math anxiety and math performance. Research shows that there is a significant negative correlation between Math Anxiety and math performance, meaning that people with high Math Anxiety have low math performance and vice versa (Barroso et al., 2021; Silver et al, 2022). In their meta-analysis from 2021, Barroso and colleagues investigated the strength of the correlation between math anxiety and math performance across studies conducted between 1992 and 2018, and they found an overall small to moderate, negative and statistically significant correlation of ($r = -.28$). The mechanism of Math Anxiety might also be explained through attentional control theory. During achievement tasks in the math context, worry and anxious thoughts take up limited cognitive resources (e.g., working memory) that are needed to complete a math task, which subsequently reduces efficiency and accuracy on the task (Eysenck et al., 2007).

As illustrated above, both language of instruction and math anxiety have an influence on math performance. Thus, the question arises if language of instruction causes language specific math anxiety which in turn has an influence on math performance. In a study conducted in the Philippines with pupils who are taught math and are tested in a language that is not their first language on one hand, and pupils who are taught and tested in their native language on the other hand, researchers found that there is a significant difference in

Math Anxiety across language of instruction. They also found that pupils with high Math Anxiety have low Math Performance and those with low Math Anxiety have high Math Performance (Barcolod-Iglesia et al., 2021).

A lot of research has already been done in the field of language, math and cognitive anxiety, however, the relation between them has not been studied in multilingual samples. This is why this study aims to take a deeper look into it from the perspective of a multilingual country such as Luxembourg, where through their school career students are taught math first in German and then from seventh grade on in French, with only a fraction of students having one of these languages as their native language. The reason why most students do not have one of these languages as their native language, is on one hand due to a high percentage (around 47.1%) of foreigners who do not have Luxembourgish nationality and mostly do not speak Luxembourgish with their children at home. On the other hand, it is due to children with Luxembourgish nationality who speak Luxembourgish or a different language at home, because their parents have Luxembourgish or a different nationality. The result of this is that for example most French speaking children prefer French and have problems with German at school, and children with other nationalities such as Luxembourgish or German prefer German and have more problems with French at school. Additionally, we can say that the mastering of languages is kind of expected from children who grow up in Luxembourg due to the fact that Luxembourg has three official languages; Luxembourgish, French, and German. With a large part of the population speaking more than one language and having contact during the Luxembourgish school curriculum with at least 4 different languages - Luxembourgish, German, French and English - studying this relationship is important, as it influences the climate in classrooms and the results of students' performance tests, which in turn can influence their future career pathways.

To integrate and analyze all these different findings in our study, we want to test the following three hypotheses.

Our first hypothesis states that Math Anxiety has an influence on math performance. In order to test this hypothesis, we will compare how participants perform in an originally composed language-based math performance test in relation to their score on a Math Anxiety scale. We would expect participants with a higher Math Anxiety to perform significantly worse in the math performance test than participants with a lower Math Anxiety.

The second hypothesis we want to test says that people who report worse proficiency in the instructional language report higher levels of Math Anxiety. In order to test this hypothesis, we will compare participants´ self-indicated level of understanding of the instructional language in relation to their score on a Math Anxiety scale. We would expect participants with a lower level of understanding of the instructional language to score higher on the Math Anxiety scale than those with a better understanding of the instructional language.

Our third hypothesis states that people who report worse proficiency in the instructional language perform worse in the math performance assessment. In order to test this hypothesis, we will compare how participants perform in the language-based math performance test in relation to their self-indicated level of understanding of the instructional language. We would expect participants with a lower level of understanding of the instructional language to perform significantly worse in the math performance test than participants with a higher level of understanding of the instructional language.

Method

Participants

Participants were recruited as part of the experimental practicum course via university's

chat groups, email, and social media platforms. Before the recruitment phase, the target sample size was fixed at 80 to 100 participants. The final number of participants being part of the study was (N = 61), although 4 participants could not take part because of personal reasons or because the criteria of the participants were not met. The actual sample included usable data of 53 participants who were between 19 and 34 years old (Mage = 22, SD = 3.4 years). Out of those, 14 identified their birth sex as male (26.4%) and 38 participants marked their birth sex as female (71.7%). One person did not respond to the question. The study's experimental procedures and methods were approved by the ethical review panel at UL. Prior to participation, written informed consent was obtained. During the testing the participants were compensated with 20-euro vouchers and/or 2 hours or course credits.

Educational and Occupational profile

The educational and occupational profile of the participants entails that most participants were university students, out of which 52.8% work or study in the field of psychology (N = 28). Other included study or work fields are Information Technology, reported by 9.4% of participants (N = 5), Medicine, which is represented by 7.5% of participants (N = 4), and Education, which 5.7% of participants (N = 3) noted as their primary study or work field. Furthermore, other participants are studying or working in the field of Law (N = 2), German Studies (N = 2), and Economics (N = 2), with each field being represented by 3.8% of the sample. Finally, there is one participant, representing 1,9% of the sample, each, who studies or works in the field of History (N = 1), Graphic Design (N = 1), Ergotherapy (N = 1), Social Sciences (N = 1), Physics (N = 1), Mathematics (N = 1), and Engineering (N = 1). Despite our best efforts to include a population as diverse as possible, the majority of the sample are psychology undergraduates. While this may have introduced an unintended sampling bias, it is not much different than the rest of the experimental psychology studies.

Linguistic profile

Regarding the native language, the most commonly reported is Luxembourgish, with 49% speaking it with their father (N = 26) and 52.8% with their mother (N = 28). The second most common native language is French, with 13.2% speaking it with their father (N =7) and 15.1% with their mother (N = 8). There are also participants having a south Slavic native language, for example 7.5% speak Bosnian with their father (N = 4) and 5.7% with their mother (N = 3), and 5.7% speak with both of their parents Serbo-Croatian (N = 3). Other native languages reported are German, with 7.5% speaking it with their father (N = 4) and 7.5% with their mother (N = 4). Additionally, 5.7% of the participants speak Chinese with both of their parents (N = 3), 7.5% speak Portuguese with their father (N = 4) and 5.7% with their mother (N = 3), 3.8 % of the participants speak Italian with their father (N = 2) and 1.9% with their mother (N = 1), and 1.9% of the participants speak Spanish (N = 1), Tamil (N = 1) or Lingala (N = 1) with both of their parents. Several participants mentioned more than one language. In sum, the language profile reported by the participants reveal a typical pattern for the diverse Luxembourgish population, with all the different native language constellations.

Subjective proficiency

When asked how they would rate their proficiency in German, the average of the participants was (M = 3.98, SD = 0.6) on a five-tier-scale, so a scale which goes from 1 to 5. While one participant rated themselves as intermediate (N = 1), 17% of participants rated themselves as advanced (N =9) and 62.3% of participants credited themselves with full proficiency (N =33). Furthermore, 18.9% of the participants reported native level proficiency (N =10).

For their proficiency in French, the average of all the participants is (M = 3.4, SD = 0.9) on the five-tier-scale. Overall, 11.3% rated themselves as intermediate (N = 6), while 56.6% of participants rated themselves as advanced (N = 30), and 15.1% of the participants credited

themselves with full proficiency (N = 8). Furthermore, 17% reported native level proficiency (N = 9).

Finally, asked for the language in which they prefer to take math exams, 56.6% of all participants reported to prefer French (N = 30), 26.4% reported to prefer German (N = 14), while 15.1% participants prefer either English or a combination of the three languages (N = 8), and 1.9% reported to prefer Serbo-Croatian (N = 1).

Design

The study consisted of two sessions for the same sample, also known as a within-sample design. In our study, we want to explore the relationship between the instructional language of the math problems as an independent variable, and the math performance shown by solving the math problems as the dependent variable. The language used in the assessment and the instructions was counterbalanced, meaning that the assessment and instructions were in German during one part of the study, and in French during the other part of the study. Overall, 47% of the participants started with French in the first session and 50.9% of participants started with German in the first session. For one participant, this information was missing.

Materials and Procedure

We collected personal, demographic (e.g., age, birth sex, mother/father tongue) and emotional-related data (State Anxiety, Trait Anxiety, Math Anxiety, Cognitive Performance Anxiety) by letting the participants fill out one general questionnaire and a Trait Anxiety questionnaire in the first session and three anxiety-related questionnaires (State Anxiety, Math Anxiety, and Cognitive Performance Anxiety) per session, as well as data about their performance in two language-based math performance tests. Participants needed to be schooled within the Luxembourgish school system for at least five years during primary and secondary school. They had to understand written German and French at least on a B2-level. Participants were excluded if they had learning difficulties (e.g., dyslexia, dyscalculia), mood disorders (e.g., anxiety, depression), attention disorders (e.g., ADD, ADHD), or were taking stress reducing or tranquillizing medication, i.e., medication that might affect their attention. They also had to have normal or corrected-to-normal vision. Regarding the regulations for gender, ethnicity, or socioeconomic status, there were no characteristics fixed.

The data collection took place from 1st November 2022 until 29th November 2022. The location of the data collection was in three different EEG Labs at the University of Luxembourg in Belval. To make sure that every participant was aware of ethical standards, safety monitoring and the process of the study, they were informed at least 24 hours in advance of their first session via email. Participants were tested in two sessions on two different dates separated by 3 to 7 days, with an approximate duration of 30 to 60 minutes per session. During the first data assessment session, participants were presented with an information sheet in English about what to expect in the study and their informed consent was obtained. After that, participants were asked to fill out questionnaires assessing demographic information, such as their biological sex and their age, possible medical conditions, as well as their perceived language proficiency in French, German, and English, and their language preferences, meaning what language they prefer to speak with their mother or father, which language they prefer for general exams, and which one they prefer for math exams. They were also asked to list all the languages they speak.

After completing the general questionnaire, they also filled out the State-Trait-Anxiety Inventory (STAI-Y, Spielberger, 1983), an instrument to measure State and Trait Anxiety in two scales, with 20 items each (e.g., “I feel nervous.”, “I am jittery.”). State Anxiety is hereby defined as the current, passing emotional state and experience of anxiety while Trait Anxiety describes the stable, general personality characteristics of anxiety. The Trait

Anxiety part of the STAI was administered prior to each session to make sure that the participants' anxiety levels were the same across both sessions, and the order of the items was randomized across sessions. The average score on each corresponding scale was used as a measurement of their State Anxiety in the first and second session and their Trait Anxiety.

After this, the TTR (De Vos, T., 1992. Tempo Test Rekenen. Arithmetic Number Fact Test.) was administered, where the participants had to solve different simple math problems in the categories “Addition”, “Subtraction”, “Division”, “Multiplication” and “Mixed”. For each category, they had one minute to solve as many problems as they could. The researchers stopped the time for the participants and instructed them when to start, when to stop, and when to turn to the next page of the TTR.

The next part was identical for both sessions. The participants did the language-based math performance test on the computer in German or French. First, they were presented with a practice problem and after completing that one, they were presented with six math problems. The experiment was self-paced, but they had a maximum of 25 minutes to solve as many problems as they could. The math problems included both arithmetic and geometry problems taken from real math exams. Furthermore, their accuracy as well as the time it took them to solve each problem were measured. After every problem, participants were presented with a question asking how confident they were that the answer they just gave was actually correct on a scale from 1 to 6, ranging from ‘certainly wrong’ to ‘certainly correct’.

After the math assessment was finished, participants were asked in the same instructional language as the language-based math performance test (i.e., French or German) to fill out a six-tier-scaled questionnaire, expressing their dis-/agreement with several statements to measure their Math Anxiety. The questionnaire was composed originally and included novel items, though it was based on items from other, previously validated tools, such as the Abbreviated Math Anxiety Scale (Hopko et al., 2003); Math Anxiety Scale (Carey et al., 2017); Children’s Anxiety in Math Scale (Jameson, 2013); the motivation questionnaire from the national school monitoring in Luxembourg (Ugen et al., 2013). The average score of participants was used to assess their Math Anxiety. After finishing the first part of the study, participants received their first ten euro voucher.

During the second data assessment session, participants again filled out the State Anxiety Questionnaire before doing the math performance test and then filled out the Math Anxiety Questionnaire again. The TTR and the Trait Anxiety were not administered for the second part. The difference was that the language-based math performance test and the Math Anxiety Questionnaire were in the other language – German if the first part was in French, and French if the first part was in German. Additionally, the order of the Math Anxiety items was randomized across sessions. The duration of the language-based math performance test was the same.

After participants finished their second session, they received their second ten euro voucher. They were also debriefed and were able to ask questions if there were some left unanswered.

Results

Descriptive Data

Table 1: Internal consistency of the Math Anxiety Questionnaires

Estimate	Cronbach's α for the German Math Anxiety Questionnaire	Cronbach's α for the French Math Anxiety Questionnaire
Point estimate	0.89	0.91
95%CI lower bound	0.84	0.87
95% CI upper bound	0.92	0.94

Note. This table shows the internal consistency of the Math Anxiety Questionnaires, which was checked with Cronbach's α, as well as the lower and upper limit of the estimated confidence intervals. Both Questionnaires, in German and in French, had high internal consistency and therefore can be assumed to have a high reliability.

Table 2: Proficiency in German, French and English

	DE	FR	ENG
Mean	3.98 (0.67)	3.37 (0.90)	3.35 (0.76)
Minimum	2.00	2.00	2.00
Maximum	5.00	5.00	5.00

Note. This table shows the mean of the self-indicated proficiency in German (DE), French (FR) and English (ENG) of the whole sample, as well as the standard deviation in brackets. It also shows the minimum and maximum proficiency indicated by participants on the five-tier scale.

Table 3: Differences in State Anxiety and Trait Anxiety in Different Language Conditions

	State Anxiety in the 1st session		State Anxiety in the 2nd session		Trait Anxiety	
	DE	FR	DE	FR	DE	FR
Mean	2.27 (0.23)	2.27 (0.18)	2.23 (0.24)	2.22 (0.19	2.36 (0.26)	2.35 (0.22)
Minimum	1.700	1.950	1.600	1.800	1.850	1.800
Maximum	2.650	2.550	2.650	2.600	2.850	2.750

Note. This table shows the average of the measured anxiety values, split by language condition of each anxiety measurement. State Anxiety was measured twice, once in each session. Trait anxiety was measured only once, in the first session. This table also shows the standard deviation of the average anxiety value in brackets and the minimum and maximum anxiety value indicated by the participants.

Table 4: Results of the Language-Based Math Performance Test

	Percentage of solved problems in the language-based performance test		Solution time of the language-based performance test	
	DE	FR	DE	FR
Mean	0.54 (0.25)	0.53 (0.26)	64.90 (26.02)	102.30 (43.98)
Minimum	0.00	0.00	11.11	35.30
Maximum	1.00	1.00	132.18	192.52

Note. This table shows the average percentage of correct answers in the language-based math performance test, split by language condition, as well as the time taken to solve the math problems. It also shows standard deviation in brackets and the achieved minimum and maximum of both correct answers and seconds taken to solve the math problems.

Table 5: Results of the Math Confidence Rating

	Confidence with the previous solution	Response time of the math confidence test

	DE	FR	DE	FR
Mean	4.44	4.02	64.90	102.30
	(1.09)	(1.10)	(26.02)	(43.99)
Minimum	1.830	1.830	11.110	35.300
Maximum	6.000	6.000	132.18	192.52

Note. This table shows the average value given by participants in response to the confidence question after every problem of the language-based math performance test and the time taken to answer the question split by language condition. It also shows standard deviation in brackets for both variables and the indicated minimum and maximum confidence with given answers as well as the minimum and maximum response time.

TTR

For the TTR, participants could achieve 40 points per section, meaning a minimum of zero points if they did not solve any problems at all and a maximum of 200 points if they solved every problem correctly. In our study, participants scored a mean of M = 139.89 (SD = 27.34) with minimum of 76 points and maximum of 197 points.

Hypothesis Testing

To test our first hypothesis regarding the influence of Math Anxiety on math performance, we conducted a correlation analysis between Trait Anxiety, Math Anxiety for both German and French, and math test accuracy. The math test accuracy was generated from the data of the language-based math performance test and indicates how accurately the participant responded to the math questions during the language-based math performance test. We found a significant positive correlation between German Math Anxiety and French Math Anxiety ($r = 0.765$, $p < 0.001$). For the German Math Anxiety and math test accuracy we found a significant negative correlation of ($r = -0.488$, $p < 0.001$) as shown in Figure 1 and for the French Math Anxiety and math test accuracy the correlation was significantly negative ($r = -0.375$, $p = 0.007$), as shown in Figure 2. This shows that the higher the Math Anxiety, the lower the test accuracy. For the trait anxiety however, we found no significant correlation with neither the Math Anxiety, nor the math performance.

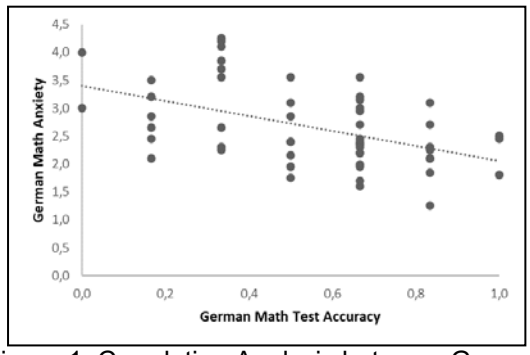


Figure 1: Correlation Analysis between German Math Anxiety and German Math Test Accuracy

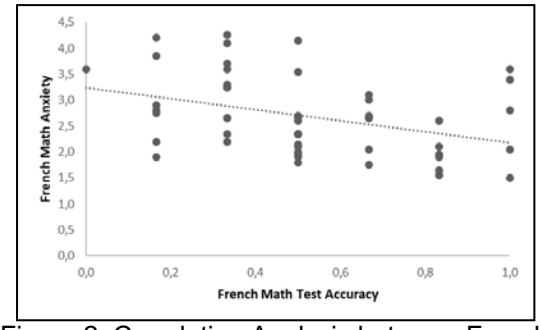


Figure 2: Correlation Analysis between French Math Anxiety and French Math Test Accuracy

For the second hypothesis, which states that participants who report worse proficiency in the instructional language report higher levels of Math Anxiety, we conducted a correlation analysis between instructional language and Math Anxiety. The correlation analysis shows that there is no significant relationship between the reported understanding of the instructional language and the reported Math Anxiety levels, neither for German nor for French. Accordingly, we did not find any significant results that support our second hypothesis.

For the third hypothesis, which states that participants perform lower in the math performance assessment when they report worse proficiency in the instructional language, we did not find any significant results. The second hypothesis already investigated the relation between math performance and math anxiety, whereas this hypothesis analyzes the relation

between math performance and language of instruction. We conducted a correlation analysis between instructional language and math performance. The correlation analysis shows that there is neither a significant relationship between the self-indicated categories of language level in French and the math performance, nor between the self-indicated categories of language level in German and the math performance. Accordingly, we did not find any significant results that support our third hypothesis, either.

Further Exploratory Analyses

The Relationship Between Language Preference, Math Anxiety and Math Test Accuracy

We wanted to observe whether the testing language preferences may have an effect on the math achievement. Therefore, we divided our sample into people who prefer French as the language of instruction for math exams (N= 36) and into people who prefer German as the language of instruction for math exams (N=17). Participants could and did mention several languages.

First, we conducted an independent samples *t*-test, comparing the French Math Anxiety, German Math Anxiety, and math test accuracy of participants who prefer French as language of instruction for math exams with participants who prefer a different language for math exams. This test shows that the French Math Anxiety of participants who prefer to do math exams in French (M = 2.54, SD = 0.78) is significantly lower than it is for those who prefer to do math exams in a different language (M = 2.99, SD = 0.58), $t(49) = 2.07$, p = 0.04.

For the German Math Anxiety and the math test accuracy however, there are no significant differences between the participants who prefer French for math exams and participants who prefer a different language.

Additionally, we also conducted an independent sample t-test, comparing the French Math Anxiety, German Math Anxiety, and math test accuracy of participants who prefer German as language of instruction for math exams with participants who prefer a different language for math exams. This test shows no significant differences at all.

Examining Language-Related Differences in Math Performance and Math Anxiety

We also conducted a paired samples t-test for Math Anxiety, math test accuracy, math test solution time, math confidence response, and math confidence solution time, for both instructional languages, German and French, showing that the math test solution time for the German version (M = 64.90, SD = 26.02) is significantly lower than the one for the French version (M = 102.30, SD = 43.98), $t(52) = - 6.85$, $p < 0.001$.

Additionally, the conducted t-test shows that the math confidence response for the German version (M = 4.44, SD = 1.09) is significantly higher than the one for the French version (M = 4.03, SD = 1.10), $t(52) = 4.35$, $p < 0.001$.

The Relationship Between Instructional Language, Math Anxiety, Math Performance, and Confidence

Beyond our hypotheses we did further data analyses to investigate the relationship between instructional language, Math Anxiety, math performance, and confidence. We conducted a correlation analysis between Math Anxiety, math test accuracy, math test solution time, math confidence response, and math confidence solution time in both German and French. The math test solution time specifies the time that the participant needed to answer the math question. The data from the math confidence response indicates how sure the participants were whether their answer to the math question was correct or not. Regarding the math confidence solution time, this shows how quick participants responded to the math confidence question.

THE RELATIONSHIP BETWEEN GERMAN INSTRUCTIONAL LANGUAGE, MATH ANXIETY, MATH CONFIDENCE AND MATH TEST ACCURACY.

THE RELATIONSHIP BETWEEN MATH ANXIETY AND MATH CONFIDENCE RESPONSE. The significant negative correlation between Math Anxiety and math confidence response (r = - 0.59, $p < 0.001$) shown in Figure 3 indicates that the higher the Math Anxiety, the less sure the participants were that they answered correctly.

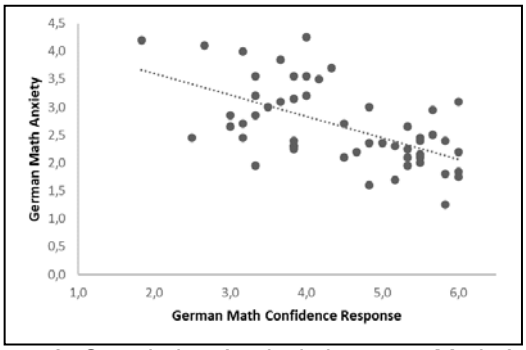


Figure 3: Correlation Analysis between Math Anxiety and Math Confidence Response

THE RELATIONSHIP BETWEEN MATH ANXIETY AND MATH TEST SOLUTION TIME. We also found a significant negative correlation between Math Anxiety and math test solution time (r = - 0.32, p = 0.02) shown in Figure 4. These results show that the higher the Math Anxiety, the faster the participants answered.

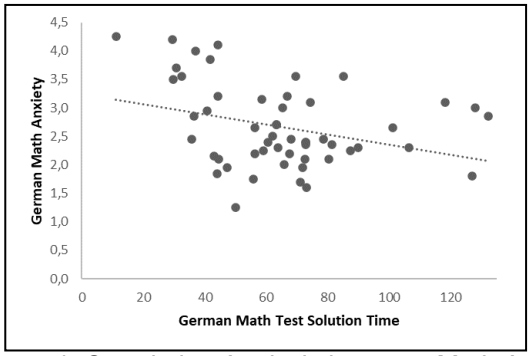


Figure 4: Correlation Analysis between Math Anxiety and Math Test Solution Time

THE RELATIONSHIP BETWEEN MATH TEST ACCURACY AND MATH TEST SOLUTION TIME.

As can be seen in Figure 5, there is a significant positive correlation between the math test accuracy and math test solution time (r = 0.30, $p = 0.03$), indicating that the higher the test accuracy, the longer participants needed to answer the math questions.

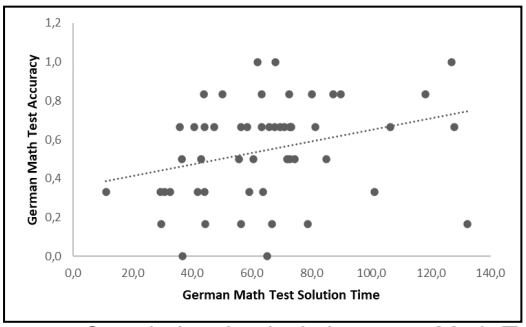


Figure 5: Correlation Analysis between Math Test Accuracy and Math Test Solution Time

THE RELATIONSHIP BETWEEN MATH TEST ACCURACY AND MATH CONFIDENCE RESPONSE. Finally, there is a significant positive correlation between math test accuracy and math confidence response (r = 0.63, $p < 0.001$), showing that the higher the test accuracy, the more sure the participants were that they answered correctly.

THE RELATIONSHIP BETWEEN FRENCH INSTRUCTIONAL LANGUAGE, MATH ANXIETY, MATH CONFIDENCE AND MATH TEST ACCURACY.

THE RELATIONSHIP BETWEEN MATH ANXIETY AND MATH CONFIDENCE RESPONSE . The significant negative correlation between Math Anxiety and math confidence response (r = - 0.56, $p < 0.001$) depicted in Figure 6, shows that the higher the Math Anxiety, the less sure the participants were that they answered correctly.

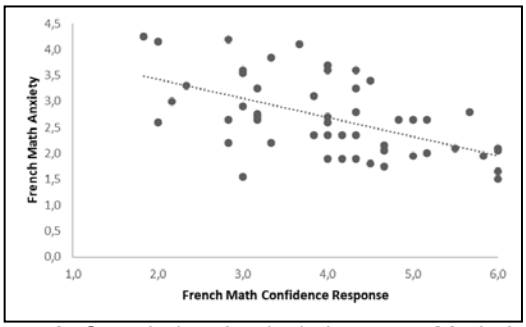


Figure 6: Correlation Analysis between Math Anxiety and Math Confidence Response

The Relationship Between Math Anxiety and Math Test Solution Time. Between the Math Anxiety and math test solution time is a significant negative correlation of (r = - 0.31, p = 0.03), shown in Figure 7, meaning that the higher the Math Anxiety, the faster the participants answered.

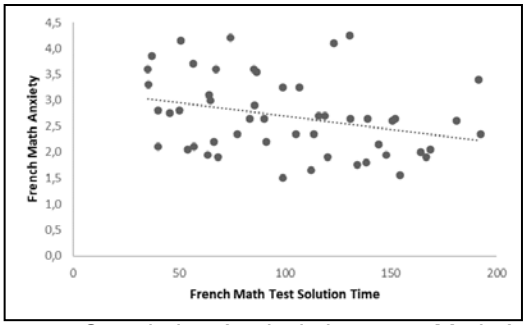


Figure 7: Correlation Analysis between Math Anxiety and Math Test Solution Time

The Relationship Between Math Test Accuracy and Math Test Solution Time. For the variables math test accuracy and math test solution time we found a significant positive correlation of (r = 0.43, p = 0.001) depicted in Figure 8, meaning the higher the test accuracy, the higher the test solution time.

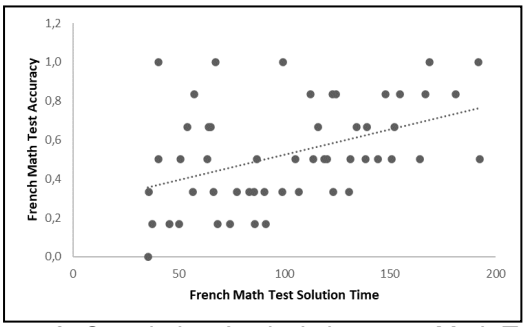


Figure 8: Correlation Analysis between Math Test Accuracy and Math Test Solution Time

The Relationship Between Math Test Accuracy and Math Confidence Response. The significant positive correlation between math test accuracy and math confidence response (r = 0.53, p < 0.001) indicates that the higher the test accuracy, the surer the participants were that they answered correctly.

The Relationship Between Math Anxiety and Math Confidence Solution Time. One result that is different from the German analysis is that for the French analysis we found a significant positive correlation between Math Anxiety and math confidence solution time (r = 0.35, p = 0.013), indicating that the higher the test anxiety, the longer the participants took to report how sure they were that they answered correctly. This correlation was not significant in the German condition. Figure 9 depicts the significant correlation for the French Math Anxiety and confidence.

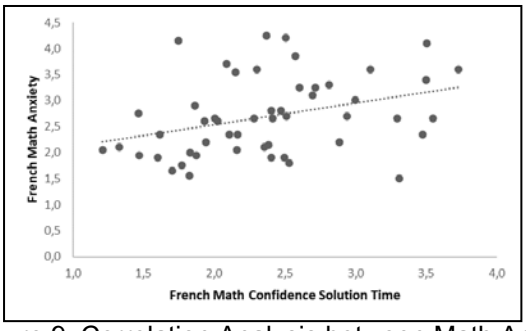


Figure 9: Correlation Analysis between Math Anxiety and Math Confidence Solution Time

The Relationship Between the TTR, Trait Anxiety, and the Math Anxiety in German and French Test Conditions. Since the language-based math performance test required more cognitive resources as it required procedural operations and knowledge, we wanted to examine whether Math Anxiety is related to math performance or to more automated math knowledge, which is typically used in the literature (i.e., the TTR). Therefore, we conducted a correlation analysis between TTR, Math Anxiety in both German and French, and trait anxiety. We found a significant negative correlation between the TTR and the German Math Anxiety (r = - 0.507, p < 0.001) and a significant negative correlation between the TTR and the French Math Anxiety (r = - 0.506, p < 0.001), which means that the higher the reported Math Anxiety, the lower the participants scored in the TTR.

For the trait anxiety we found no significant correlations with the TTR or with the Math Anxiety.
Figure 10 depicts the correlation between the TTR and the German Math Anxiety and Figure 11 shows the correlation between the TTR and the French Math Anxiety.

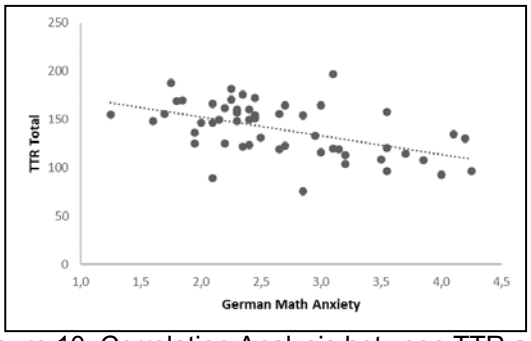


Figure 10: Correlation Analysis between TTR and German Math Anxiety

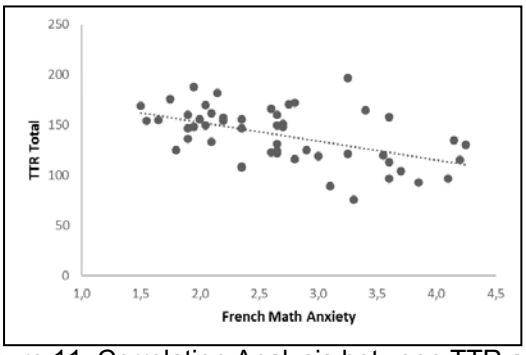


Figure 11: Correlation Analysis between TTR and French Math Anxiety

Summary of the Results

For our first hypothesis we found a significant positive correlation between German Math Anxiety and French Math Anxiety (r = 0.765, p < 0.001). For the German Math Anxiety and math test accuracy we found a significant negative correlation of (r = - 0.488, p < 0.001) and for the French Math Anxiety and math test accuracy the correlation was significantly negative (r = - 0.375, p = 0.007).

For both the second and the third hypothesis we did not find any significant results. With an independent t-test we found that French Math Anxiety of participants who prefer to do math exams in French (M = 2.54, SD = 0.78) is significantly lower than it is for those who prefer to do math exams in a different language, t(49) = 2.07, p = 0.04. We did not find a significant equivalent in the German preferring sample.

The paired samples t-tests show that the math test solution time for the German version (M = 64.90, SD = 26.02) is significantly lower than the one for the French version (M = 102.30, SD = 43.98), t(52) = - 6.85, p < 0.001, and that the math confidence response for the German version (M = 4.44, SD = 1.09) is significantly higher than the one for the French version (M = 4.03, SD = 1.10), t(52) = 4.35, p < 0.001.

For the instructional language German, the correlation analyses show a significant negative correlation between Math Anxiety and math test accuracy (r = - 0.49, p < 0.001), a

significant negative correlation between Math Anxiety and math confidence response (r = - 0.59, $p < 0.001$) and a significant negative correlation between Math Anxiety and math test solution time (r = - 0.32, p = 0.02). Additionally, they show a significant positive correlation between math test accuracy and math test solution time (r = 0.30, p = 0.03) and a significant positive correlation between math test accuracy and math confidence response (r = 0.63, $p < 0.001$).

For the instructional language French, the correlation analyses show a significant negative correlation between Math Anxiety and math test accuracy (r = - 0.38, p = 0.007), a significant negative correlation between Math Anxiety and math confidence response (r = - 0.56, $p < 0.001$) and a significant negative correlation between the Math Anxiety and math test solution time (r = - 0.31, p = 0.03). Additionally, they show a significant positive correlation between math test accuracy and math test solution time (r = 0.43, p = 0.001), a significant positive correlation between math test accuracy and math confidence response (r = 0.53, $p < 0.001$) and a significant positive correlation between Math Anxiety and math confidence solution time (r = 0.35, p = 0.013).

Finally, between the TTR and the German Math Anxiety we found a significant negative correlation (r = - 0.507, $p < 0.001$) and between the TTR and the French Math Anxiety a significant negative correlation (r = - 0.506, $p < 0.001$). For trait anxiety we found no significant correlations with the TTR or with the Math Anxiety.

Discussion

Hypothesis Testing

The results for our first hypothesis for which we found a significant positive correlation between German Math Anxiety and French Math Anxiety were expected because the Math Anxiety questionnaires are standardized and are supposed to be correlated. We can also say, for both German and French, that participants who reported higher Math Anxiety performed worse in the math performance test than those who reported lower Math Anxiety. Additionally, for both German and French, the Trait Anxiety has no significant influence on the Math Anxiety or the math test performance. These results support our first hypothesis and are in line with previous studies (e.g., Barroso et al., 2021; Silver et al, 2022), where it has been shown that Math Anxiety has a negative influence on math performance.

For our second hypothesis which states that people who report worse proficiency in the instructional language report higher levels of Math Anxiety, we do not have any significant findings that would either support the hypotheses or contradict it, meaning that the results are inconclusive. Considering that worse proficiency might add another hurdle to solving math problems and score well in the testing situation, these results were surprising to us. But, as we will mention in our limitations as well, since a lot of our participants reported advanced to full proficiency in all testing languages, they might not have had to worry about understanding at all.

Coming to the third hypothesis which states that people who report worse proficiency in the instructional language perform worse in the math performance assessment, we also do not have any significant findings that would either support the hypothesis or contradict it. Therefore again, we can neither accept the null hypothesis, nor the alternative hypothesis. These results also make sense in the light of our assumption that understanding the instructions was not a source of worry for the participants since they could understand both languages on a sufficient level.

Further Exploratory Analyses

The Relationship Between Language Preference, Math Anxiety and Math Test Accuracy. Since most participants were schooled in the Luxembourgish School System and should be used to solving math problems in French, it would be expected that they show less Math Anxiety and solve more prob-

lems correctly in the French condition. This was observed for the French preferring sample. This shows that there might be an effect of language on Math Anxiety, but as we suspect, this effect might exist for other languages, but we did not find it because our sample consists mainly of people who prefer French as language of instruction for Math exams. Nevertheless, participants took generally longer to solve the problems in French. This mechanism might be unique to multilingually educated samples and has not been explored before. Additionally, it does seem logical that focusing on only one area of difficulty might lower the anxiety surrounding solving a set of problems, when people who are being tested do not have to think about their language comprehension and can shift all of their attention on solving the math problems. Research has shown that those with high anxiety divide their attention between the task and ruminations stimulated by high anxiety. Those with lower anxiety however, devote less of their attention to anxiety related preoccupations and more of their anxiety to the task, thus performing better (e.g., Sarason, 1988; Wine, 1971).

Examining Language-Related Differences in Math Performance and Anxiety. We found that participants needed longer to answer the math questions in French, than they did for the German part. It is possible that some tasks in the French condition were unintentionally harder than in the German condition, since math difficulty is quite difficult to match evenly across different tasks. This would be supported by participants being more confident with the answers given in the German language-based math performance test. It might also be possible that participants, even though they were proficient enough in both languages to understand the problems, read slower in French than in German.

The Relationship Between Instructional Language, Math Anxiety, Math Performance, and Confidence. Participants who reported lower Math Anxiety took longer to answer in the language-based math performance test, but also scored higher in the language-based math performance test.

Additionally, participants who reported higher Math Anxiety were faster to answer in the language-based math performance test, but also scored lower in the language-based math performance test. Accordingly, we can say that the participants who reported higher Math Anxiety, were faster in the language-based math performance test, indicating a speed-accuracy trade-off. This could be because participants with a high Math Anxiety might be overwhelmed by the math questions and therefore might give up more easily without even really trying to find the correct solution to the question or might want to leave the uncomfortable situation. This then might lead to them just selecting a random answer, which takes way less time than actually doing calculations to find the right answer. This behavior might be explained through attention control theory, which states that worry and anxious thoughts take up limited cognitive resources that are needed to complete an achievement task, which subsequently reduces accuracy and efficiency on the task (Eysenck et al., 2007). Participants who took longer to solve the math problems on the other hand had more time to be thorough, which might explain why they answered more accurately.

In addition, participants who reported higher Math Anxiety were less sure that they answered correctly in the language-based math performance test, and participants who were more confident that they answered correctly in this test, actually scored higher in this test.

Lastly for this analysis, we can only state for the French part that participants who reported higher Math Anxiety needed longer to report how sure they were that they answered correctly. For the German part this result was not significant.

The Relationship Between the TTR, Trait Anxiety, State Anxiety and the Math Anxiety in German and French Test Conditions. We found that the participants who reported higher Math Anxiety scored lower in the TTR, irrespectively of the

language condition of the session in which the TTR was administered.

Additionally, the correlation between the TTR and the German Math Anxiety, and the correlation between the TTR and the French Math Anxiety are nearly equal. As the TTR is not language-based and is a more automatized process, we can assume that the Math Anxiety of the participants is not due to language effects. Furthermore, we did not find a significant correlation between the TTR and Trait Anxiety, or between Math Anxiety and Trait Anxiety, probably meaning that Trait Anxiety is not pivotal to an individual showing Math Anxiety. The nonsignificant finding between Math Anxiety and Trait Anxiety is not in line with the majority of previous studies that investigated this relationship and showed a significant positive correlation between Math Anxiety and Trait Anxiety (e.g., Betz, 1978; Hembree, 1990; Paechter, 2017; McAuliffe & Trueblood, 1986; Hopko et al., 2003) However, there are a few studies that also found no significant relationship between Math Anxiety and Trait Anxiety (e.g., Wu et al., 2012; Klados et al., 2015; Grezo & Sarmány-Schuller, 2018). Additionally, our correlation analysis for the results of the TTR and State Anxiety yielded no significant results, meaning that there probably was no influence from either side. This is interesting because, as we showed earlier, there was a correlation between lower TTR results and higher Math Anxiety. This difference, as well as the missing relationship with Trait Anxiety, might give a hint as to how specific Math Anxiety can be. Based on these results, it seems to be not only a general increase of stress, but a feeling of anxiety centered completely on the math problems and it seems that a generally higher anxiety level might not influence the accuracy and speed with which participants solve easy math problems.

Limitations and Future Directions

The bias of our sample consisting of mostly females, mostly students, especially psychology students, as mentioned beforehand is typical for psychological study samples but could possibly influence the reliability of the observed results. Especially considering the typicality of these biases, future studies could benefit from bigger samples, in which biases are easier to filter out. If future studies aim to derive usable and constructive recommendations for a broad population, researchers should make efforts to gather a gender balanced sample that reflects all types of study fields and education paths in their target population, as well as a sample that reflects well the age structure of the population that is supposed to be examined.

Additionally, students who did not study math or math-related topics had to rely on memories to answer some of the questions included in the Math Anxiety Questionnaire, since especially older participants did not take a math exam in (possibly) several years. This difference between younger and older participants might also influence their familiarity with the test topics, as participants who just recently finished high school might have been more used to them. However, there were not enough participants to compare age groups. Therefore, future studies should either aim to include more participants from different age groups or to limit the age criteria even more.

Since we specifically searched for participants who speak French and German, the actual (not self-proclaimed) proficiency might not differ that much between participants. Therefore, our results are not applicable to a classroom in which all levels of language proficiency are existent. One improvement could be made by adding a short test to assess participants' language skills to assess the actual proficiency and include participants with vastly differing skill levels, which in turn might paint a different picture of the triggered Math Anxiety. This is important to yielding results that can claim to reflect the reality of students of all ages in Luxembourg.

Additionally, most of our participants were recruited and tested by someone they know, and many of our participants might have been

used to testing situations since taking part in psychological studies is a mandatory part of the study program in psychology. Both of these factors might have lowered their feeling of anxiety.

Furthermore, in our data analyses, we found several interesting effects not covered by our hypotheses, which could be interesting reference points for future studies. To explore the speed-accuracy trade-off in future studies, as well as other possible explanations for participants with higher Math Anxiety taking less time in the language-based math performance test, we recommend post-study interviews examining participants' feelings and inner processes during the study.

Another possible focus point could be the effect that participants who prefer French as an exam language show less Math Anxiety when taking the French language-based math performance test than participants who prefer a different exam language. If and in which dimensions this effect might be found for different languages. and how it relates back to performance should be tested with better-balanced samples, as well as different multilingual samples.

On a broader scale, we would look at our results with caution. Nevertheless, from the effects observed in our sample, we feel comfortable advising lecturers in almost every position to provide exams in a language that students feel comfortable with or prefer, wherever possible, even though we are aware that this might not be possible since translation of exams can be a difficult process without sufficient resources. In general, we can also advise lecturers and students to incorporate actions aiming to empower students to manage their Math Anxiety, as there is no possibility to completely eliminate uncomfortableness around math exercises.

Conclusion

The goal of our study was to examine existing relationships between Math Anxiety, language of instruction and math performance that might influence the performance shown by multilingual students in language-based math performance tests.

Overall, our results provide evidence for the hypothesis that Math Anxiety influences math performance negatively, for both, language- and non-language-based math performance tests. These results were equivalent for German and French. We did find that students with higher Math Anxiety showed a speed-accuracy trade-off pattern and were less sure with the answers they gave. Moreover, our results show as well that Math Anxiety is specific, i.e. not influenced by the general or current anxiety shown by an individual. Surprisingly, we did not observe evidence for the influence of language of instruction on either Math Anxiety or math performance in general. However, our results also imply that Math Anxiety could be lowered by participants taking the test in the language they preferred, even though this was only true for one condition. These seemingly contradicting results should be further investigated.

We can conclude that the language of instruction, even though it did not play a major role in our analyses, might still be a noteworthy factor in the complex processes influencing students' performance, and more research will be necessary to build a more reliable body of knowledge.

While providing students with their preferred language of instruction might give them a better chance to perform to the best of their abilities, we should also consider enabling them to handle their Math Anxiety, if present. For this, it would be vital to understand the speed-accuracy trade-off even better. We nevertheless encourage lecturers to consider our results when thinking about the design of exams and language-based math performance tests.

References

Angelidis, A., Solis, E., Lautenbach, F.,van der Does, W., & Putman, P. (2019). I'm going to fail! Acute cognitive perfor-

mance anxiety increases threat-interference and impairs WM performance. PloS one, 14(2), e0210824

Bacolod-Iglesia, A. B., Dundain, H. H., Vega-re-Miñoza, M., & Casimiro,A. B. (2021). Teaching Mathematics in 'different tongues': An analysis of mathematics anxiety and proficiency among elementary grade learners. Linguistics and Culture Review, 5(S2), 860-891. https://doi.org/10.21744/lingcur.v5nS2.1523

Barroso, C., Ganley, C. M., McGraw, A.L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between Math Anxiety and math achievement. Psychological Bulletin, 147(2),134. https://doi.org/10.1037/bul0000307

Betz, N. E. (1978). Prevalence, distribution, and correlates of math anxiety in college students. Journal of Counselling Psychology, 25(5), 441–448.

De Vos, T. (1992). Tempo Test Rekenen (TTR). Arithmetic Number Fact Test. Nijmegen: Berkhout.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. Emotion, 7(2), 336–353. https://doi.org/10.1037/15283542.7.2.336

Hembree, R. (1990). The nature, effects And relief of mathematics anxiety. Journal for Research in Mathematics Education, 21(1), 33–46.

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The Abbreviated Math Anxiety Scale (AMAS): Construction, validity, and reliability. Assessment, 10(2), 178–182.

Horwitz, E. (2001). Language anxiety and achievement. Annual review of applied linguistics, 21, 112-126.

Greisen, M., Georges, C., Hornung, C., Sonnleitner, P., & Schiltz, C. (2021). Learning Mathematics with shackles: How lower reading comprehension in the language of mathematics instruction accounts for lower mathematics achievement in speakers of different home languages. Acta Psychologica, 221, 103456. https://doi.org/10.1016/j.actpsy2021.103456

Grežo, M., & Sarmány-Schuller, I. (2018). Do Emotions Matter? The Relationship Between Math Anxiety, Trait Anxiety, and Problem Solving Ability. Studia Psychologica, Vol. 60, No.4, 2018, 226-244. https://doi.org/10.21909/sp.2018.04.764

Klados, M. A., Simos, P., Micheloyannis, S., Margulies, D. & Bamidis, P. D. (2015). ERP measures of math anxiety: How math anxiety affects working memory and mental calculation tasks? Frontiers in Behavioral Neuroscience, 9:282

Kleemans, T., & Segers, E. (2020). Linguistic precursors of advanced math growth in first-language and second-language learners. Research in Develop-mental Disabilities, 103, 103661. https://doi.org/10.1016/j.ridd.2020.103661

McAuliffe, E. A., & Trueblood, C. R. (1986). Factor analysis: A tool for studying mathematics anxiety. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, USA.

Paechter, M., Macher, D., Martskvishvili,K., Wimmer, S., & Papousek, I. (2017).Mathematics anxiety and statistics anxiety. Shared but also unshared components and antagonistic contributions to performance in statistics. Frontiers in Psychology, 8:1196.

Sarason, I. G. (1988). Anxiety, self-preoccupation, and attention. *Anxiety Research, 1*(1), 37. https://doi.org/10.1080/10615808808248215

Silver, A. M., Elliott, L., Reynvoet, B., Sasanguie, D., & Libertus, M. E.(2022). Teasing apart the unique contributions of cognitive and affective predictors of math performance. Annals of the New York Academy of Sciences. https://doi.org/10.1111/nyas.14747

Spielberger, C. D., Goruch, R. L.,Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1983). Manual for the state-trait inventory STAI

(form Y). Mind Garden, Palo Alto, CA, USA.

Teimouri, Y., Goetze, J., & Plonsky, L.(2019). Second language anxiety and achievement: A meta-analysis. Studies in Second Language Acquisition, 41(2), 363-387.

Wilson, J., Komba, S. C. (2012). TheLink between English Language Proficiency and Academic Performance: A Pedagogical Perspective in Tanzanian Secondary Schools. World Journal of English Language. https://doi.org/10.5430/wjel.v2n4p1

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin, 76*(2), 92104. https://doi.org/10.1037/h0031332

Wu, S. S., Barth, M., Amin, H., Malcarne,V., & Menon, V. (2012). Mathanxiety in second and third graders and its relation to mathematics achievement. Frontiers in Psychology, 3:162.

Exploring gender differences in math performance and financial literacy

Charlotta WELLING, Clémentine OFFNER, Chloé MENDES, Medina MUSIC, Anita ADROVIC, Angelika STEIN

Supervisor: Doctoral Researcher Styliani POLITI

In the following study, we examined the problem of gender differences in mathematical performance and financial literacy. There are several different factors that might have contributed specifically to the gender differences in mathematics.
In a review paper, Zheng Zhu (2007) addressed the different approaches men and women take to mathematical problem solving. He as well mentioned not only the fact that in some investigation's women perform worse than men, but also that in some cases women even perform better in mathematics, depending on the type of test. Since mathematics and finance are inextricably linked (Baron & Lorraine, 2015; Scott, 2018) we examined both in our study. For this reason, differences in financial knowledge as well as behavior regarding finance between genders were also considered. We recruited university students aged 18 to 35 from STEM and non-STEM field. We used computerized self-assessment questionnaires. The questionnaires contained standardized tasks in financial literacy and behavior as well tasks measuring mathematical performance. Results showed that there were no differences between genders in terms of achievement scores. However, a significant effect could be observed in the arithmetic task and in the math mastery task within the non-STEM students. As being a very complex situation of gender differences in mathematical problem solving it would be wise to investigate these differences in more depth, focusing on STEM and non-STEM students, as more differences could be found in larger samples.

Introduction

Mathematics

The gender gap in mathematics has already been discussed for a long time (Hedges & Nowell, 1995). Previous research has shown, that at an early age there are no intrinsic gender differences in numerical abilities (Kersey et al., 2018). The study by Else-Quest et al. (2010) drew the conclusion that discrepancies usually develop with advancing age and occur mainly in societies with inequalities in education and gender equality between men and women. Due to this, greater equality in terms of educational achievement, would lead to gender similarities in mathematics (Guiso et al., 2008).

In the research by Rodríguez et al. (2020) they investigate whether there are significant gender differences in intrinsic motivation, perceived competence, negative feelings, and mathematics anxiety.

Furthermore, according to recent research (Lee, & Anderson, 2015), the suggestion that gender differences in attitudes toward mathematics are more pronounced in coeducational schools than in single-sex schools, raises the larger issue of gender stereotyping and the possible impact of the school environment. It could be that students in a coeducational school are more likely to conform to gender stereotypes, while in single-sex schools there

is more latitude for students not to conform to gender expectations.

Stereotypes related to gender and mathematical ability state that women are less mathematically talented than men. A growing body of research suggests that these gender stereotypes of mathematics influence women's interest as well as performance in mathematics (stereotype threat), and that gender differences in mathematics performance occur primarily in gender stereotype environments (Franceschini et al., 2014).
For example, there have been found gender differences in exams like SAT (Combs et al., 2009).

As gender inequalities in STEM still exist in most populations nowadays (Casad et al., 2019), the question arises whether, and to what extent, the gender gap among higher education students could be eliminated.

Financial Literacy

There are also significant indirect effects of gender in financial literacy through such factors as financial anxiety. Women feel greater anxiety about financial matters because of stereotype threats or other unexplained mechanisms like the direct effect of gender (Tinghög et al., 2021).

According to the study from Bottazzi and Lusardi (2021) using data from more than 140 countries on financial literacy showed that there are gender differences everywhere, from developing countries to advanced economies.
It also shows that, on one hand, parental background, particularly the role of mothers, matter for girls' financial literacy. On the other hand, social and cultural environment in which girls and boys live, also play a crucial role in explaining gender differences.

Moreover, the research from Sunderaraman et al. (2020) showed that numerical reasoning had a strong association with financial literacy. For this reason, in our study except from possible gender differences in math performance, the relationship between math performance and financial literacy is explored.

Present study

All the points listed above inspired us to conduct a study on gender differences in math performance and financial literacy among university students.
Therefore, we have conducted following hypotheses:

1. Male participants' overall math performance will be better than female participants' overall math performance.
2. Male participants will perform better at the problem-solving tasks and the math fluency task than the female participants. There will not be gender differences for the math mastery task.
3. Male participants will perform better at the financial literacy measure.
4. There will be a positive correlation between financial literacy and math performance.

Methodology

University students (18-35 years old) who are not diagnosed with any attention disorder or learning difficulty were recruited. Participants' scores were compared in three different mathematics tasks, in financial literacy and financial behavior questionnaires (see complete description in Measures). The reliability coefficients for the two financial questionnaires, financial literacy questionnaire ($\alpha = 0.51$) and financial behavior questionnaire ($\alpha = 0.19$) were low. That means that questionnaires used were not reliably measuring the concepts of financial literacy and financial behavior. This could be due to a low number of questions, as we attempted to keep the assessment time as short as possible (approximately 60 minutes). The participants were tested either in a laboratory or at the library of University of Luxembourg. In total 44 participants were included in the final analysis, 2 participants were excluded because they scored too low in the Advanced Progressive Matrices Test,

which was used as a screening test of high-level observation
skills. Further, two participants were excluded for being highly distracted during the data collection. The population is presented in Figure 1. The distribution of the socioeconomic status is illustrated for men and women in Figure 2 and Figure 3 correspondingly.

Figure 1
Participants' mean age, gender, and study field distribution.

Participants	N=44	
Age mean (σ)	23,97 (3,812)	
Gender	STEM Students	Non-STEM Students
Male(N=21)	12	9
Female (N=23)	11	12

Figure 2
SES distribution in female population

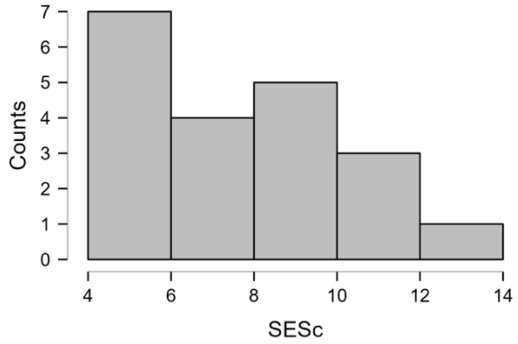
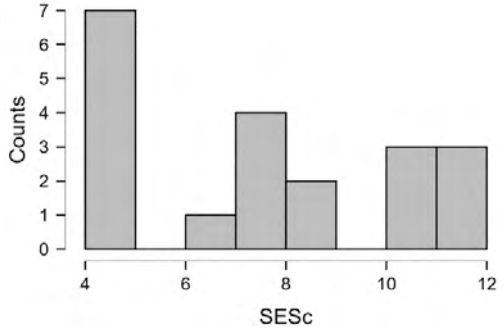


Figure 3
SES distribution in male population



Measures

To test whether there is a significant difference between female and male university student in terms of mathematical performance and financial literacy, the participants had to fill out some computerized self-assessment questionnaires and perform computerized tasks using the platform SosciSurvey.

High-level observation skills

In the first place a short form of the Raven's Advanced Progressive Matrices, which was developed by Arthur and Day in 1994, was presented to the participants to check their cognitive abilities and intelligence level. More precisely, it assessed their reasoning skills by having them determining which element, out of multiple possibilities, is the piece that is missing in the pattern that is presented to them. This short form consists of 12 items and has a duration of approximately 15 minutes.

Mathematics

Furthermore, the participants had to complete the three math performance measures, more precisely the Geometrical Problem-Solving Task and the Arithmetical Problem-Solving Task, which were presented in form of multiple-choice items and have a similar format as the problems used in SAT exams, as well as the Math Mastery task (Figure 4).
To measure the participant's geometrical problem-solving abilities, they had to solve 4 geometrical problems with an average processing time of 8 minutes. These tasks consisted for example of determining the perimeter or the area of shapes such as a pyramid. Therefore, the solving of these problems required knowledge on geometric principles.
The Arithmetical Problem-Solving tasks contained four arithmetical problems with a focus on algebra and took approximately 8 minutes. In this task participants had to work with whole numbers, precents as well as fractions and decimals. These tasks included problem-solving with additions, substructions, multiplications, and divisions.

The Math Mastery tasks is an adapted version of the brief math assessment created by Steiner & Ashcraft (2012) and which has a duration of approximately 10 minutes. The 10 items were developed to be similar in difficulty and type as the items of the standardized WRAT3 (Snelbaker et al., 2001) and had an increasing difficulty. In the adapted version the items 9 and 10 were replaced by a complex multiplication problem, which required knowledge of fractions. These items, that had a focus on algebra, were adapted by the researcher as these items were resembling items measured in the Arithmetic Problem-Solving Task. At the beginning, the participants had to complete items with very easy calculations of addition, subtraction, multiplication, and division. This approach has been used to motivate the participants to try and complete as many items as possible.

Financial Literacy and behavior

Finally, the participants had to complete a questionnaire linked to financial literacy and another one linked to financial behavior, which measured the participant's ability to make financial decisions and to ensure their financial wellbeing.

First, the participants had to fill out a Financial Literacy questionnaire. The items were created by Atkinson and Messy (2012) according to OECD suggestions and are the same as the ones used in the study of Preston and Wright (2019). The questionnaire contained 5 item concepts such as the interest rate, inflation, diversification, investment risk and money illusion.

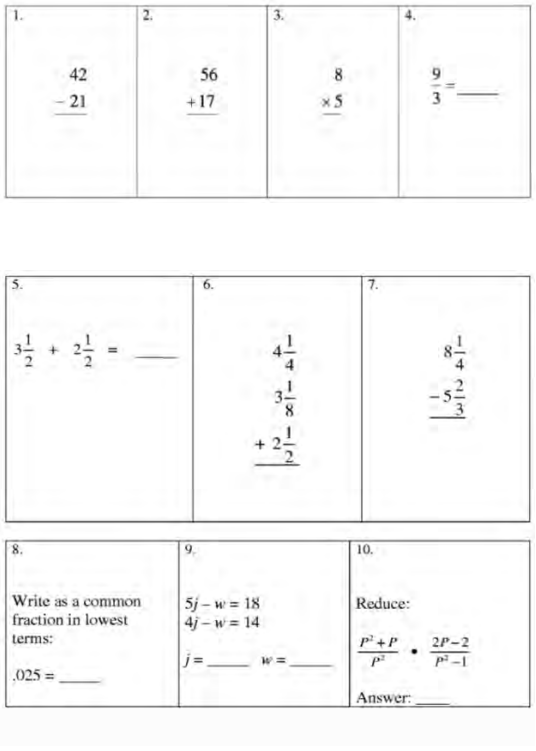
In addition to that, a Financial Behaviour questionnaire, with items that have been used in Financial Literacy and Inclusion OECD/INFE Survey across Countries and by Gender (2013) was presented to the participants. This questionnaire asked more specifically about how the participants are handling their finances. The questionnaire measures for example the responsibility of the household budget, the consideration of affordability, the management of their financial affairs, the setting and achievement of their financial goals, and their informed financial decisions. There were also questions on whether they have borrowed money, whether they are saving money and whether they pay their bills on time.

Demographics and attitude

At the end a demographics questionnaire concerning the participant's biological sex, their gender, socioeconomic status, and field of study as well two questions about the participant's attitude on gender, science and mathematics had to be filled out. To get an idea about participants' attitudes we asked them for instance whether they think that one gender is better in math than the other. The items of that questionnaire were based on items used in the study of Hall, Robinson and Limin Jao (2020).

Figure 4

Math Mastery Task by Steiner & Ashcraft (2012)



Results

To assess the overall performance between the women and men, an independent t-test

was conducted. The results of the independent samples t-test revealed that in terms of overall math performance there was no significant difference between women (M=-.575, SD=2.808) and men (M=.630, SD = 1.481), t (42) =-1.756, p = 0.086.

Further to evaluate male and female's performance in the problem-solving tasks as well as the gender differences in the math mastery task, a repeated measures ANOVA was conducted.
Using the Greenhouse-Geisser correction, the results of the repeated measures ANOVA showed that there was no significant interaction between gender and math performance task. (F= (1, 1.563) =0.977, p =0.364). see Figure 6, Figure 7, and Figure 8 for the performance of each gender of the three math tasks.

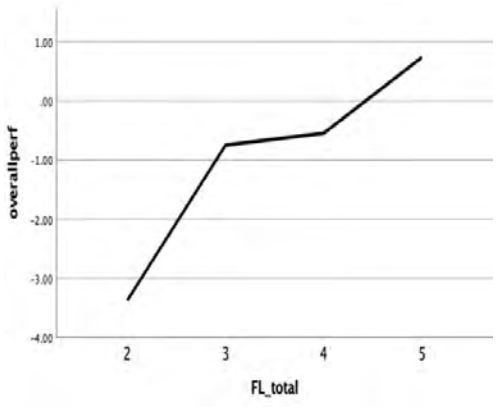
Moreover, to analyze male and female participants' performance190. (The financial literacy measure and the financial behavior measure, two individual independent samples t-tests were conducted. The results revealed that in terms of financial literacy that there was no difference between female (M=4.304, SD=0.926) and male participants (M=4.429, SD=0.926), t (42) = -0.444, p=0.659. Regarding financial behavior there was no difference for female (M=61.304, SD=14.555) and male (M=62.857, SD=16.475) participants neither t (42) = -0.332, p= -0.742. (See participants' performance in Figure 9 and Figure 10).

Between financial literacy measure and math performance a correlation analysis revealed that there is a moderate correlation between the arithmetic problem-solving task and financial literacy, r =.470, p = .001. In addition, there was a moderate association between the geometry problem-solving task and financial literacy, r =.429, p = .004. However,

there was no correlation between the Math Mastery task and financial literacy, r =174, p = .258. Contrary to financial literacy, financial behavior was not significantly correlated with any of the math tasks.

Figure 5

Correlation between overall math performance and financial literacy measure (y-axis= overall math performance, x-axis= Financial Literacy measure)



To further explore the gender differences in each of the math tasks, students' performance was compared between genders in each of the math performance tasks.

Three individual independent samples t-tests were conducted to compare student's performance in the respective math tasks. The t-tests showed that there was a significant effect in the arithmetical problem-solving task between female students (M=-.3024, SD= 1.173) and male students (M=.3312, SD=.644), t (42) =-2.190, p= .034. (See figure 6)
Further the t-test revealed that there was no significant effect in the geometrical problem-solving tasks between female students (M=-.082, SD=1.192) and male students (M=.090, SD=.754), t (42) =-.569, p=0.573 (see Figure 4). Moreover, there was no significant effect found in the Math Mastery Task in female (M=-.190, SD=1.078) and male students (M=.208, SD=.885), t (42) =-1.333, p=.190. (See Figure 8)

To explore and compare students' performance in the financial measures two independent samples t-tests were conducted. Bartell, T-G. Test revealed that there was no effect in the financial literacy measure between females (M=4.30, SD=.926) and males

(M=4.43, SD=.926), t (42) = -.444, p=.659. (See figure 9). The second t-test revealed that there was no effect in the financial behavior measure between females (M=61.30, SD=14.555) and males (M=62.86, SD=16.475), t (42) =-.332, p=.742 (See Figure 10).

Figure 6

Performance of female and male students in arithmetical problem-solving tasks. (1=female, 2=male)

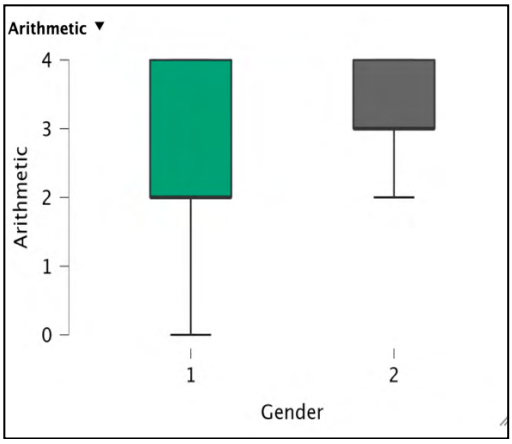


Figure 7

Performance of female and male students in Geometrical problem-solving task. (1=female, 2=male)

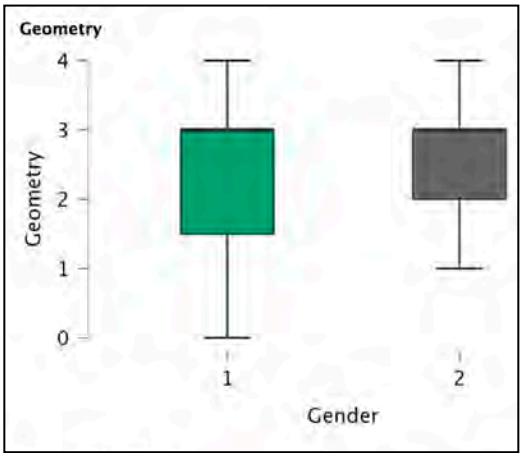


Figure 8

Performance of female and male students in Math Mastery Task. (1=female, 2=male)

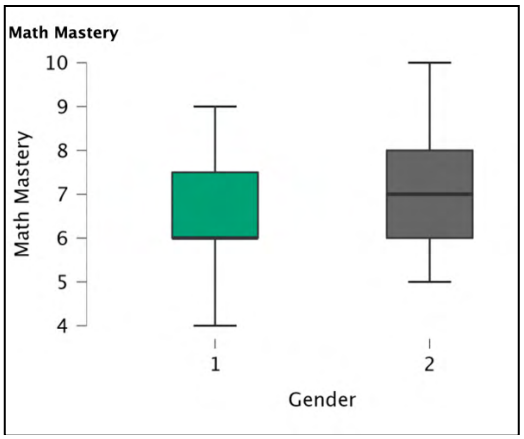


Figure 9

Performance of female and male students in Financial Literacy measure (1=female, 2=male)

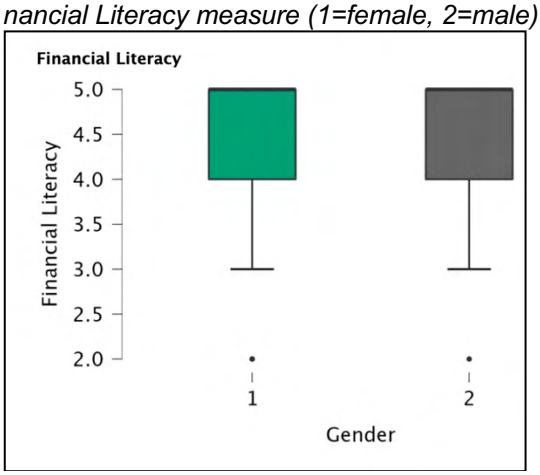


Figure 10

Performance of female and male students in Financial Behaviour measure. (1=female, 2=male)

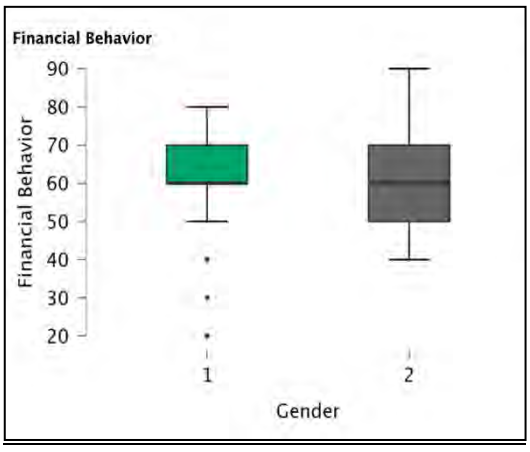
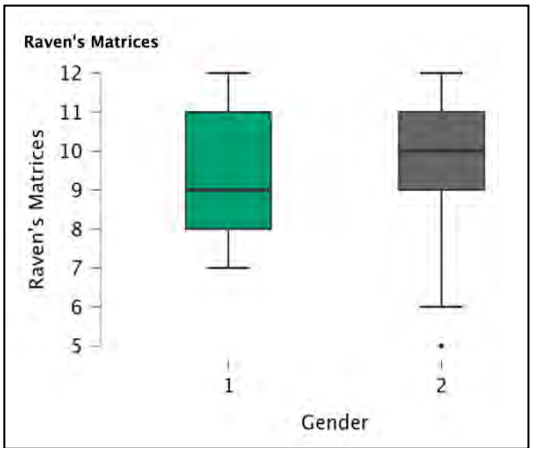


Figure 11

Performance of female and male students the Raven's Matrices task (1=female, 2=male)



Additional Data Analysis

Subsequently, we decided to conduct further analysis and compared the genders by separating them into an additional category. Indeed, on the one hand we compared the results of STEM students and on the other hand those of non-STEM students, to see if there are significant gender differences in the respective study fields.

More specifically, first, we compared their overall performance, secondly their performance in arithmetic mathematics, then in geometry and finally the mathematics mastery tasks. For this purpose, we conducted mainly independent sample t-tests, which allowed us to draw the following results.

Beginning with Non-STEM students: the results of the t-test comparing arithmetic task outcomes only revealed a tendency toward a significant difference between females (M=-.639, SD=1.201) and males (M=.151, SD=.474), t (15.143) =-2.073, p=.056. In terms of geometry performance, the results revealed no significant difference between females (M=-.221, SD=1.413) and males (M=-.053, SD=0.712), t(17.025)=-.356, p=.727. However, the results in terms of mathematics mastery task revealed a significant difference between females (M=-.521, SD=1.118) and males (M=.500, SD=1.076), t (17.738) =-2.116, p=0.049.

To conclude with the overall performance, the results didn't reveal any significant differences between females (M=-.4603, SD=1.053) and males (M=.1992, SD=0.466), t (15.998) =-1.931, p=.071.

Proceeding with the STEM students: the t-test results comparing arithmetic task outcomes revealed no significant difference between females (M=.064, SD=1.076) and males (M=.466, SD=.738), t (17.511) =-1.035, p=.315. Likewise, the results concerning geometric performance did not reveal any significant difference between females (M=.068, SD=.0940) and males (M=.198, SD=.798), t (19.742) =-.355, p=.726. The same conclusion can be drawn regarding the math mastery task, as the results did not reveal any significant difference between females (M=.170, SD=.954) and males (M=-.010, SD=.678), t (17.916) =.519, p=.610.

By looking at their overall performance the conclusion is identical, as there were no significant differences between females (M=.1011, SD=.725) and males (M=.2182, SD=.533), t(18.298)=-.438, p=.667

Additionally, by analyzing the Ravens Matrices results for all participants, STEM and non-STEM, we found no significant differences which shows us that we tested participants with the same intelligence levels. This means the participants can be compared on the same level.

Discussion

To sum up, overall, there was no difference in the results between genders regarding math and financial literacy performance, nor in financial behaviour scores. More specifically, no gender differences were found in the geometry task, which is in line with previous findings (Capraro, 2001), and neither in the brief assessment of mathematics achievement (here called Math Mastery task). These results

agree with the recent evidence of the elimination of the gap in more gender-equal societies (Guiso et al., 2008b).
However, the arithmetic problem-solving task was significant (r =.470, p = .001) between all the students, which could be of concern considering the multiple factors contributing to gender differences in math performance (2007) and previous findings showing the evidence of gender differences in strategy use in mathematics (e.g., Gallagher et al., 2000). Moreover, we also found a tendency to a significant difference in non-STEM students' Arithmetic Task (t (19) =-1.858, p=.079), as well as a significant effect in non-STEM students' Math Mastery Task (t (19) =-2.104, p= .04). This could suggest that the gender differences tend to outwash themselves depending on the populations. However, in terms of geometry performance, the results between non-STEM students revealed no significant difference between females (M=-.221, SD=1.413) and males (M=-.053, SD=0.712), t (17.025) =-.356, p=.727.
The differences that were found solely in the non-STEM students highlight the importance of education and exposure to mathematics in the elimination of gender differences (United Nations, n.d.).
Regarding financial literacy in contradiction to other studies (Tinghög et al., 2021) we did not find significant gender differences. This could possibly be explained based on the characteristics of our sample of students and based on limitations of our instrument for financial literacy. Therefore, gender differences in financial literacy should be further explored.
Finally, the results exploring the relationship of financial literacy and mathematics showed that financial literacy is connected to specific math tasks but not to all of them. Future studies should clarify what parts of mathematics education could support strengthening the financial literacy of men and women.

Limitations

It is an important consideration and a necessary component for research studies to have a justified and adequate sample size. Therefore, one of our limitations is our small sample size for comparing groups.
In addition, the sample consisted exclusively of university students. In this sample, there is little variability regarding socioeconomic factors, which have been found to explain to a great extent the gender gap (Breda et al., 2018).
Furthermore, we used items that are widely used to measure financial literacy (Preston & Wright, 2019). However, these items seemed to be quite easy in the present study and for this reason, were possibly not ideal for uncovering individual differences in financial literacy among students.
Additionally, we have limited ourselves to only a few math tasks to test mathematical performance. However, there are more facets of math than can be further explored. For example, timed math tasks may uncover gender differences. For instance, previous studies showed that men performed faster under pressure than women in timed tasks (Shurchkov, 2012) this could lead to further differences in their performance.
Moreover, both reliability coefficients of our two questionnaires (i.e., the financial literacy and financial behaviour questionnaire) were very low and not acceptable. It could be due to the low number of questions or the poor interrelatedness between items or possible heterogeneous constructs. This limitation suggests that there is a need for standardized and validated measures on financial literacy and financial behaviour that could be used in short assessment time.

Outlook

Even though overall we didn't find any differences between gender, the topic as explained in the introduction, is still open to be explored further. A wider population can be broken down into different smaller groups and therefore be better explored. For example, you could look at gender differences when looking further into the socio-economic status of a population (Lee et al., 2016). For this reason, they should be explored in regard to typical

sociodemographic factors (e.g., age, socioeconomic status, immigration background etc.).
It would also be very interesting to do a longitudinal study and explore the role of siblings. Since siblings grew up in the same environment it would be interesting to examine if differences in their mathematical and financial literacy performance as well as their financial behaviour develop over time.
Furthermore, it would be interesting to set up a longitudinal study starting in primary school. This would help to see the evolution of math performance and financial literacy throughout childhood to adulthood and explore the differences that arise for both men and women as result of socialization conditions. Moreover, one could focus in such a longitudinal study on the family factors like the home math environment (Daucourt et al., 2021) or look at different teaching styles and approaches (Nelson L., 2011) This would help us to gain a better insight into the connection between education systems and gender differences that may be significant. Previous studies showed that strategy flexibility is a source of gender differences in mathematical ability (e.g., Gallagher, 2000). With this said, it could be possible that the way problems are presented or taught in school also influence gender differences. In conclusion, there is no clear one final answer regarding the topic of gender inequalities.

References

Scott, E. (2018, September 06). All About the Relationship Be-tween Mathematics and Eco-nomics. Essay-Corp. https://www.essaycorp.com/blog/relationship-between-mathematics-and-economics

Arthur, W., & Day, D. V. (1994). Development of a Short form for the Raven Advanced Progressive Matrices Test. Educational and Psychological Measurement, 54(2), 394–403.

Bartell, T.G., Bieda,K.N.,Putnam,R.T.,Bradfield,K.,&Dominguez, H. (Eds.).(2015).Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. East Lansing, MI: Michigan State University

Bottazzi, L., & Lusardi, A. (2021c). Stereotypes in financial literacy: Evidence from PISA. Journal of Corporate Finance, 71, 101831.

Bonawitz, E. (2019). Faculty Opinions recommendation of No intrinsic gender differences in children's earliest numerical abilities. Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature.

Breda, T., Jouini, E., & Napp, C. (2018). Societal inequalities amplify gender gaps in math. Science, 359(6381), 1219–1220.

Casad, B. J., Franks, J. E., Garasky, C. E., Kittleman, M. M., Roesler, A. C., Hall, D. Y., & Petzel, Z. W. (2020). Gender inequality in academia: Problems and solutions for women faculty in STEM. Journal of Neuroscience Research, 99(1), 13–23.

Combs, J. P., Slate, J. R., Moore, G. W., Bustamante, R. M., Onwuegbuzie, A. J., & Edmonson, S. L. (2010). Gender differences in college preparedness: A statewide study. *The Urban Review*, *42*(5), 441-457.

Daucourt, M. C., Napoli, A. R., Quinn, J. M., Wood, S. G., & Hart, S. A. (2021). The home math environment and math achievement: A meta-analysis. *Psychological Bulletin*, *147*(6), 565.

Franceschini, G., Galli, S., Chiesi, F., & Primi, C. (2014). Implicit gender–math stereotype and women's susceptibility to stereotype threat and stereotype lift. Learning and Individual Differences, 32, 273–277.

Financial knowledge questions in the 2015 OECD/INFE Toolkit for Measuring Financial Literacy and Financial Inclusion. (2016). Financial Education in Europe, 66–67.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008b). Culture, Gender, and Math. Science, 320(5880), 1164–1165.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender Differences in Advanced Mathematical Problem Solving. Journal of Experimental Child Psychology, 75(3), 165–190.

Hedges, L. V., & Nowell, A. (1995). Sex Differences in Mental Test Scores, Variability, and Numbers of High-Scoring Individuals. Science, 269(5220), 41–45.

Hall, J., Robinson, T., & Jao, L. (2020). Non-binary people's views of gender and mathematics. Mathematics Education Across Cultures: Proceedings of the 42nd Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.

Kersey, A. J., Braham, E. J., Csumitta, K. D., Libertus, M. E., & Cantlon, J. F. (2018). No intrinsic gender differences in children's earliest numerical abilities. Npj Science of Learning, 3(1).

Lee, K., & Anderson, J. (2015). Gender Differences in Mathematics Attitudes in Co-educational and Single Sex Secondary Education. Mathematics Education Research Group of Australasia.

Lee, Y. S., Park, Y. S., & Ginsburg, H. (2016). Socio-economic status differences in mathematics accuracy, strategy use, and profiles in the early years of schooling. ZDM Mathematics Education, 48, 1065–1078.

Nelson Laird, T. F., Garver, A. K., & Niskodé-Dossett, A. S. (2011). Gender gaps in collegiate teaching style: Variations by course characteristics. *Research in Higher Education*, *52*(3), 261-277.

OECD (2013). Financial literacy and inclusion: Results of OECD/INFE survey across countries and by gender

Preston, A. C., & Wright, R. E. (2019). Understanding the Gender Gap in Financial Literacy: Evidence from Australia. Economic Record, 95(S1), 1–29

Rodriguez, S., Regueiro, B., Piñeiro, I., Estévez, I., & Valle, A. (2020). Gender differences in mathematics motivation: Differential effects on performance in primary education. *Frontiers in psychology, 10, 3050.*

Robert M. Capraro. (2001). Exploring the Influences of Geometric Spatial Visualization, Gender, and Ethnicity on the Acquisition of Geometry Content Knowledge.

Steiner, E. T., & Ashcraft, M. H. (2012). Three brief assessments of math achievement. Behavior Research Methods, 44(4), 1101–1107.

Sunderaraman, P., Barker, M., Chapman, S., & Cosentino, S. (2020). Assessing numerical reasoning provides insight into financial literacy. Applied Neuropsychology: Adult, 29(4), 710–717.

Snelbaker, A. J., Wilkinson, G. S., Robertson, G. J., & Glutting, J. J. (2001). Wide range achievement test 3 (wrat3). In *Understanding psychological assessment* (pp. 259-274).

Shurchkov, O. (2012). Under pressure: gender differences in output quality and quantity under competition and time constraints. Journal of the European Economic Association, 10(5), 1189-1213.

Tinghög, G., Ahmed, A., Barrafrem, K., Lind, T., Skagerlund, K., & Västfjäll, D. (2021). Gender differences in financial literacy: The role of stereotype threat. Journal of Economic Behavior &Amp; Organization, 192, 405–416.

United Nations. (n.d.). Education as the Pathway towards Gender Equality. https://www.un.org/en/chronicle/article/education-pathway-towards-gender-equality

Zheng Zhu. (2007). Gender Differences in Mathematical Problem Solving Patterns: A Review of Literature. International Education Journal, 8(2), 187–203.

The Positive Influence of Video Games on Mental Well-being

De Cillia Lisa, Lara Donner, Mahr Michel, Schiltz Anne-Sarah, Thill Noah, Weber Raphael, Will Vanessa

supervision: M.Sc. Gary Lee Wagener, Asst.-Prof. Dr. André Melzer

This paper proposes an opposing view to the misconception that videogames only have negative effects on mental well-being as it demonstrates that enhancing effects certainly exist. Based on previous literature of Seligman's (2011) PERMA theory of well-being, this research examined positive effects of videogames on well-being using the WHO-5 Index. It was subsequently analyzed how the assessed well-being varied according to variables such as player types and play time. In addition to that, gender analysis was conducted, focussing on the influence of gender on play time and game preference. A sample of 967 participants gathered by an online questionnaire served to demonstrate the findings of this article, which established that well-being correlates positively with the amount of fulfilled PERMA columns, as well as a moderate play time (7-10 hours per week). Among all the player types, it was found that the 'High Performer' recorded the highest well-being. Furthermore, there were no differences found regarding the play time between genders. Moreover, the gender analysis of this report has shown that there was a potential bias towards specific game genres preferences between male and female participants.

Introduction

Video games are largely known in the news and academic literature for their bad reputation, which results in the fact that mostly there is a fundamentally negative attitude towards video games and electronic games in general (Przybylski & Weinstein, 2016). Findings by Ferguson (2015) suggest that even among clinicians, debates about video games are influenced by historical patterns of generational conflict, with harmful beliefs held primarily by older individuals who are hostile to younger generations. In this regard, one can find many studies which emphasize the negative aspects, for example on the correlation between violent video games and aggression (Calvert et al., 2017; Anderson et al., 2010) or video game disorders and addiction (Rosenberg, 2020; Bean et al., 2017). As a result of this polarizing representation, just a few studies seem to address the potential positive attributes of video games, and thus much research remains to be done (Granic et al., 2014).

Accordingly, this study aims to take a different perspective with this research project and contribute to the few positive findings related to video games and mental health. Therefore, the aim of the present study is to investigate the relationships between video games and psychological well-being and to address various aspects in detail.

Theoretical Background

The theoretical background of the study is based on Seligman's PERMA model (Seligman, 2011). The model is an acronym for factors that contribute to well-being. The PERMA model is not empirically proven or derived from a study. However, all five factors are confirmed by scientific research. The 5 pillars of the model consist of Positive Emotions (P), Engagement (E), Relationships (R), Meaning (M), and Accomplishment (A). Well-being is interpreted not only as the mere absence of mental disorders, but as the opposite (Jones et al., 2014). Video games might inherently have design elements that focus on attributes of well-being, and playing video games

provides opportunities for mental health, considering the PERMA Model, this can be shown using the 5 pillars.

All scientific statements and knowledge about the PERMA Model in relation to video games, are based on the paper by Jones et al (2014) and will be explained in the following.
Positive emotions, in this context, refer to relaxation, stress reduction and steam off, which can be felt due to the distraction while playing video games.
Engagement refers to an emotional involvement or commitment to an object or domain of interest and to the experiential intensity of a relationship or interaction. It also refers to temporal engagement or interactions with activities and social partners in the immediate environment, where a strong relationship between engagement and positive well-being has been demonstrated.
Relationships describes that playing with real-life friends allows players to share their sense of accomplishment and success in order to strengthen and restore their self-esteem and esteem as gamers. The friendships that gamers make online are in some cases comparable to or better than their real-life friendships. Likewise, playing with others can reduce problem gambling behavior. It appears that video games give players the opportunity to express themselves in ways that might make them uncomfortable in real life because of their appearance, gender, sexuality, and/or age. In addition, video games encourage people to stay in touch.
Meaning is described as participation in meaningful activities gives a sense of purpose and fulfilment to daily life. The search for meaning has been associated with participation in activities that contribute to something greater than self. Video games also have the potential to eliminate loneliness; even when friends and family are far away, video games can allow players to interact, share, and be social.
Accomplishment and a sense of satisfaction in life can be achieved by working towards goals and it contributes to a sense of competence and well-being. Ryan et al. (2006) believe that the psychological appeal of games is primarily due to the fact that they evoke feelings such as competence and thus increase psychological well-being.

Furthermore Jones et al. (2014) discovered in their findings that video games can contribute to mental health, especially when played for a moderate amount of time. In their paper they describe a moderate playing time of 7-10 hours per week. Below this level, video games are not important enough in the lives of consumers to have a real effect on the psyche. Above this time (over 10 hours per week), other important areas of life may be neglected.

Further studies, like Phan et al. (2012) and Hamlen (2010), found gender differences regarding game preferences and playing times. In the studies of Hamlen's (2010) it was shown that men prefer shooter games, in contrast to women who play mostly casual games. In Phan et al. 2012 study, it was found that men play video games more often compared to women.

In addition, the MOPS, motivation to play scale (Holl et al., 2022), was included in the online study. The scale assesses gaming motivation and allows a division of gamers into four different player types: 'Casual Player', 'High Performer', 'Crafter' and 'Highly involved player'. This study used the MOPS questionnaire to classify the participants into the four gamer types.

Hypothesis

Based on the theoretical background, the following hypotheses were created:
1) The different types of gamers differ in terms of the 5 columns of the PERMA model. (Jones et al., 2014).
2) Men play more video games than women (Hamlen, 2010).
3) Men and women differ in game preference (Phan et al., 2012).
4) Moderate play time leads to a positive effect on well-being (Jones et al., 2014). According to Jones et al.,

around 7-10 hours per week is moderate playing time.
5) The more columns of the PERMA model are fulfilled, the better the well-being.

Research questions

From these hypotheses the following research questions were derived:
1) Do people spent more time playing video games, when gaming has a higher status for them? Status means the importance of video games compared to other aspects of life.
2) Is the relation between time spent playing and well-being moderated by the video game status? People for whom video games have a higher status, play more video games, and therefore have greater well-being. Playing more video games means a longer and more intense playing time.
3) Does the well-being of the four player types depend on the number of pillars of the PERMA model they fulfill?

Methods

Participants

Using G*Power 3.1.9.7 (Faul et al., 2007) a statistical power of 0.9 and an alpha of 0.05 the sample size for the online study needed to be 270 participants. We reached a sample size of 2147 finished data sets, ending in 1586 valid cases. The participants were recruited via the Moodle portal of the University of Luxembourg and through personal contacts. Moreover, the study was promoted on social network sites (e.g. Facebook groups, Discord servers, Instagram stories) and interactive online gaming sites (e.g. Twitch).

Participants (n = 1586) were assessed for careless responding. Participants with unusually short completion time (i.e., relative speed index ≥ 2; Leiner, 2019), straightlining (Schonlau & Toepoel, 2015; i.e., ≥ 15 identical ratings in a row and an average longstring of ≥ 5) and/or extremely high or low response variability (e.g., Dunn et al., 2018) were excluded, resulting in a remaining number of N = 967.

Among those participants 70.1% were male (n = 678) and 29.7% were female (n = 287), one participant was diverse and another participant indicated “other” but didn’t specify. Most of the participants were between 25 and 26 years old (n = 259; 26.8 %) and the overall mean value was 25.2 (min = 18, max = 34, SD = 3.42).

The study includes only people in their youth (from 18th to 34th year of life). This results in the fact that this group has the most amount of free time and therefore plays more video games than other age groups: Statistics published by the entertainment software association show that people aged from 18 to 34 are the largest group of gamers in the US (Paaßen et al., 2017; Entertainment Software Association, 2022; Ratan et al., 2015).

Since the study was mostly aimed towards active gamers, every participant who indicated that they play less than an hour per week were excluded.

Design and materials

In order to create and conduct the online study, the online platform SoSci Survey was used. The first page of the survey consisted of a brief introduction and a declaration that participation is voluntary, and data will be treated anonymously. The survey contained items on basic demographics (e.g., age, gender), gaming habits/preferences, gaming motives (MOPS) and mental well-being (PERMA-Profiler and WHO-5). The survey took 10 to 15 minutes and was conducted in English.

Inspired by the paper 'Gaming well: links between videogames and flourishing mental health' (Jones et al., 2014), this study incorporated the five elements of PERMA (Positive Emotion, Engagement, Relationship, Meaning and Achievement) to test for correlations between video gaming and 'flourishing mental well-being' (Jones et al., 2014, p. 1). With that intent, the PERMA-Profiler by Butler and Kern (Butler et al., 2016), a brief measure of PERMA, was also integrated in the study's questionnaire. The PERMA Profiler consists of 15 questions (3 items per PERMA domain). The response possibilities range from 0, never, to 10, always (Butler et al., 2016). The overall internal consistency of the PERMA-Profiler, measured in this study, is excellent, Cronbach's alpha lies at $\boldsymbol{\alpha}$ = .94 (Positive emotion*:* α = .77; Engagement: α = .76; Relationships: α = .79; Meaning: α = .78; Accomplishment: α = .76).

Furthermore, this online study included the World Health Organisation's Well Being Index (WHO-5, 1998 version) which is composed of five items (example: 'I have felt cheerful and in good spirits.'). The internal consistency of the WHO-5, measured in this study is excellent with Cronbach's alpha α = .89. The five statements measure well-being over the past two weeks with a scale from 5, all the time, to 1, at no time.

In addition, the MOPS, Motivation to Play Scale (Holl et al., 2021), was included in the online study. The MOPS assesses gaming motivation on a comprehensive and fully validated scale. It comprises 58 items with a 10- factor structure (creativity/exploration, escapism, competition, prestige, enjoyment, achievement, socializing, boredom, aggression, and skill). The MOPS' items are based on a systematic bottom-up literature search. The MOPS allows to divide video game players into four player types: 'Casual Player', who plays casually and enjoys games and does not play primarily for achievement or competition (lowest means for all gaming motives, but high on boredom and enjoyment motivation), 'High Performer', who wants to achieve things, compete and show their skills (high motivation for aggression, prestige, and competition, but lowest motivation for enjoyment), 'Crafter', who enjoys to create things and to achieve things by building and crafting (low aggression, prestige, and escapism motivation, but high in achievement, creativity and enjoyment motivation) and 'Highly involved player', who is fully focused on stories and achievements (highest overall motivation and highest motivation for escapism). The items are rated on a five-point likert scale from 1, strongly disagree to 5, strongly agree. The internal consistency of the MOPS, measured in this study, is excellent $\boldsymbol{\alpha}$ = .95.

Procedure

The participants could access the online survey by clicking on a specific link that was provided. The collected data was on SoSci Survey, but later on the data was extracted to SPSS. The survey can be divided into three parts.

On the first two pages of the questionnaire, participants gave informed consent.

The first part is evolving around some general demographics. First of all, the participants were asked to give us their consent and allow the use and publication of the data gained from this questionnaire for the Expra. In the next the age of the participants was asked as well as their gender (male/female/divers/other). Furthermore, the participants were asked if they considered themselves as active gamers.

The second part about the questionnaire is all about video games. It starts by asking the participants how many hours a week they play on average and on what device (PC/Console/VR/Mobile/other). Then the participants were asked to rank the

status of video games in contrast to other hobbies and leisure activities (e.g sports) from least to most important. Then 10 genres of games were presented, defined by HP®[2], to the participants and they needed to choose the three genres they play most and rank them from most to least played.

Moving on to the third and last part of the questionnaire which is about the participants' well-being. In the last part, participants completed the WHO-5 items and the PERMA profiler. In addition, the participants were also asked to fill out the MOPS questionnaire which consists of 58 items.

At the end of the questionnaire, participants were thanked and were able to take part in a raffle of one of ten 20€ coupons by providing a valid email address (stored separately from all other data).

After getting the results from the survey, they were analyzed to prove or to falsify the presented hypotheses as well as answer the given research questions. To achieve this goal, several calculations were made such as correlations, most of them are linear regressions, and doing a comparison between the results and the corresponding literature.

Descriptive statistics

ACTIVE GAMER

A total of 96.5 % of the participants described themselves as active gamers (*n* = 933).

HOURS PER WEEK PLAYING VIDEO GAMES

The lowest amount of hours played per week was 0.4 and the highest amount was 75 hours.
Min = 0.4, *Max* = 75, *M* = 14.55, *SD* = 12.9

PLAYER TYPES

Among all the participants, 273 identified themselves as Casual Players (28.2 %), 399 participants considered themselves as High Performers (41.3 %), 172 regarded themselves as Crafters (17.8 %) and 123 participants self-identified as Highly involved Players (12.7 %).

PLAYING DEVICE

A total of 52.7 % of the participants mostly play on PC (*n* = 510), 28.7 % use their mobile (*n* = 278), 12.3 % use a console (*n* = 119), and 6.2 % the VR (*n* = 60).

GAME GENRE PREFERENCE

Multiplayer online battle arena (MOBA) with 38.9% and Role-Playing with 34.9 % (RPG, ARPG and More) were the two most distributed video game genres in the Top 3 ranking.

Results

To evaluate if the well-being is greater if more columns of the PERMA model are fulfilled, a bivariate Pearson Correlation was conducted. There was a significant moderate positive correlation between the amount of PERMA columns fulfilled and the well-being, $p < .001$, $r = .45$.

To test if the gamer type groups differ in terms of the 5 columns of the PERMA model, a multivariate analysis of variance was conducted. Using Pillai's trace, there was a significant effect of the 5 columns of the PERMA model on the 4 gamer types, $V = 0.8$, $F(15,2883) = 5.30$, $p < .001$. The pairwise comparison showed that High Performers have significantly higher scores in the column accomplishment then Casual Player. Additionally, an ANOVA was calculated with the four player types as independent variable and the sum score of the PERMA profiler as dependent variable. There was a significant group difference for the player types, $F(3, 963) = 19.92$, $p < .001$, $\eta_p^2 = .06$. The means are displayed in figure 1.

[2] https://www.hp.com/us-en/shop/tech-takes/video-game-genres

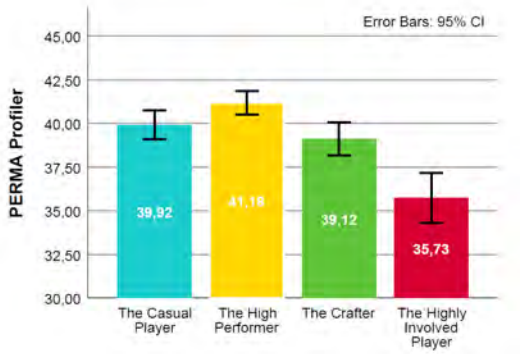


Figure 1: Mean values of the individual player types for the overall PERMA Profiler

To analyze each gamer group's correlation with well-being depending on the number of pillars of the PERMA model they fulfill, a filter was applied to test each group individually and a bivariate Pearson Correlation was used.

THE HIGH PERFORMER'S well-being has a low significant correlation across all the pillars of the PERMA model. The strongest correlation was shown on the pillars Accomplishment ($r = .27$, $p < .01$) and Meaning ($r = .29$, $p < .01$). The weakest correlation was shown on the pillar Positive Emotion ($r = .24$, $p < .01$).

THE CASUAL PLAYER'S well-being has a strong correlation across all the pillars of the PERMA model. The strongest correlation was shown on the pillar Positive Emotion ($r = .62$, $p < .01$).The weakest correlation was shown on the pillar Positive Relationships ($r = .56$, $p < .01$).

THE CRAFTER'S well-being has a strong correlation across most of the pillars of the PERMA model. The strongest correlation was shown on the pillars Engagement ($r = .55$, $p < .01$) and Meaning ($r = .56$, $p < .01$). The weakest correlation was shown on the pillar Accomplishment ($r = .49$, $p < .01$) with the only moderate correlation.

THE HIGHLY INVOLVED PLAYER'S well-being has a moderate correlation across all the pillars of the PERMA model. The strongest correlation was shown on the pillar Positive Emotion ($r = .45$, $p < .01$). The weakest correlation was shown on the pillar Positive Relationships ($r = .38$, $p < .01$).

To test if men play more video games than women, a one-way ANOVA was calculated with gender as independent variable and average hours spent playing video games per week as dependent variable. As Levene's test of homogeneity of variances was significant ($p = .007$), Welch's Test was used. There was no significant difference between men ($M = 14.28$; $SD = 12.39$) and women ($M = 15.18$; $SD = 14.03$) for the amount of video game play per week, $W(1, 483.65) = 0.90$, $p = 3.43$, $\eta^2 < .01$.

To measure how men and women differ in game preference, a frequency analysis was conducted to determine 3 most preferred game genres. For men, the most featured game genre in their top 3 were the Multiplayer Online Battle Arena (MOBA) genre with 37%, the Role-Playing Games (RPG) and Action Role-Playing Games (ARPG) with 35.1%, and Shooters, like First-Person Shooters (FPS) and Third-Person Shooter, with 33.8%. For women however, the 3 most featured game genres in their top 3 were MOBA's with 43.6%, Sandbox games with 32.4% and RPG's with 31,7%. To further investigate differences between men and women in game preference, the frequency analysis was further used to determine the game genre mostly put on first place. For men Shooters were most featured in first place (12.7%). For women RPG's were most featured in first place (14.3%).

To prove the effects of play time on well-being, a bivariate Pearson correlation was conducted for each group of play time with well-being. The low play time group ($M = 2.92$; $SD = 1.60$) has a significant weak negative correlation with well-being: $r = -.12$, $p < .05$. The moderate play time group ($M = 8.89$; $SD = 1.33$) registered a significant weak positive correlation with well-being: $r = .29$, $p < .01$. The high playtime group ($M = 24.73$; $SD = 10.84$) has a weak negative correlation with well-being: $r = -.08$, $p = .06$.

To test if higher video game status is associated with more time spent play-

ing video games per week, a bivariate Spearman correlation was calculated. There was a significant weak positive correlation between video game status and time spent playing video games per week, $r_s = .07$, $p < .05$.

To test, if video game status moderates the relationship between time spent playing video games per week and well-being, a moderated regression analysis was calculated. There was a significant interaction between video game status and time spent playing video games on the outcome of well-being, $b = 0.01$, $SE < 0.01$, $p = .040$, 95% CI [<0.01, 0.02]. However, there was no significant effect of time spent playing video games per week on well-being for any rank of video status, ($ps \geq .072$).

Discussion

Video games have had a reputation in research for having a mostly negative effect on the affective and mental state of players (Przybylski & Weinstein, 2016). Consequently, the aim of this study was to create a balance of perspectives in the scientific field, focusing on the existing positive aspects and influences of video games, and to contribute to this field of research. In particular, this study focused on the positive influence of video games on well-being.

To capture mental well-being, Seligman's PERMA model and the WHO-5 were used. Additionally, the MOPS scale (4 gamer types), play time per week, video game status and gender differences were utilized in order to identify which aspects lead to a positive effect, regarding the correlation between the independent variable video games and the dependent variable mental well-being.

It was identified in this study that the four player types differentiated compared to the 5 pillars of the PERMA model. The comparison showed that High Performers have the highest scores in the overall columns ($M = 41.18$) (see Figure 1). However, it should be noted that the High Performer, unlike the other player types, has a low significant correlation across all the pillars of the PERMA model. Which means that High Performers do not show a clear preference regarding the 5 PERMA pillars. Moreover, the group allocation was a self-evaluation which limits the generalization and could be related to the preference of the pillars. Nevertheless, it has been found that well-being is greater when more pillars of the PERMA model are fulfilled and, accordingly, concluding that High Performers represent the gamer type group that exhibits the highest well-being.

In terms of gender differences, no significant differences were found in terms of playing time per week. This means that women and men do not differ in their playing time per week. This corresponds to the opposite of what was assumed, thus Hamlen's (2010) findings to gender differences were disproved. These results may be due to the fact that Hamlen's (2010) study is outdated. The Entertainment Software Association analyzed video game players and found that in 2022, 48% of the players in the U.S. identified themselves as female. In addition, the video game industry is in constant change and growth, resulting in more games being produced for various target groups, increasing interest in video games and playtime regardless of gender (Paaßen et al., 2017).
In terms of video game preference, some gender differences could be identified. Two out of three game genres were equally listed in the top 3 for both genders. These are the genres: Multiplayer Online Battle Arena and Role-Playing Games, which due to their gameplay appeal most to the player types High Performer and Highly Involved Player. Given that the majority of the participants classified themselves as High Performer and Highly Involved Players (54%) support these results.

However, this does not seem to be consistent with the existing literature in this area, as these genres have been proven to be heavily male-dominated (Paaßen et al., 2017; Entertainment Software Association, 2022; Ratan et al., 2015). Further research would be needed to determine and substantiate the gender preference of these particular genres. Shooters on the other hand are aimed at the male audience and owned the 1st place. The violent video games were thus as suspected and proven by other studies the most preferred genre of men. The violent video games were thus as suspected and proven by other studies the most preferred genre of men. Women, on the contrary, had the Sandbox games, in 3rd place. These results are consistent with previous analyses, in which especially the genre preferences regarding shooters for men and build-up/casual games for women stood out (Paaßen et al., 2017; Phan et al., 2012). Those results imply that video game preferences are influenced by learned stereotypes regarding learned gender roles and learned gendered media preferences (Wühr et al., 2017; Melzer, 2019).

It was significantly proven that moderate play time (7-10 hours per week) of video games has a positive effect on mental well-being, and low and higher playing time have a negative influence. These results confirmed the assumptions and showed similar findings as previous studies (Jones et al., 2014). Accordingly, Jones et al. (2014) found that this moderate amount of play time was associated with all the positive effects of videogames, and excessive play or no play at all reduced or negated these benefits.
Furthermore, it was found that the higher the viewing status of video games in life, the higher the individual playing time per week. Finally, there was a significant interaction between the two variables and well-being on the outcome of well-being. However, there was no significant effect of time spent playing video games per week on well-being for any rank of video status. These results thus confirm the previous results on playing time and well-being, that more playing time leads to a negative influence of well-being. The higher the status, the higher the playing time, which has a negative effect on well-being. Likewise, these results can be linked to addictive behavior to video games, where high status encourages higher play time, which has a negative impact on well-being (Rosenberg, 2020; Bean et al., 2017).

Limitations

Since the study was an online survey, no causal interpretations were possible, only correlational interpretations of the data were allowed and possible.

Another major source of uncertainty is the self-categorization from the Motives to Play Scale (MOPS). The participants were asked to categorize themselves in one of the four given player types: ‘Casual Player’, ‘High Performer’, ‘Crafter’ and ‘Highly involved player’. This could lead to a wrong categorization and to other results. The best possible approach would have been to assign each participant to one of the 4 groups based on the results of the MOPS. This was, however, beyond the scope of the present work but would be valuable to assess in future projects.

Furthermore, the four given player types: ‘Casual Player’, ‘High Performer’, ‘Crafter’ and ‘Highly involved player’, were described with a short sentence referring to each player type. They could have been described more precisely in the online survey. This could have an influence on the self-categorization and on the high number of the ‘High Performer’ match.

Another limitation refers to the 10 genres of gamers that were defined by HP®. The participants needed to choose the three genres they play most, and they had to rank them from most to least played. In fact, the 10 gamer genres given by HP® are not a scientific reference and so the validity

could be criticized. The use of these genres is justified by the fact that the scientific classifications are outdated and in no way address today's novel game genres. Thus, categorisations like Wolf''s (2001) only address traditional genres and do not include online games.

Conclusion and outlook

To conclude, almost all the study's hypotheses were confirmed and followed what other studies had shown. The positive correlations between well-being and video games were all confirmed. Only the gender differences were not found as expected, in terms of playing time and in terms of genre preference they were not evident. This suggests that there may have been or may be a turnaround in which gender differences and gender roles in video games no longer play the role they once did. Accordingly, future studies can focus on this aspect and analyze it further. In the future, it might also be interesting to use the Motives to Play Scale (Holl et al., 2022) to categorize participants into player types and further calculate the correlations between the player types, Seligman's PERMA Model (2011) and well-being. Likewise, these positive findings on video games will encourage further studies to create a more positive image on video games.

References

Anderson, C. A. , Shibuya, A. , Ihori, N. , Swing, E. L. , Bushman, B. J. , Sakamoto, A. , Rothstein, H. R. & Saleem, M. (2010). Violent Video Game Effects on Aggression, Empathy, and Prosocial Behavior in Eastern and Western Countries. *Psychological Bulletin, 136* (2), 151-173.

Bean, A. M., Nielsen, R. K. L., van Rooij, A. J., & Ferguson, C. J. (2017). Video Game Addiction: The Push to Pathologize Video Games. *Professional Psychology: Research and Practice*, *48*(5), 378–389.

Butler, J. & Kern, M. (2016). The PERMA-Profiler: A Brief Multidimensional Measure of Flourishing. *International Journal of Well-being, 6*(3), 1–48.

Calvert, S. L., Appelbaum, M., Dodge, K. A., Graham, S., Nagayama Hall, G. C., Hamby, S., Fasig-Caldwell, L. G., Citkowicz, M., Galloway, D. P., & Hedges, L. V. (2017). The American Psychological Association Task Force Assessment of Violent Video Games: Science in the Service of Public Interest. *American Psychologist*, *72*(2), 126–143.

Dunn, A. M., Heggestad, E. D., Shanock, L. R., & Theilgard, N. (2018). Intra-individual Response Variability as an Indicator of Insufficient Effort Responding: Comparison to Other Indicators and Relationships with Individual Differences. *Journal of Business and Psychology*, *33*(1), 105–121.

Entertainment Software Association (2022). *Essential Facts about the Video Game Industry.* Retrieved from https://www.theesa.com/wp-content/uploads/2022/06/2022-Essential-Facts-About-the-Video-Game-Industry.pdf

Ferguson, C. J. (2015). Clinicians' Attitudes toward Video Games vary as a Function of Age, Gender and Negative Beliefs about Youth: A Sociology of Media Research Approach. *Computers in Human Behavior, 52,* 379–386.

Granic, I., Lobel, A., Engels, R.C. (2014). The Benefits of Playing Video Games. *The American Psychologist, 69*(1), 66-78.

Hamlen, K. R. (2010). Re-Examining Gender Differences in Video Game Play: Time Spent and Feelings of Success. *Journal of Educational Computing Research, 43*(3), 293–308.

Holl, E., Sischka, P., Wagener, G. L., & Melzer, A. (2022, September). Development and Validation of the Motivation to Play Scale

(MOPS). *Poster presented at the 52nd Congress of the German Psychological Society (DGPs)*, Hildesheim.

HP® Tech Takes (2020). *Video Game Genres Everything You Need To Know,* Retrieved 21 November 2022, from https://www.hp.com/us-en/shop/tech-takes/video-game-genres

Jones, C., Scholes, L., Johnson, D., Katsikitis, M., & and Carras, M. (2014). Gaming Well: Links between Videogames and Flourishing Mental Health. *Frontiers in Psychology, 5*:260.

Leiner, D. J. (2019). Too Fast, Too Straight, Too Weird: Non-Reactive Indicators for Meaningless Data in Internet Surveys. *Survey Research Methods*, *13*(3), 229–248.

Melzer, A. (2019). Of Princesses, Paladins, and Player Motivations: Gender Stereotypes and Gendered Perceptions in Video Games. In D. Pietschmann, B. Liebold, B. Lange & J. Breuer (eds.), *Digital Hunter-gatherers: An Evolutionary Psychology Approach to Digital Games* (pp. 205-220). London: Routledge.

Paaßen, B., Morgenroth, T., & Stratemeyer, M. (2017). What is a True Gamer? The Male Gamer Stereotype and the Marginalization of Women in Video Game Culture. *Sex Roles*, *76*(7–8), 421–435.

Phan, M. H., Jardina, J. R., Hoyle, S., & Chaparro, B. S. (2012). Examining the Role of Gender in Video Game Usage, Preference, and Behavior. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56*(1), 1496–1500.

Przybylski, A. K., & Weinstein, N. (2016). How We See Electronic Games. *PeerJ*, *4*, e1931.

Ratan, R. A., Taylor, N., Hogan, J., Kennedy, T., & Williams, D. (2015). Stand by Your Man: An Examination of Gender Disparity in League of Legends. *Games and Culture*, *10*(5), 438–462.

Rosenberg*, K.* (2020). Is Video Game Addiction Real? *AJN, American Journal of Nursing, 120*(8), 48–48.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion*, *30*(4), 344–360.

Schonlau, M., & Toepoel, V. (2015). Straightlining in Web Survey Panels over Time. *Survey Research Methods, 9*(2), 125–137.

Seligman, M. E. P. (2011). Flourish: A Visionary New Understanding of Happiness and Well-being. *Free Press.*

Wolf, M. J. P., & Baer, R. H. (2001). Genre and the Video Game. In M. J. P. Wolf (Ed.), *The Medium of the Video Game* (pp. 113–134). University of Texas Press.

World Health Organization. (1998). *Well-Being Measures in Primary Health Care: The DepCare Project.* WHO Regional Office for Europe, Copenhagen.

Wühr, P., Lange, B. P., & Schwarz, S. (2017). Tears or Fears? Comparing Gender Stereotypes about Movie Preferences to Actual Preferences. *Frontiers in Psychology*, *8*.

Gaming & Emotions 2: Testing Mood Repair Effects of Meaningful Video Games

Lili Amphiarus, Sarah Kimmlinger, Lisa Weber, Carmen Margot Elisabeth Zinnen

Supervisor: M.Sc. Gary L. Wagener

Prior research examined casual video game effects on players' mood and well-being, and their potential for therapeutic interventions.
There has been limited research conducted on meaningful video games, such as Gris, which offer non-hedonic, eudaimonic content, and thus may be better suited for mood management. This study investigated if playing Gris could have a positive effect on players' mood, alleviate sad mood, and if this is related to the extent of depressive symptoms. Additionally, physiological measurements were taken to explore a physiological relaxation effect.
In a lab experiment (*N* = 49), participants played Gris after they underwent either a sad- ness or a neutral autobiographical recall task. At different time points during the study, participants provided self-reported data for depressive symptoms and subjective mood levels. Additionally, heart rate variability data was collected as a measure for physiological effects throughout the study. Results showed that even though there was no mood improvement from playing Gris, there was a significant mood repair effect for people in the sadness condition. Additionally, playing Gris provided participants with physiological relaxation (i.e., an increase in heart rate variability), independent of condition.
Findings are discussed considering the additional purpose of video games in the clinical and therapeutic domain for mood management and overall well-being.

Introduction & Theoretical Background

As of 2022, around 3.2 out of approxi- mately 8 billion people worldwide played video games. In Germany, approxi- mately half of the population engages in this behavior (Newzoo Global Games Market Report, 2022). Playing video games has therefore long since ceased to be a marginal phenomenon and is now an essential entertainment tool in our modern, increasingly fast-paced everyday lives. Furthermore, half of all gamers are female (48%), and the aver- age is 34.7 years. Not only younger generations play video games, but 10% of all gamers are between 60 and 69 years old (Tenzer, 2023). Due to its overgrowing popularity among the broad population, it seems evident that re- search should investigate what effects video games have on their players. In the last decade, research in the field of media psychology has shown that playing video games appears to be effective for decreasing negative emotions like stress (Weber et al., 2020), sadness (Rieger et al., 2015), and frustration (Rieger et al., 2014). Moreover, re- search indicates that playing video games may enhance emotional intelli- gence and the mastering of emotional regulation strategies, as well as help with recovering from negative emotional states (Villani et al., 2018). They can also be used for more active mood re- pair, which means distraction from neg- active moods and the modification of un- pleasant emotional states, especially in negative states, such as depression (Rieger et al., 2015). In addition, it was shown that video games applied in a therapeutic context can reduce symp- toms of depression and

anxiety, as well as decrease negative affect in depres- sive mood, by promoting enjoyment, flow states, and motivation (Kowal et al., 2021; Pine et al., 2020). Furthermore, entertainment experiences are often conceptualized in terms of hedonic pleasure, a positive affect elicited by me- dia content (Oliver et al., 2011). However, as shown by Reinecke (2016) and Rieger et al., (2014), playing video games is not only based on hedonistic reasons, but is also essentially used for mood management and for eudaimonic purposes, referring to feelings of mean- ingfulness and the experience of moral and intellectual virtues.

Up to this point, in the context of mood improvement, repair, and potential use in therapeutic contexts, research has been limited to casual video games, and little to no research has been conducted on meaningful video games. A study by Reinecke (2016) concluded that media selection is a function of the media us- ers' current affective state and follows the principle of mood optimization (mood management theory). These findings are extended by the results of Rieger et al., (2014) who found that both hedonis- tic and eudaimonic media entertainment experiences are associated with psy- chological detachment and relaxation. Specifically, eudaimonic media enter- tainment experiences are associated with mastery experiences (Rieger et al., 2014). Therefore, it is worth considering further research specifically on mean- ingful video games to examine the ef- fects on mood repair, improvement, and physiological relaxation to potentially
find a better alternative for an indication tool in therapeutic contexts. Meaningful video games such as Gris are character- ized by more complex experiences, such as feeling inspired, touched, or moved, which has been described as "meaningful affect" (Oliver et al., 2012), and engagement with deeper themes. In addition, research on role-playing video games emphasizes that they promote self-identification and can be used as an effective therapeutic tool (Reinecke et al., 2012). Furthermore, they can lead to changes in entrenched thought patterns, promote the development of positive al- ternatives, and serve as a means of challenging self-schemas in patients un- dergoing cognitive behavioral therapy (Kowal et al., 2021).
Gris offers a flow and melancholic experience, as well as relaxing discovery op- portunities and increased immersion (García Catalán et al., 2021). It deals with non-hedonic, eudaimonic content and trauma processing due to self-healing and personal development. Therefore, conducting research on the video game Gris might be able to show effects on mood improvement, repair, and relaxation.

In view of previous research, this study was designed to test if playing Gris can improve players´ mood and sadness lev- els and whether this is moderated by the extent of depressive symptoms. Addi- tionally, a physiological relaxation effect should be examined.
Based on these research questions, the following hypotheses were tested:

H1: Playing Gris improves positive mood and decreases negative mood from T1 (before gameplay) to T2 (after game- play).

H1a: The improvement in mood is greater after sadness has been induced, than it is in a neutral mood (mood repair effect).

H1b: The improvement in positive mood and decrease in negative mood are greater for people with higher DASS-21 scores.

H2: The playing of Gris has a relaxation effect: an increase in heart rate variability from t1 (baseline) to t3 (during gameplay) and to t4 (after gameplay).

H2a: For people in the sadness condition, the HRV decreases from t1 to t2 (during the autobiographical recall task) and increases from t3 to t4.

These hypotheses were tested in a lab experiment in which participants under- went an autobiographical recall task that either induced sadness or a neutral mood. Afterward, participants played the video game Gris for 20 minutes. Self-re- ported mood was assessed

before and after game play, whereas sadness and serenity levels were measured after the autobiographical recall task and before the game play. Physiological relaxation was measured at the beginning of the study, before, during, and after game play.

Methods

Sample

An a priori power analysis was conducted using G*Power (Version 3.1.9.4) (Faul et al., 2007) for sample size estimation. Results indicated a minimum sample size of N = 46 to achieve 95% power with a significance criterium of $\alpha = .05$, $\eta^2 = .25$.
Participants were recruited via flyers in the buildings of the campus Belval at the University of Luxembourg. In total, 49 participants aged between 18 and 65 years (M = 25.63; SD = 8.80) took part in the present study, with 55.1% being female (N = 27) and 44.9% being male (N = 22). Additionally, participants reported the average time they spent playing video games per week. On average, participants played video games for 5.55 hours per week (SD = 7.16; range: 0 to 32).

Material

SELF-REPORTED SADNESS-LEVELS

Depressive symptoms – DASS 21
Depressive symptoms were measured using the short 21-item version of the Depression, Anxiety, and Stress Scale (DASS-21); (Antony et al., 1998; Nilges et al., 2021). Participants rated how many statements (e.g., “I felt that life was meaningless”) applied to them in the past week on a 4-point-scale (0 = did not apply to me at all to 3 = applied to me very much/most of the time). The DASS-21 consists of three subscales: stress, anxiety, and depression. The scale had excellent internal consistency, with Cronbach´s α ranging from .75 to .82.

Emotion Regulation – DERS-SF
Emotion regulation skills were assessed with the short version of the Difficulties in Emotion Regulation Scale (DERS-SF); (Kaufman et al., 2016). The DERS-SF has six subscales (non-acceptance, difficulties engaging in goal-directed behaviour, impulse control difficulties, lack of emotional awareness, listed access to emotion regulation strategies, lack of emotional clarity) and contains 18 items (Kaufman et al., 2016). Participants were asked to indicate how often the items, such as “I am confused about how I feel” apply to themselves, with responses ranging from 1 to 5 (1 = almost never to 5 = almost always). The scale had excellent internal consistency, with Cronbach´s α = .87.

Positive and Negative Affect – PANAS
The Positive and Negative Affect Schedule (PANAS); (Watson et al., 1988) was used to measure the positive and negative affect of participants. The test contains a 20-item mood scale with adjectives describing different sensations and feelings. Ten items each capture the dimension of Positive Affect (PA), (e.g., attentive, interested, enthusiastic, inspired) and Negative Affect (NA), (e.g., distressed, upset, scared, afraid). Respondents were asked to indicate to what extent the adjectives described their current state of mind, with responses ranging on a 5-point-scale (1 = not at all to 5 = extremely). The scale had excellent internal consistency, with Cronbach´s α ranging from .77 to 87.

Visual Analogue Scale – VAS
The VAS (Williams et al., 2010) is a psychometric scale used to quantify a person’s subjective and behavioral state (e.g., pain, sadness etc.) and its intensity. In this study, the VAS was applied to measure participants' subjective level of sadness and serenity. It consisted of two items, in form of a 10 cm line with descriptive anchors at each end, such as "not sad at all" to "extremely sad" and “not serene at all” to “extremely serene”. Participants were asked to mark a vertical line through a point on the scale that best represents their experience at the moment.

GAMING STIMULUS

Gris

Gris is a single-player jump-'n'-run ad- venture game developed by Nomada Studios and published in 2018 by De- volver Digital. In this game, the player accompanies the protagonist through the five stages of grief (denial, anger, bargaining, depression, and acceptance) based on the Kübler-Ross model (Sandra et al., 2022). Gris was chosen for this study since it is characterized as a meaningful video game that offers a non-hedonic, eudaimonic gratification, for instance, feelings of meaningfulness, contemplation, flow and melancholic experiences, as well as relaxing discovery opportunities and increased immersion (García Catalán et al., 2021). More precisely, the game presents a young girl's journey of self-healing and personal growth due to a traumatic occurrence. The game offers an experience with almost no text but is rich with colorful and delicate art, detailed animation, and elegant music scores, building greater immersion and engagement (Sandra et al., 2022). It can also be identified as a role-playing video game emphasizing self-identification (Reinecke et al., 2012). The beginning of the game was selected out of practical reasons. It pro- vides an easy introduction to the game regarding the story line and game navigation. Playing from the beginning also gave all participants the same chance to reach approximately equally far in the game, as advanced levels may be more demanding and incomprehensible. Participants played the beginning of the game in which they experienced the first two stages of grief (denial, anger).

PHYSIOLOGICAL MEASUREMENTS

Heart Rate Variability – HRV

Heart rate variability (HRV) refers to variations in heart rate over a shorter or longer measurement period in a heart- beat-to-heartbeat analysis. It is a param- eter of the autonomic function of the nervous system, which is subdivided into the sympathetic and parasympa- thetic (Löllgen, 1999). The heart is dually innervated by the autonomic nervous system, such that relative increases in sympathetic activity are associated with heart rate increases and relative in- creases in parasympathetic activity are associated with heart rate decreases. Thus, relative sympathetic increases shorten the time between heart beats (the inter-beat interval) and relative parasympathetic increases lengthen inter- beat interval (Lane et al., 2009). A healthy heart is characterized by much variability of long and short intervals. More frequent long intervals, resulting from the activation of the parasympathetic nervous system, indicate relaxa- tion. Heart rate variability was measured by disposable electrodes (Cardinal Health ECG electrodes with solid gel) at- tached to the upper body (right and left under the collarbone, under the left rib) and were connected to the BIOPAC MP36 Device. The software AcqKnowledge (Version 5.0.6) was used to view the data in real time on the com- puter. For the analysis, the raw data were first converted into Microsoft Ac- cess Table (MAT) files and then trans- ferred to PhysioData Toolbox (Version 0.6.3). ECG Signal Analyzer was used to clear raw data. The final data were an- alyzed by SPSS (Version 27).

MOOD INDUCTION PROCEDURE - MIP

Autobiographical Recall Task - AR

Sad and neutral mood were induced using the AR, which is one of the two most commonly applied methods for mood induction (Devilly et al., 2021). The task included recalling and writing for ten minutes as detailed as possible about either an event in which participants experienced sadness (sadness condition) or about the clothes they wore in their last week (neutral condition).

Procedure

The study was conducted as a single session in the MExLAB at the Belval Campus at the University of Luxem- bourg and could be conducted in two languages: German or English. First, participants were informed about the

study procedure, received an infor- mation sheet, and were asked to sign a consent form. Afterwards, participants placed three electrodes on their upper body (right and left under the collarbone and under the left rib) to record heart rate variability before, during, and after play, as well as when completing ques- tionnaires. Prior to attaching the elec- trodes, it was ensured that participants do not suffer from photosensitive epi- lepsy. While participants filled out ques- tions on a PC, baseline measurements for HRV (t1) were taken. The first part of the computer-based questionnaire in- cluded basic demographics (age, gen- der, nationality, education level), gaming habits, PANAS (T1), DASS-21 and DERS-SF. After answering these ques- tions, t1 measurements were stopped. Then, mood induction was conducted using AR task. During the 10 minutes of the AR task, the second measurements for HRV (t2) were recorded. In addition, participants in the sadness condition were informed that they could stop the task at any time if it was too stressful for them, and that they could contact UMat- ter (psychological help from the Univer- sity of Luxembourg) if they did not feel well after participating in the study. After
the task, participants filled out a first VAS (T1). Then, t2 measurements were stopped. After this, the participants re- ceived gaming instructions and game control sheets to read. Participants then played for 20 minutes on a *Nintendo Switch*, beginning with the start se- quence of the video game. While partic- ipants played the video game, the exper- imenter left the room and third measure- ments for HRV (t3) were taken. Partici- pants were alone in the room with dimmed lights to create a suitable gam- ing atmosphere. After 20 minutes, t3 measurements were stopped and the fourth and last measurements for HRV (t4) started. Participants filled out a sec- ond VAS (T2) and the second part of the computer based PANAS questionnaire (T2). This questionnaire explored the ex- perience during the game and partici- pants' emotional state. After answering these questions, the t4 measurements were stopped, and the study was over. Participants could remove the cables and attached electrodes. Before leaving the MEx-LAB, a debriefing followed, and participants were remunerated with a €10 SODEXO voucher and received study participation credits if needed.

Results

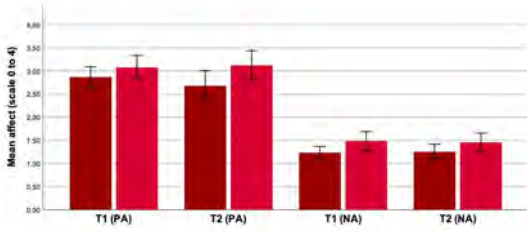
To investigate if our manipulation worked, we calculated a one-way ANOVA with the VAS score for sadness at T1 as dependent variable and condi- tion (sadness condition vs. neutral con- dition) as independent variable. Leven´s test of homogeneity of variances was significant (p < .001), and therefore Welch's test was used. Participants in the sadness condition (M = 42.90; SD = 30.15) felt significantly more sadness af- ter the autobiographical recall than par- ticipants in the neutral condition (M = 12.51; SD =18.07). Welch's $F(1, 37.34) = 18.13$, $p < .001$, $\eta = .28$.

1) Does playing Gris improve positive mood and decrease negative mood? (mood improvement effect)

Initially, a mixed ANOVA with positive affect as within-subject factor (positive affect at T1 vs. positive affect at T2) and condition as between-subjects fac- tor (sadness condition vs. neutral condi- tion) was conducted. There was no sig- nificant within-subjects effect for positive affect $F(1, 47) = .69$; $p = .410$; $\eta 2p = .02$; and no interaction affect with condi- tion $F(1,47) = 1.97$; $p = .167$; $\eta 2p = .04$. Next, a mixed ANOVA with negative affect as within-subjects factor (negative affect at T1 vs. negative affect T2) and condition as between-subjects fac- tor (sadness condition vs. neutral condi- tion) was conducted. There was no significant within-subjects effect for negative affect $F(1, 47) = .018$; $p = .893$; $\eta 2p = .01$; and no interaction affect with condi- tion $F(1,47) = 4.06$; $p = .527$; $\eta 2p = .009$. These results are shown graphically in Figure 1.

Figure 1

Comparison of mean positive and negative affect at T1 and T2



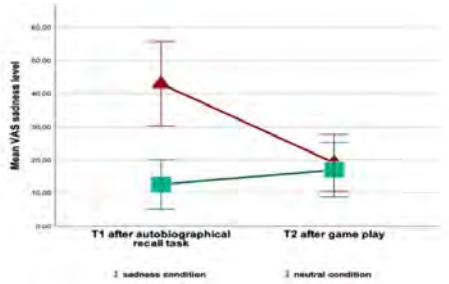
Note. error bars 95% CI

2) Can the video game Gris be used for active mood repair? (mood repair effect)

To investigate if participants in the sad- ness condition feel less sad after playing the video game Gris (mood repair ef- fect), we performed a two-step analysis. As a first step, a one-way ANOVA was conducted to further investigate whether there is a significant difference in the means of the VAS of participants' sad- ness levels at T2 between the two conditions (sadness vs. neutral condition). The results indicate that there was no significant difference in the sadness condition (*M* = 19.07; *SD* = 20.25) com- pared to the neutral condition (*M* = 16.96; *SD* = 19.95), *F*(1, 48) = 0.14, *p* = .72, η2 = .003. Participants in sadness and neutral conditions felt approximately the same level of sadness after playing Gris, which is shown in Figure 2.

Figure 2

Comparison of Mean VAS sadness level after autobiographical recall task to after game play between the sadness and neutral condition



Note. error bars 95% CI

In a second step, a mixed ANOVA with subjective sadness level measured by the VAS as within-subject factor (sad- ness levels at T1 vs. T2) and condition as between-subjects factor (sadness vs. neutral condition) was conducted. For the feeling of sadness, there was a sig- nificant within-subject effect with participants feeling significantly less sad after playing Gris (*M* = 17.99; *SD* = 19.92) compared to after the autobiographical recall task (*M* = 27.40; *SD* = 28.89), *F*(1, 47) = 13.01, *p* < .001, η2p = .22. However, the within-subject effect was qualified by a significant interaction effect *F*(1,47) = 27,67, p = .00, η2p = .37. To explore more precisely in which direction the effect exists, separate analyses were conducted and showed that for participants in the sadness condition, there was a signifi- cant decrease in sadness feeling from after the autobiographical recall task (*M* = 42.90; *SD* = 30.15) to after playing Gris (*M* = 19.07; *SD* = 20.25), *F*(1, 23) = 22.12, *p* < .001, η2p = .49, which is shown in Figure 2. However, participants in the neutral condition felt significantly sadder after playing Gris (*M* = 16.96; *SD* = 19.95) compared to after the autobiographical recall task (*M* = 12.51; *SD* = 18.07), *F*(1, 24) = 4.82, *p* = .038**,** η2p = .17, which is shown in Figure 2. Hypothesis H1a can be accepted.

3) Do depression symptoms influence the effect on players´ mood?

Firstly, to analyze if the extent of depres- sive symptoms is associated with changes in players´ mood after playing Gris, change variables for positive and negative affect, and for subjective sad- ness and serenity levels were calculated by subtracting T2 scores from T1 scores. Then, Pearson correlations were calcu- lated with these change variables and the

depression score of the DASS-21. There were no significant correlations between the DASS-21 depression score and changes in positive or negative af- fect from before to after gameplay, nor with changes in subjective sadness or serenity levels from after the autobiographical recall task to after gameplay (see Table 1). Results of Pearson corre- lation are shown in Table 1. Hypothesis H1b is, based on these results, rejected.

Table 1

Pearson Correlation between DASS-21 Depression Score and change variables

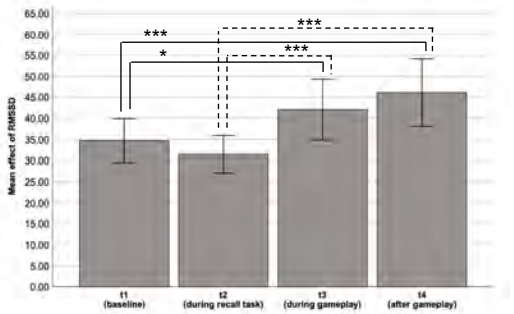
	DASS-21 Depression Score	
	r	*p*
Change in PA	-.08	.567
Change in NA	.16	.280
Change in Sadness	-.17	.235
Change in Serenity	.21	.148

4) Is there a physiological relaxation effect from playing Gris and does this variate for people in the sadness condition?

To investigate the physiological relaxation effect RMSSD was analyzed. We expected an increase in heart rate variability from t1 (baseline: before gaming) to t3 (during gaming) and to t4 (after gaming). A mixed ANOVA was conducted. There was a significant within-subjects effect for RMSSD across the four measurement points, $F(2.02, 87.09) = 17.56$, $p < .001$, $\eta^2_p = .29$. However, there was no interaction effect with condition, $F(2.03, 87.09) = 0.31$, $p = .816$, $\eta^2_p = .01$. Hypothesis H2a is rejected. Bonferroni corrected pairwise comparisons showed that RMSSD significantly increased from baseline (t1) ($M = 34.69$; $SD = 17.36$) to during gameplay (t3) ($M = 42.07$; $SD = 24.12$), $Mdiff = 7.40$, $SE = 2.47$, $p = .027$, 95% CI [0.58, 14.22] and to after gameplay (t4) ($M = 46.14$; $SD = 26.81$), $Mdiff = 11.43$, $SE = 2.43$, $p < .001$, 95% CI [4.69, 18.16], which is shown within Figure 3. Furthermore, there was also a significant increase in RMSSD from during the autobiographical recall task ($M = 31.49$; $SD = 14.71$) to during gameplay, $Mdiff = 10.59$, $SE = 2.29$, $p < .001$, 95% CI [4.26, 16.93] and to after gameplay, $Mdiff = 14.62$, $SE = 2.77$, $p < .001$, 95% CI [6.95, 22.29]. There were no significant changes in RMSSD levels from baseline to during the autobiographical recall task, $Mdiff = 3.20$, $SE = 1.36$, $p = .141$, 95% CI [-0.57, 6.96] or from during gameplay to after gameplay, $Mdiff = 4.03$, $SE = 1.96$, $p = .274$, 95% CI [-1.39, 9.44]. Based on these results hypothesis H2 is accepted.

Figure 3

Comparison of mean RMSSD across the four measurement points



NOTE. * = P < .05, *** = P < .001, ERROR BARS 95% CI

Discussion

The aim of the present study was to test mood repair effects of a meaningful video game after sadness induction. It was investigated if playing Gris could have a positive effect on players' mood and sadness levels, and whether

this is related to the extent of depressive symptoms. To sum up the results, gameplay had no effect on positive and negative affect (no mood improvement). Playing Gris did not improve players' positive mood or decrease negative mood from T1 (before playing) to T2 (after playing). However, our results showed that there was, in fact, a significant mood repair effect by playing Gris which is in line with findings by Rieger et al., (2015). In their study, a sad mood was induced in participants, who then played a casual video game, and a significant mood repair effect was found. In this study design, on the other hand, the meaningful video game Gris was played. Participants in the sadness condition played after their mood was broken and experienced a reduction in sadness subsequent to playing the game. Consequently, mood of participants in the sadness condition was restored after playing Gris. In addition, improvement in positive mood and decrease in negative mood was not greater for people with higher DASS-21 scores. There was no effect of the extent of depressive symptoms. An explanation for this finding could be that the sample of this study was not clinically relevant. There was little variance of health state in the sample, meaning that participants, on average, showed low depression scores. Furthermore, playing Gris had a physiological relaxation effect independent of the condition (sadness condition vs. neutral condition). The increase in RMSSD from before gameplay to after gameplay demonstrates that it is playing Gris that has brought a physiological relaxation effect. Prior to gameplay, no relaxation effect has been observed. Therefore, the game can be seen as the decisive point for physiological relaxation. Playing Gris, did, in fact, offer relaxing discover opportunities (García Catalán et al., 2021). There was no significant decrease in HRV from baseline to autobiographical recall task (ART), but there was an increase in HRV through video game play in both conditions. Several limitations of the present study should be noted. First, participants only played the beginning of the video game. It is conceivable that different parts could have stronger or other impacts, such as reaching the acceptance phase of Gris could lead to stronger mood repair and physiological relaxation effects. Furthermore, participants solely played for 20 minutes. Due to time, monetary, and organizational reasons, however, it was not possible to extend the game play duration. Playing from the beginning was the easiest way to provide an introduction to the game, and it gave the same chance to all participants. Then, it can also be argued that some participants, such as male participants, may be prone to a lack of identification with the female protagonist. Self-identification in role-playing video games can be described as "users' immersion with an avatar or a played role and (…) achieving goals according to rules within a virtual world" (Kowal et al., 2021). As stated earlier in this report, previous research has proposed that role-playing video games can promote self-identification, increase intrinsic motivations, and potentially be used as therapeutic tools (Reinecke et al., 2012). Self-identification effects of participants with the grieving main character in Gris were not assessed in the present study and could be investigated in future research. In addition, all participants exclusively played Gris and there was no neutral condition. Furthermore, another limitation lies in the generalizability of the present findings. Participants in the present study only played the video game Gris. Future studies should, therefore, aim to replicate the present findings with different game titles and moreover consider including a neutral condition (e.g., participants completing a puzzle).

Conclusion

Some important insights have been gained into the effects of meaningful video games on mood alteration in this study. It has been revealed that, like it is the case for casual video games (Rieger et al., 2015), playing meaningful video games like Gris can achieve a mood

repair effect as well as a physiological relaxation effect. Despite some limitations, these findings are fundamentally relevant for future research on video games, health, and well-being. It may be plausible that the effectiveness of casual video games in alleviating depression and anxiety symptoms (Rieger et al., 2015) extends to meaningful video games as well. To further explore the use of meaningful video games as a therapeutic tool, it would be necessary to conduct research on a clinically relevant sample. Further research is needed before application. Another outlook for future research would be to compare the mood repair effects of meaningful video games with casual video games. Nevertheless, the overall findings of the present study indicate that meaningful video games like Gris can have a mood repair and physiological relaxation effect.

References

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the depression anxiety stress scales in clinical groups and a community sample. *Psychological Assessment*, *10*(2), 176–181. https://doi.org/10.1037/1040-3590.10.2.176

Devilly, G. J., & O'Donohue, R. P. (2021). A video is worth a thousand thoughts: Comparing a video mood induction procedure to an autobiographical recall technique. *Australian Journal of Psychology*, *73*(4), 438–451. https://doi.org/10.1080/00049530.2021.1997553

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. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

García Catalán, S., Rodríguez Serrano, A., & Martín Núñez, M. (2021). Aprender de la caída, hacer con el desgarro: Paradojas de la melancolía lúdica en «Gris». *Artnodes*, *27*. https://doi.org/10.7238/a.v0i27.374909

Kaufman, E. A., Xia, M., Fosco, G., Yaptangco, M., Skidmore, C. R., & Crowell, S. E. (2016). The difficulties in emotion regulation scale short form (DERS-SF): Validation and replication in adolescent and adult samples. *Journal of Psychopathology and Behavioral Assessment*, *38*(3), 443–455. https://doi.org/10.1007/s10862-015-9529-3

Kowal, M., Conroy, E., Ramsbottom, N., Smithies, T., Toth, A., & Campbell, M. (2021). Gaming your mental health: A narrative review on mitigating symptoms of depression and anxiety using commercial video games. *JMIR Serious Games*, *9*(2), e26575. https://doi.org/10.2196/26575

Lane, R., Mcrae, K., Reiman, E., Chen, K., Ahern, G., & Thayer, J. (2009). Neural correlates of heart rate variability during emotion. *Neuroimage*, *44*(1), 213–222. https://doi.org/10.1016/j.neuroimage.2008.07.056

Löllgen, H. (1999). *Löllgen (1999), Herzfrequenzvariabilität*.

Newzoo. (2022). *Global Games Market Report* https://newzoo.com/games-market-reports-forecasts

Nilges, P., & Essau, C. (2021). *DASS. Depressions-Angst-Stress-Skalen—Deutschsprachige Kurzfassung*. https://doi.org/10.23668/PSYCHARCHIVES.4579

Oliver, M. B., Hartmann, T., & Woolley, J. K. (2012). Elevation in response to entertainment portrayals of moral virtue. *Human Communication Research*, *38*(3), 360–378. https://doi.org/10.1111/j.1468-2958.2012.01427.x

Oliver, M. B., & Raney, A. A. (2011). Entertainment as pleasurable and meaningful: Identifying hedonic and eudaimonic motivations for entertainment consumption. *Journal of Communication*, *61*(5), 984–

1004. https://doi.org/10.1111/j.1460-2466.2011.01585.x

Pine, R., Fleming, T., mccallum, S., & Sutcliffe, K. (2020). The effects of casual videogames on anxiety, depression, stress, and low mood: A systematic review. *Games for Health Journal*, *9*(4), 255–264. https://doi.org/10.1089/g4h.2019.0132

Reinecke, L. (2016). Mood management theory. In P. Rössler (Hrsg.), *The International Encyclopedia of Media Effects* (S. 1–13). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118783764.wbieme0085

Reinecke, L., Tamborini, R., Grizzard, M., Lewis, R., Eden, A., & David Bowman, N. (2012). Characterizing mood management as need satisfaction: The effects of intrinsic needs on selective exposure and mood repair. *Journal of Communication*, *62*(3), 437–453. https://doi.org/10.1111/j.1460-2466.2012.01649.x

Rieger, D., Frischlich, L., Wulf, T., Bente, G., & Kneer, J. (2015). Eating ghosts: The underlying mechanisms of mood repair via interactive and noninteractive media. *Psychology of Popular Media Culture*, *4*(2), 138–154. https://doi.org/10.1037/ppm0000018

Rieger, D., Reinecke, L., Frischlich, L., & Bente, G. (2014). Media entertainment and well-being-linking hedonic and eudaimonic entertainment experience to media-induced recovery and vitality: Media entertainment and well-being. *Journal of Communication*, *64*(3), 456–478. https://doi.org/10.1111/jcom.12097

Rieger, D., Wulf, T., Kneer, J., Frischlich, L., & Bente, G. (2014). The winner takes it all: The effect of in-game success and need satisfaction on mood repair and enjoyment. *Computers in Human Behavior*, *39*, 281–286. https://doi.org/10.1016/j.chb.2014.07.037

Sandra, N. P. A. A., & Mutiaz, I. R. (2022). *The Embodiment of Kübler-Ross Model through Game Elements in GRIS and Its Impact on Player Experience:* ICON ARCCADE 2021: The 2nd International Conference on Art, Craft, Culture and Design (ICON-ARCCADE 2021), Bandung, Indonesia. https://doi.org/10.2991/assehr.k.211228.033

Tenzer, F. (2023). *Computerspieler—Anzahl in Deutschland bis 2020*. [Data set]. Statista. https://de.statista.com/statistik/daten/studie/712928/umfrage/anzahl-der-computerspieler-in-deutschland/

Villani, D., Carissoli, C., Triberti, S., Marchetti, A., Gilli, G., & Riva, G. (2018). Videogames for emotion regulation: A systematic review. *Games for Health Journal*, *7*(2), 85–99. https://doi.org/10.1089/g4h.2017.0108

Watson, D., Anna, L., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology,* 54(6), 1063-1070.

Weber, S., Mast, F. W., & Weibel, D. (2020). Experiencing presence in a gaming activity improves mood after a negative mood induction. *International Journal of Gaming and Computer-Mediated Simulations*, *12*(4), 22. https://doi.org/10.4018/JGCMS.2020100101

Williams, V. S., Morlock, R. J., & Feltner, D. (2010). Psychometric evaluation of a visual analog scale for the assessment of anxiety. *Health and Quality of Life Outcomes, 57*(8). https://www.hqlo.com/content/8/1/57

The impact of language during an emotional eliciting situation on the cognitive performance of young autistic people

Arian Afroze Al Kabir, Tina Büchler, Diana dos Santos Trindade, Irena Hildebrand

Supervision: Dr. Andreia Costa, M. Sc. Maïte Franco

The aim of our study is to investigate the difference in cognitive performance in neurotypical adolescents vs. autistic adolescents during an emotional eliciting situation using their first language (L1) vs. second language (L2). Language affects the way we think, make decisions, and other cognitive processes (Perlovsky, 2009), while L2 can be used to distance ourselves from those intense emotions (Zentella, 1997). Autistic adolescents have more emotional difficulties than neurotypicals (Sofronoff et al., 2007) and the emotional difficulties are linked to lower performance in cognitive processes (Richards, 2004). We recruited 12 autistic adolescents, who were between 10- and 16- years old verbal and without intellectual disability. We recruited 12 neurotypical adolescents for the control group. One relative was recruited for each child as well. The study consisted of two parts. In the first part, the relatives were asked to fill out the following questionnaires, about the adolescent: AQC-P, ERSSQ, SRS-2 and LHQ-3. In the second part, the adolescent completed the RPM and was placed in an emotional eliciting situation through Cyberball. In the next step, the adolescent answered to questions, which were translated into seven languages (French, English, German, Luxembourgish, Spanish, Russian, Portuguese), about the Cyberball situation in either their L1 or their L2, which was assigned randomly. Lastly, the adolescent watched an educational video and was quizzed on it afterwards (memory task). Autistic adolescents had a higher ($M = 13.75$, $SD = 2.56$) memory task score than neurotypicals ($M = 10.58$, $SD = 2.50$), $t(22) = 3.06$, $p<.05$. Adolescents in the L1 condition had a higher ($M = 13.42$, $SD = 2.47$) memory task score than adolescents in the L2 condition ($M = 10.92$, $SD = 2.97$), $t(22) = 2.84$, $p<.05$. These findings highlight the significance of languages after the emotional eliciting situation on cognitive performance, further research should be focused on the complex interaction between emotion regulation and cognitive processes. These findings show that language condition and group condition show a significant effect on cognitive performance.

Introduction

The academic performance of autistic individuals is an area of growing interest and concern (Dijkhuis et al., 2020). While some autistic individuals demonstrate exceptional cognitive abilities, many still face significant challenges in the academic performance (Uddin 2022, Dijkhuis et al., 2020). Understanding the factors that contribute to these difficulties is crucial for developing effective interventions and support systems.

The aim of this study is to explore possible aspects that may impact the academic performance of autistic children, like the executive function skills. Executive function refers to a set of cognitive processes that regulate and control one's thoughts, actions, and emotions (Diamond, 2014). This plays an important role in various academic tasks, such as attention, working memory, planning and self-regulation (Diamond, 2014).

Despite the importance of executive skills in academic performance, little is known about the specific role and how they relate to the academic challenges experienced by autistic in-

dividuals (Dijkhuis et al., 2020). By investigating the factors of autistic individuals with varying academic abilities, this study seeks to shed light on the potential contributions of function skills to their academic performance. In the following sections, we will review the different aspects, that may have an influence on the cognitive processes and academic performance.

Autism

According to the American Psychiatric Association (APA, 2013), individuals may demonstrate ongoing difficulties in each of three areas of social communication, interactions, and at least two of the four types of restricted, repetitive behaviours, ongoing deficits in social communication and social interaction. With social communication difficulties is meant, that individuals might struggle to understand social cues (APA, 2013). It may also include reduced sharing of interests, emotions, or affect, as well as difficulty initiating or responding to social interactions (APA, 2013). Autistic individuals may show difficulties in verbal and nonverbal communication, and deficits in relationships (APA, 2013). Repetitive patterns of behaviour are stereotyped or repetitive motor movements, object usage, or speech, such as repetitive hand flapping or body rocking, as well as specific behaviours like lining up toys or flipping objects (APA, 2013). It can also involve echolalia, where the individual repeats words or phrases, and the use of idiosyncratic or unique expressions (APA, 2013). Individuals may experience extreme distress when faced with even minor changes, have difficulties transitioning between activities or environments, exhibit rigid thinking patterns, engage in specific greeting rituals, or insist on following the exact same route or eating the same food every day (APA, 2013). According to the APA Dictionary of Psychology (n.d.), there is a difference between autistic traits and symptomatology. Traits are persisting personality

characteristics describing how the person behaves across different situations (APA Dictionary of Psychology, n.d.). The symptomatology refers to the combined signs or indications of psychopathology (APA Dictionary of Psychology, n.d.).

Cognitive performance

It was shown, that regarding academic achievement in autism, autistic people have a significant variability, at individual academic achievement in autism, autistic people have a significant variability, at individual level (Keen et al. 2016). But at a comparative level, they are generally not as good as neurotypicals regardless of their IQ (Ashburner et al., 2010). It was shown that autistic people are 20x more likely to be excluded from school than neurotypicals (Barnard et al. 2000). The major causes for this are behavioral and emotional problems (Barnard et al. 2000). The educational development of autistic adolescents can be influenced by their challenges in managing their emotions and behaviors, as well as their ability to stay focused and engaged in a task (Eaves & Ho, 1997). The link between cognitive abilities and emotions was researched using neurotypicals participants that demonstrated that emotional suppression was linked to impairments in cognitive processes (Davis & Levine, 2012), learning (Blair, 2002) and academic success (Graziano et al., 2007) as well as in autistic participants that showed that emotional abilities are linked to better sustained attention (Costa et al. 2017).

Emotional difficulties

It is known that some autistic adolescents often tend to struggle with emotional difficulties (Morie et al., 2019; Mazefsky et al., 2013). These emotional difficulties can include many aspects such as emotional reactivity, which means how easily and strongly someone reacts to a stimulus or a situation (Karrass et al., 2006; Schapero

et al., 2016). Autistic adolescents are more likely to experience higher emotional reactivity which means they can feel stronger reactions compared to neurotypical peers (Beck et al., 2020; Conner et al., 2021). If a change in a routine or a social interaction occurs, they are more likely to feel more overwhelmed or anxious than neurotypicals (Lausman, 2019). Also, alexithymia is highly prevalent in autistic adolescents and is known to be related to some emotional difficulties (Cook et al., 2013). They may struggle with understanding and decoding important components of human interaction such as facial expressions and body language (Niedenthal, 2007). This makes it difficult for them to handle social interactions and relationships (Bird & Cook, 2013). Furthermore, they may experience incongruence in their emotional expression (Mazzoni et al., 2022) because their gestures or tone of voice can be incongruent with the emotions they feel internally (Trevisan et al., 2018). This can lead to problems conveying their emotions and feelings to be understood in a social context (Brewer et al., 2016). Another difficulty that some autistic people may face is emotional regulation which is the ability to regulate and adapt your regulation depending on the environmental context (Gross & Thompson, 2007). It is a construct that refers to the process by which a person consciously or automatically adjusts their emotional state to handle situations better (Thompson, 1994). Some strategies to regulate emotions are, for example, expressive suppression, the tendency to try and ignore emotional responses, or emotional reappraisal, in which individuals change their thought patterns (Gross & John, 2003). Children with autism can face challenges in regulating their emotions effectively and can have difficulties managing emotions such as frustration or anger (Cai et al., 2018). This can lead to consequences such as emotional outbursts or meltdowns (Samson et al., 2014). Also, children with autism often experience depression, anxiety, or anger because of difficulties

with emotion regulation (Lecavalier, 2006). Further, emotional regulation is interrelated with language (Eisenberg et al., 2006).

Language

The language we speak has a big impact on individuals' lives. It affects thinking, decision, and other cognitive processes (Perlovsky, 2009). It helps individuals learn new concepts, in early life (Xu, 2002), which may enhance other cognitive processes (Lupyan, 2012). Language also influences how we perceive things due to rapid and reciprocal connections between early sensory brain areas and the orbitofrontal cortex (Lamme & Roelfsema, 2000; Pessoa & Adolphs, 2010). It simplifies the organization and use of concepts, which may be an important component of how we perceive emotions (Barrett et al., 2007; Lindquist & Gendron, 2013). So, language can play an important role in how we experience emotions (Lindquist & Barrett, 2008). There are beliefs that bilingual households confuse language learning process (Kay-Raining, 2005). However, it was shown that bilinguals may have neurocognitive advantages (Bialystock, 2001). The bilingual advantage hypothesis means that they have an advantage in cognitive functions (Gonzales-Barrero & Nadig, 2017). Autis-

tic bilinguals have an advantage in set-shifting tasks but not working memory (Gonzales-Barrero & Nadig, 2017). Intense emotions are frequently expressed in the L1 and the L is used to decrease distance (Zentella, 1997). The emotional contents of language affect cognition and foster the dual connection between language and cognition (Perlovsky, 2009). This means that the language in which children manage their emotions influences the cognitive process. Therefore, expressing emotions in an L2 might ease a distance from the situation, which reduces its cognitive impact (Zentella,1997).

The present study

This study aims to explore aspects that may impact the academic performance of autistic adolescents compared to neurotypical adolescents, focusing on an emotional eliciting situation using their L1 vs. L2. Our first hypothesis is, that autistic adolescents will score not as good on the memory task compared to neurotypical adolescents. Existing literature suggests that autistic individuals may exhibit difficulties in memory tasks (Dijkhuis et al., 2020), due to emotional regulation difficulties (Scheibe & Blanchard-Fiels, 2009). Therefore, we predict that autistic adolescents will perform worse in the memory task compared to their neurotypical peers. Our second hypothesis is bidirectional, as we assume that the use of either L1 or L2 will enhance memory task performance in adolescents. Previous research on bilingualism indicates potential cognitive benefits associated with the use of multiple languages (Gonzales-Barrero & Nadig,2017; Bialystock, 2001). Distance from the situation, which may occur when using a L2 (Zentella, 1997), has been suggested to facilitate emotion regulation and positively influence cognitive performance (Perlovsky, 2009). Based on this literature, we propose that utilizing the L1 or L2 will result in improved memory task performance in adolescents.

Methods

Participants

For this study, we recruited 24 adolescents, who were between 10 and 16 years old (M = 12.25, SD = 2.39),
verbal, and without intellectual disability. Neurotypical adolescents have been recruited as a control group, also aged between 10 and 16 years (M = 11.67, SD = 2.43). One relative was recruited for each adolescent with ages ranging between 23 and 56 years (M = 40.62, SD = 8.57).

The sample consisted of 12 participants in each group, with 2 girls (16.70%) and 10 boys (83.30%) in the autistic and 5
girls (41.70%) and 7 boys (58.30%) in the neurotypical group. Out of the 24 relatives we had in total, 20 were adult women (83.33%) consisting of 19
mothers (79.20%) and 1 sister (4.15%), and 4 adult men (16.66%) consisting of 3 fathers (12.50%) and 1 tutor (4.15%).
Among our participants, the nationalities were German, French, Luxembourgish, Portuguese, Bosnian, English and American of which 20.83% had a double nationality.
In the sample, we had 15 participants who went to a regular school without any special assistance (62.50%), 6 participants went to regular a school with personal in-class assistance (25.00%), 2 participants went to a regular school and additional private assistance (8.30%) and 1 participant went to a private school (4.20%). For the neurotypical group, we had 11 participants who went to a regular school without any special assistance (91.70%), and 1 went to a school with additional private assistance (8.30%). For the autistic group, we had
4 participants who went to a regular school without any special assistance (33.30%), 6 participants went to a regular school with per-

sonal in-class assistance (50.00%), 1 participant went to a regular school with additional private assistance (8.30%) and 1 participant went to a private school (8.30%).

In the autistic group, we had 11 (91.70%) adolescents who went to school and 1 (8.30%) adolescent who was employed. For the neurotypical group we had 12 (100.00%) adolescents, who went to school.

In the autistic group, all the participants had a diagnosis of Autistic Spectrum Disorder with or without Asperger Syndrome. Some of the participants had additionally an Attention-Deficit/Hyperactivity Disorder and/or attention deficit disorder.

The mean age at which the participants were diagnosed was M = 7.83 (SD = 4.04). 16.70% of the participants are taking regular medication.

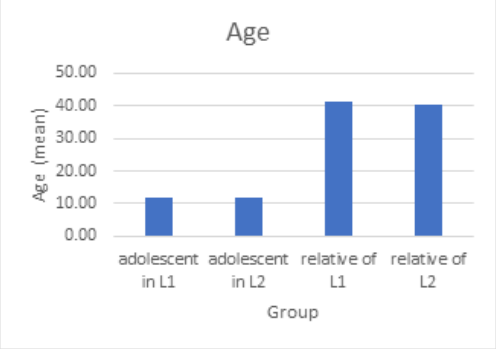
Regarding their income, 8.30% of the relatives reported that they were very satisfied, 37.50% of the relatives were satisfied with their income, 37.50% of the relatives were neutral regarding their income, 12.50% of the relatives were unsatisfied with their income and 4.20% of the relatives were very unsatisfied with their income.

Concerning the education of the relatives, 8.30% had a primary school diploma, 25.00% had a high school diploma, 25.00% had a professional diploma, 20.8% had a bachelor's degree, 12.50% had a master's degree and 8.30% had a PhD.
Regarding the exclusion criteria's, we decided to recruit adolescents between the ages of 10 and 16 who were capable of speaking multiple languages. Our focus was on verbal adolescents without intellectual disabilities. Additionally, we included for each adolescent a family member or relative.

Preliminary analyses

Figure 1

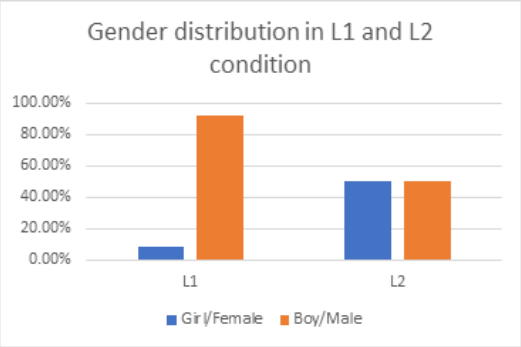
Mean age of adolescents in L1 and L2 condition



Adolescents were assigned to L1 and L2 condition randomly with each condition consisting of 12 adolescents. The average age of adolescent in L1 condition was M = 11.92 (SD = 2.67) while the average age of adolescent in L2 condition was M = 12.00 (SD = 2.17). The average age of the relative in L1 condition was M = 41.08 (SD = 8.90) while the average age of the relative in L2 condition was M = 40.17 (SD = 8.51). There were no significant differences between L1 (Mage = 11.92, SD = 2.68) and L2 (Mage = 40.17, SD = 8.51), $t(22) = .084$, $p > .05$ regarding their age.

Figure 2

Gender distribution in L1 and L2 condition

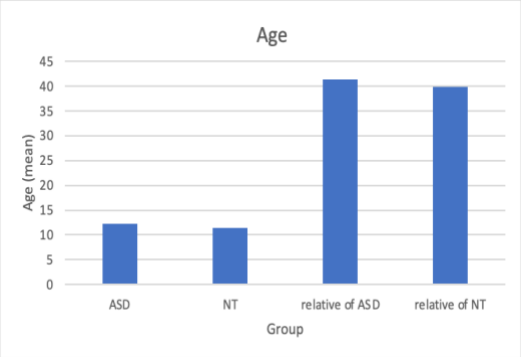


We were also interested in analysing how gender was distributed among the L1 and L2 condition using Chi-square test. We found that in L1 condition, there were significantly more boys (n = 11, 91.7%) than girls (n = 1, 8.30%)

than in L2 condition (boys: n = 6; girls: n = 6), $\chi 2(1, N = 24) = 5.04$, $p<.05$.

Figure 3

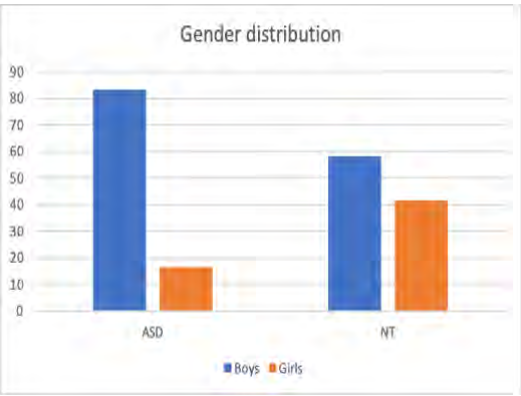
Mean age of autistic and neurotypical adolescents



We had 12 adolescents in each group. The average age of autistic adolescents was M = 12.25 (SD = 2.42) while the average age of neurotypical adolescents was M = 11.67 (SD = 2.43). The average age of the relatives of autistic adolescents was M = 41.42 (SD = 9.00) while the average age of the relatives of the neurotypical adolescents was M = 39.83 (SD = 8.43). There were no significant differences between autistic adolescents (Mage = 12.25, SD = 2.42) and neurotypical adolescents (Mage = 11.67, SD = 2.43), $t(22) = .590$, $p > .05$ regarding their age.

Figure 4

Gender distribution of autistic and neurotypical adolescents



We also analyzed how gender was distributed among autistic and neurotypical adolescents using Chi-square test. In the autistic group, we had n = 10 boys (83.30%) and n = 2 girls (16.70%). In the neurotypical group, we had n = 7 boys (58.30%) and n = 5 girls (41.70%) where we found that there was no significant difference between the groups. $\chi 2(1, N = 24) = 1.82$, $p > .05$.

DISTRIBUTION OF AUTISTIC AND NEUROTYPICAL ADOLESCENTS IN L1 AND L2 CONDITION

Regarding the language conditions, participants were randomly assigned as follows.
In the L1 condition, autistic adolescents accounted for 58.30% of the group, while neurotypical adolescents comprised 41.70%. In L2, condition, the group consisted of 41.70% of autistic adolescents and 58.30 of neurotypicals. There was no significant relationship between the group the adolescents were in and their language condition $\chi 2(1, N = 24) = .667$, $p > .05$.

MEASURING INSTRUMENTS RELATIVES' INSTRUMENTS

The survey started with the relatives'
choice of language. They could choose one language among German, French, or English. After that, they were asked to read and agree to the consent form. Next, relatives indicated the language proficiency of the adolescent. Lastly, the relatives were instructed to complete the demographics questionnaire, as well as the AQC-P, ERSSQ-P, LHQ-3, and SRS-2 questionnaires.

DEMOGRAPHICS QUESTIONNAIRE

The relatives had to answer demographic questions about themselves for example their relationship with the adolescent, their age, gender, nationality, financial resources, and educational level. The relatives were also required to answer demographic questions about the adolescent, for example, age, gender, nationality, occupation, education setting, psychological diagnosis, IQ assessment and medication.

ALEXITHYMIA QUESTIONNAIRE FOR CHILDREN (AQC-P)

The relatives completed a 20-item AQC-P questionnaire to assess the alexithymia traits in adolescents (Brown & al., 2004). AQC-P is the parent's version of the alexithymia questionnaire, although, in our study, the questionnaire was also filled out by relatives other than parents. It entails three original subscales of the TAS-20 (Bagby et al., 1994); "difficulty identifying feelings" (DIF), "difficulty describing feelings" (DDF), and "externally oriented thinking" (EOT). The AQC-P is rated on a 3-point Likert scale (1 = not true to 3 = true) (Rieffe et al., 2006). The AQC-P has an internal consistency of 0.70, supporting the reliability (Brown et al., 2021). The AQC-P score of each participant was calculated as the sum of the 20 items with a higher score indicating a greater degree of alexithymic traits (Brown et al., 2021).

EMOTION REGULATION AND SOCIAL SKILLS QUESTIONNAIRE (ERSSQ-P)

Relatives filled out the Emotion Regulation and Social Skills Questionnaire (Beaumont & Sofronoff, 2008), which is a rating scale, designed specifically to assess the social skills and emotion regulation. The ERSSQ-P includes 27-item for parents (ERSSQ-P) to rate their adolescent's behaviour on a 5-point scale, ranging from never 0 to always 4 (Beaumont et al., 2008). The ERSSQ-P has an internal consistency of 0.89 (Beaumont et al., 2008). To calculate the ERSSQ-P score of each participant, the 27 items were summed up to represent a child's social skills; higher scores indicate higher levels of social skill (Butterworth et al., 2014). Although ERSSQ-P is the parent version of the emotion regulation and social skills questionnaire, ERSSQ-P was also filled out by relatives other than parents.

LANGUAGE QUESTIONNAIRE (LHQ-3)

Relatives first had to rate the adolescents' general skills regarding learning new languages in comparison to other peers. Then the relatives had to rate on a scale from 1 = very poor to 7 = excellent the child's ability in terms of Listening, Speaking, Reading and-Writing for each language the adolescents speak. LHQ-3 is a self- report instrument used for assessing the linguistic background and language proficiency of people who speak or are learning multiple languages (Li et al., 2019, Bedore et al., 2011). For the LHQ-3 we calculated the sum of each language the adolescent spoke based on the proficiency levels indicated for the different components (Li et al., 2019).

AUTISTIC SYMPTOMATOLOGY (SRS-2)

Relatives were asked to answer to the Social Responsiveness Scale, Second Edition (SRS-2; Constantino & Gruber, 2012) to report on the adolescent's autistic symptomatology and traits. The SRS-2 is a parent-report questionnaire for a quantitative measure of autistic traits in 4- to 18- years old (Constantino & Gruber, 2012). It is composed by 65 items using a Likert scale, (0=not true to 3=almost always true). The SRS-2 has good internal consistency of 0.91 –0.97 (Constantino & Gruber, 2012). We calculated the sub-sum of the social awareness, social cognition, social communication, social motivation and autistic mannerisms. The final SRS-2 score of each participant was calculated as the sum of the sub-sums. A score below 40 indicates a high level of social reactivity, from 40 to 60 indicates a normal level of social reactivity, from 61 to 75 indicates a weak to moderate impairment of social reactivity, and a score above 76 indicates a severe impairment of social reactivity (Constantino & Gruber, 2012).

ADOLESCENTS' INSTRUMENTS RAVEN'S™ STANDARD PROGRESSIVE MATRICES

Adolescents were asked to complete a shortened version of Raven's™ Standard Progressive Matrices (RPM, Langener et al., 2022). The shortened RPM test is used for measuring cognitive ability. This non-verbal test has 15 items and has a very high degree r = 0.89 (for 9 -12 years old) and r = 0.93 (for 13 -16 years old). This test takes around 11 minutes to complete (Langener et al., 2022). The survey had two versions of the RPM. The 9 -12 version was used with adolescents aged between 10 and 12 years, and the 13 -16 version with adolescents aged between 13 and 16. The

RPM score of each participant was calculated as the sum of the 15 items. A higher score means, the stronger were the cognitive abilities of the adolescent (Langener et al., 2022).

CYBERBALL

Cyberball is a ball-tossing game designed to elicit negative emotions such as sadness, anger, or frustration, where adolescents were made to believe that they were playing with real participants (Williams & Jarvis, 2006). The game includes 2 rounds of different conditions respectively, where each round consists of 15 throws. In the first round, the game throws the ball at the same rate for each one of the participants. Our participant himself can choose to whom he wants to throw the ball. The second round of the Cyberball game was manipulated, whereas after the fourth throw, the adolescent was excluded and did no longer receive the ball. This is supposed to induce negative feelings as it is believed that those feelings become stronger if the participant is excluded from the game. Moreover, adolescents were asked to visualize the appearance and feelings of the players.

INTERVIEW

The adolescents recorded themselves answering 13 questions about their emotional experience during the Cyberball game either in L1 or L2 condition, which was chosen pseudo- randomly by the online platform. The purpose of this interview was that the adolescents had to think about those questions in the specific language. Some questions were for example:" What gender do you think were the other two players?" and "How old do you think they were?".

EDUCATIONAL VIDEO

Adolescents watched a nonverbal 6-minute and 14 seconds long educational video on how to collect water from the desert (Westdeutscher Rundfunk, 2022)

MEMORY TASK

The memory task consisted of 13 visual memory questions about the educational video. For example, in some questions they had to choose the correct pictures that they saw in the video or put them in the right order. The score of each participant regarding the memory task was calculated as the sum of the 13 questions. The higher the score was, the better the adolescent performed regarding this task.

RELIABILITY QUESTIONS

We included 4 control questions to detect the participant's reliability. One control question was included in the demographic questionnaire, one in the ERSSQ, one in the LHQ-3 and one control question was included in the memory task. The 4 questions were in the style of the following example: "What is your child's favorite colour? To show that you're still attentive, write Aquamarine."

Procedure

The study was a 50-minute online survey launched on Qualtrics © from April 10, 2023 to May 10, 2023. It comprised two sections which were available in different languages. The relatives' part was available in three languages (English, German, and French) and the part for the adolescents has been translated into seven languages: German, French, Luxembourgish, English, Portuguese, Spanish, and Russian. We contacted autism foundations, parent associations, schools, and forums in those country speaking languages within Europe such as Portugal, Spain, France, Belgium, Luxembourg, Germany, and the United Kingdom. To recruit the neurotypical control group, we sent our flyers to youth houses, day-cares, parent associations and to our friends and family. After compiling a list of contact data from all over Europe, we additionally prepared a template, for our emails and social media posts. It is also important to note, that along the emails and social media posts, flyers directed towards relatives was attached as well. The flyers have been translated in French, English and German to reach a wider audience. The online questionnaire started with an information page, where relatives were briefly informed about the purpose and the goals of the study. Subsequently, informed consent was collected for their data collection

and sharing of their anonymized data for research purposes.

The survey was split into two parts, where one part was answered by the relative and the other part was answered by the adolescent. In the first part, the relatives were asked to fill out the following questionnaires: demographics questionnaire, AQC-P, ERSSQ, SRS-2 and LHQ-3 questionnaires. In the second part, the adolescent was administered a nonverbal IQ-test, RPM to measure their cognitive ability. Adolescents were then asked to engage in the Cyberball game in L1. In the next step, the adolescents answered questions about how their emotional experience was during the Cyberball task. In this part, they answered the questions, either in their L1 or L2, which was randomly chosen. The language in which they had to answer the questions, were randomly chosen by the Qualtrics© software, while keeping an even presentation of conditions. In the last part, the adolescent had to watch a non- verbal educational video and then perform a memory task in L1 condition, consisting of 12 visual memory questions about the educational video. The participants received a compensation of 20 euros in form of a GoGift voucher of their choice, that can be redeemed on a global gift card website.

Results

Analysis

We performed several statistical analyses to address the aim of the study using the software SPSS version 27. For each statistical analysis, the significance level was set at $p < .05$. We conducted an independent t-test to confirm whether the adolescents belong to the according group, to examine analyses that are relevant for the aim of our study and the outcome measures. We also conducted a two-way ANOVA to examine the combined effects of group condition and language condition on cognitive performance. Tukey test was used as post hoc test to determine significant differences in both groups. It is important to note that the independent t-test and two-way ANOVA were used after checking assumptions of normality and homogeneity.

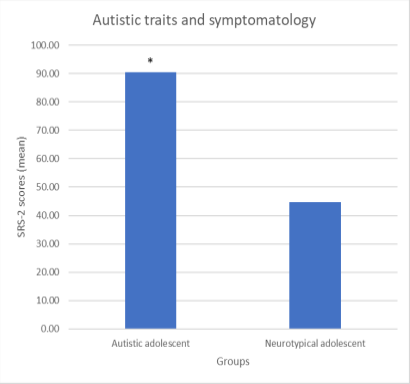
Autistic vs Neurotypical adolescents

Our study concentrates on the impact of languages during emotion eliciting situations on the cognitive performance between autistic adolescents and neurotypicals adolescents. Comparing these two groups can help to show the specific features that distinguish autistic adolescents from neurotypical adolescents, which could be an important factor in explaining our outcome results. Moreover, these findings can help identify and emphasize the characteristics and problems that people with autism deal with. In the following section, we will illustrate the differences between autistic adolescents and neurotypicals regarding autistic symptomatology and traits, cognitive ability, socio-emotional skills, alexithymia and memory task by using an independent t-test.

Autistic traits & symptomatology

Figure 1

SRS-2 scores in autistic adolescents and neurotypical adolescents



We included an SRS-2 questionnaire in the study to confirm the presence of higher autistic traits among the autistic adolescents, compared to the neurotypical adolescents. Results showed that autistic adolescents had significantly higher autistic traits ($M = 90.50$, $SD = 10.95$) than neurotypicals ($M = 44.75$, $SD = 23.19$), $t(22) = 6.197$, $p < .05$.

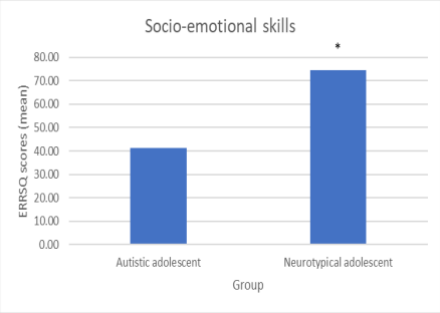
Cognitive ability

We compared the cognitive ability, which was assessed with the shortened RPM between the autistic adolescents and the neurotypicals. Results revealed that there is no significant difference between autistic adolescents (M = 10.67, SD = 3.62) and the neurotypicals (M = 10.58, SD = 4.78), t(22) = .048, p > .05.

Socio-emotional skills

Figure 2

ERSSQ scores in autistic adolescents and neurotypical adolescents

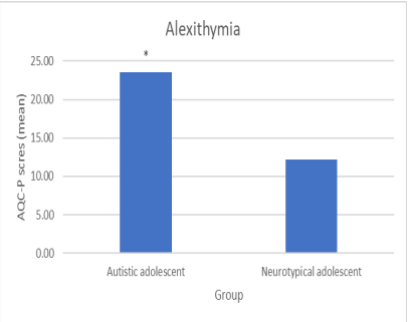


We examined whether there are significant differences in socio- emotional skills between autistic adolescents and neurotypicals. We compared the scores that were obtained in the ERSSQ questionnaire. As expected, it revealed that autistic adolescents had on average significantly less socio-emotional skills (M = 41.42, SD = 14.54) than neurotypicals (M = 74.42, SD = 16.47), t(22) = -5.203, p < .001.

Alexithymia

Figure 3

AQC-P scores in autistic adolescents and neurotypical adolescents

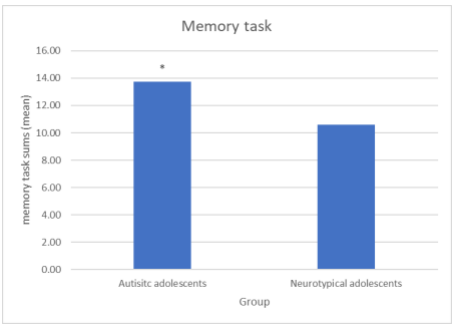


We wanted to examine the presence of alexithymia traits in autistic adolescents by comparing the means that the two groups received after filling out the AQC- P questionnaire. The results are in line with the literature, since in our sample we found that autistic adolescents had significantly more advanced alexithymia traits (M = 23.50, SD = 6.27) than neurotypicals (M = 12.17, SD = 5.41), t(22) = 4.740, p < .001.

Memory task

Figure 4

Memory task sums in autistic adolescents and neurotypical adolescents



One of the main aims of our study is to investigate whether there is a significant difference between autistic adolescents and neurotypicals in their cognitive performance using a memory task. According to our first hypothesis, we expected that neurotypicals score higher in the memory task than autistic adolescents. We compared the means of the memory task among the groups. Comparing the scores of both groups, we found that autistic adolescents scored higher (M =13.75, SD = 2.56) in the memory task than neurotypical adolescents (M = 10.58, SD = 2.50), t(22) = 3.062, p < .05.

Language 1 vs Language 2

Preliminary analyses and main analyses

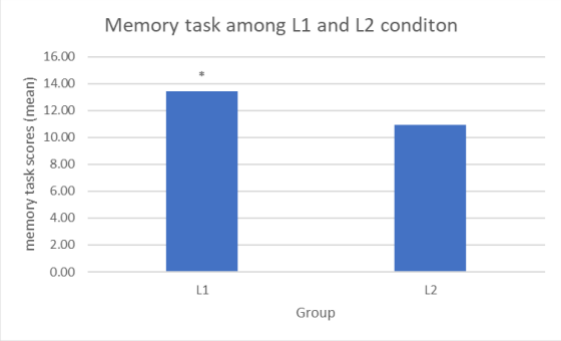
We examined the differences between L1 conditionand L2 condition regarding autistic traits and symptomatology, cognitive ability, socio-emotional skills and alexithymia using an independent t- test. Results revealed that

there were no significant differences regarding autistic traits and symptomatology between L1 (M = 69.31, SD = 31.73) and L2 (M = 65.64, SD = 27.62), t(22) = 0.299, p > .05, cognitive ability between L1 (M = 2.85, SD = .69) and L2 (M = 2.45, SD = .82), t(22) = 1.272, p > .05, socio-emotional skills between L1 (M = 53.46, SD = 25.76) and L2 (M = 63.18, SD = 18.22), t(22) = -1.048, p > .05 and alexithymia between L1 (M = 17.25, SD = 8.84) and L2 (M = 18.42, SD = 7.73), t(22) = -.344, p > .05.

Memory task

Figure 5

Memory task scores in L1 and L2 condition



One of our main aims of our study is to investigate whether there is a difference between speaking the L1 vs L2 about an emotion-eliciting situation on the subsequent performance on a cognitive task. For our second hypothesis, we took a bi-directional approach by assuming that adolescents score higher in the memory task, after using either the L1 or L2 during an emotion eliciting situation. We compared the cognitive performance of the memory task in participants who were in L1 condition to participants in L2 condition. Results showed adolescents had higher memory task scores when they spoke about the emotional situation in their L1 (M=13.42, SD=2.47) than when they spoke in their L2 (M = 10.92, SD = 2.97) t (22) = 2.244,
p < .05.

Effect of groupand language condition on memory task

Table 1

Tests of Between-Subject Effects Dependent Variable: Sum of the memory task

Source	*df*	*Mean Square*	*F*	*Sig.*
Corrected Model	3	31.416	5.867	.005
Intercept	1	3392.076	633.525	.000
Group	1	46.671	8.717	.008
Condition	1	24.005	4.483	.047
*Group * Condition*	1	10.076	1.882	.185
Error	20	5.354		
Total	24			
Corrected Total	23			

a. R Squared = .468 (Adjusted R Squared = .388)

A two-way ANOVA was performed to examine the effect of group (autistic or neurotypical), and language condition (L1 or L2), as well as their interaction on memory task scores. The means and standard deviations for independent variables (group and language condition) are presented in table 1. A two-way ANOVA revealed that there is no significant interaction between the effects of group and language condition F(1,20) = 1.882, p > .05. Simple main effects analysis showed that group had a statistically significant effect on memory task F(1,20) = 8.72, p < .01. Simple main effect analysis also showed that language condition had a statistically significant effect on memory task F(1,20) = 4.48, p < .05.

Discussion

Alexithymia, Socio-emotional skills in Autism

The aim of this study was to investigate the difference in cognitive performance in autistic and neurotypical adolescents during an emotional eliciting situation using L1 or L2. According to existing literature, autistic adolescents tend to have more problems expressing and identifying their emotions as well as understanding the emotions of others, also known as alexithymia (Hill et al., 2004). We wanted to examine whether the traits in alexithymia were present in autistic adolescents in our sample. Our study revealed that autistic adolescents showed higher levels of alexithymia compared to neurotypicals. These findings are consistent with existing literature suggesting that alexithymia is commonly observed in autism (Hill et al., 2004; Fitzgerald & Bellgrove, 2006) and that autistic individuals have trouble identifying and describing their feelings (Hill et al., 2004; Shah et al., 2006). Moreover, evidence suggests increased levels of alexithymia in autistic individuals compared to neurotypicals (Berthoz & Hill, 2005; Hill et al., 2004; Maisel et al., 2016; Rieffe et al., 2007; Rieffe et al., 2010; Samson et al., 2012). High levels of alexithymia traits that are often observed in autistic adolescents can negatively impact their subjective emotion regulation (Connelly & Deney, 2007). As expected, our study revealed that autistic individuals have more issues with socio-emotional skills compared to neurotypical adolescents. These findings go in line with existing literature stating that reduced emotional self-awareness may hinder the autistic individual's ability to understand and regulate their difficult emotions (Hill et al., 2004).

Performance in memory task between autistic vs neurotypical adolescents

Previous existing literature demonstrated that emotional difficulties are associated with lower performance in cognitive processes (Richards, 2004). Since autistic adolescents tend to have emotional difficulties (Morie et al., 2019; Mazefsky et al., 2013), this could mean that they might have worse performance in cognitive tasks than neurotypicals. Contrary to our hypothesis, our results showed that autistic adolescents scored higher in the visual memory task than neurotypicals. An alternative explanation as to why autistic adolescents scored higher in the memory task, could be that some autistic adolescents have cognitive strengths regarding their visual perception (Dakin & Faith, 2005). There is evidence suggesting that autistic individuals often demonstrated to have enhanced perceptual abilities (Dakin & Faith, 2005) when engaging in visual search (Joseph et al., 2009; Kemner et al., 2008; O'Riordan, 2004) and visual discrimination tasks (Bertone et al., 2005; Plaisted et al., 1998). Overall, individuals with autism have shown to have a superior performance in visual search tasks (Plaisted et al., 1998). Furthermore, there are also studies that suggest that autistic individuals showed superior performance when it comes to tasks that are detail related, which is linked to enhanced memory (Happé & Frith, 2006). Although autistic adolescents have been shown to have high alexithymia traits, less emotional ability and socio-emotional skills, these factors in fact did not impair their cognitive performance and autistic adolescents even scored higher than neurotypicals.

Language condition and memory

Evidence shows that language is an important tool for understanding and regulating children's emotions and using it to find appropriate ways to do so (Kopp, 1989; 1992). Moreover, there is a positive correlation between the use of language as a distraction in a frustrating situation (Stansbury & Zimmerman, 1999). Findings suggest that our L2 can be used to distance ourselves from intense emotions (Zentella, 1997). However, there is also evidence that using L1 can be beneficial as well as L1 is often used to express intense emotions (Zentella, 1997). Expressing our emotions in our L1 can help categorize them (Lindquist et al., 2015), which consequently shape our bodily sensations (Lindquist & Barrett, 2008). In our study, we were particularly interested in examining whether the adolescents would score higher in memory task after L1 or in their L2 during an emotion eliciting situation. Supporting existing literature, our results revealed that adolescents in L1 condition

had a higher memory task score than adolescents in L2 condition. These findings imply that it is more beneficial to use L1 during an emotion eliciting situation. It can help us regulate our emotions if we express them in our L1. By regulating our emotions in our L1, more cognitive resources are reserved for tasks that require cognitive resources in other areas.

Effect of group and language condition on memory tasks

It is important to note that our sample in L1 and L2 condition consisted of adolescents from both autistic and neurotypical groups. The combination of group condition and language condition did not have a statistically significant effect on the memory. However, the results revealed that language condition and group condition separately had a statistically significant effect on the memory task. These findings indicate that whether the adolescent was neurotypical or autistic as well as the language in which the interview was conducted after the emotion eliciting situation had an impact on their cognitive performance.

Limitations

Although the study has shown valuable insights, it is important to consider the limitations that may affect the generalizability of the findings. This discussion highlights the limitations of the study, including the fact that the data collection is online, long tasks, a small sample size, and the exclusion of participants with intellectual disabilities. We decided to conduct a study online to reach a broader and more diverse range of participants throughout Europe. However, the online environment introduces certain limitations that need to be considered. For instance, it was not possible to find out whether the relatives accompanied the children or if they did the study by themselves or with friends. This reduced experimental control can lead to falsified data in the questionnaires that needed to be originally filled out by the relatives. Additionally, the absence of a direct supervisor may affect the engagement in the questionnaire. Another disadvantage of the study is the length of the cognitive tasks which can lead to boredom, fatigue, and a reduced attention span especially in children. Investing too much time in these tasks can influence the quality of the data, leading to results with worse quality and accuracy. Thus, it is important for future studies to shorten the task duration to improve participants' engagement. Furthermore, the small sample size of 24 participants makes it difficult to generalize the findings of the study to the overall population. Due to having a small group of participants, it is not representative. Future studies should aim for larger sample sizes to improve the validity of the results. The decision to exclude participants with intellectual disability was made to ensure a homogeneous sample because the participants should be comparable with each other. However, this exclusion affects the generalizability in a negative way since it is not representative for the population. Future research should include a broader range of participants to understand the impact of emotion on cognition across different populations. If the study included adolescents with intellectual disabilities, it would be necessary to do some changes that need to be done in order to adapt to their needs. It is important to find out the intellectual disability as well as the severity of it to be able to understand its impact on the study. Further, the instructions can be made in a very easy and understandable language and additional support like videos or voice notes can be provided so that the study can be accessible to a wide range of people. It reduces the risk of misinterpretation since technical language can be challenging for people with intellectual disabilities. Moreover, it is necessary to shorten tasks with long duration such as the video task, and break it into smaller pieces due to a probably shorter attention span than in neurotypicals, so that engagement can be maintained with as little loss in quality as possible. Another aspect would be to have several options to respond, if a person struggles with conveying a message, they can either respond verbally or use videos. By incorporating these strategies, it is possible to include more individuals in our study which could also lead to more participants, which makes it more representative.

Future direction

Future research should be conducted by further examining the interaction between socio-emotional difficulties and cognitive abilities to get a better understanding of the factors that could explain the enhanced cognitive performance observed in autistic adolescents. However, since there is evidence suggesting that using either the L1 or L2 have a positive impact on cognitive performance after an emotion- eliciting situation, another area that could be further explored are factors or mechanisms that contribute to the positive impact of using either L1 or L2 on cognitive performance following an emotional-eliciting situation by further investigating the complex interaction between language and cognitive abilities.

Conclusion

To summarize our important findings, autistic adolescents scored higher in memory task than neurotypical adolescents. These findings challenge the assumption that socio-emotion difficulties lead to poor cognitive performance in autistic adolescents. Additionally, it is important to consider the heterogeneity within the autism spectrum when discussing cognitive deficits however, it is also crucial to acknowledge cognitive strengths too. Autistic adolescents may have performed better in visual memory tasks compared to neurotypical adolescents, likely due to their enhanced visual perception associated with autism. Apart from the significant difference observed between autistic and neurotypicals in their cognitive performance, our findings also revealed that the use of L1 led to better cognitive performance suggesting the significance of language use during an emotional eliciting situation on cognitive performance.

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

APA Dictionary of Psychology. (n.d.). https://dictionary.apa.org/sympt omatology

APA Dictionary of Psychology. (n.d.-b). https://dictionary.apa.org/trait

Ashburner, J., Ziviani, Jenny & Rodger, Sylvia. (2010). Surviving in the mainstream: Capacity of children with autism spectrum disorders to perform academically and regulate their emotions and behaviour at school. Research in Autism Spectrum Disorders. 4. 18-27. https://doi.org/10.1016/j.rasd.2009.07.002.

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The Twenty Item Toronto Alexithymia Scale – I. Item selection and cross validation of the factor structure. Journal of Psychosomatic Research, 38(1), 23–32. https://doi.org/10.1016/0022-3999(94)90005-1

Barnard, Judith & Prior, Aidan & Potter, David. (2000). Inclusion and Autism: Is It Working?

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. Annual review of psychology, 58, 373–403. https://doi.org/10.1146/annurev.psych.58.110405.085709

Beaumont, R., & Sofronoff, K. (2008). A multi-component social skills intervention for children with Asperger syndrome: the Junior Detective Training Program. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 49(7), 743–753. https://doi.org/10.1111/j.1469-7610.2008.01920.x

Berthoz, S., & Hill, E. L. (2005). The validity of using self-reports to assess emotion regulation abilities in adults with autism spectrum disorder. European Psychiatry, 20(3), 291–298. https://doi.org/10.1016/j.eurpsy.2004.06.013

Bertone, A., Mottron, L., Jelenic, P., & Faubert, J. (2005). Enhanced and diminished visuo-spatial information processing

in autism depends on stimulus complexity. Brain: a journal of neurology, 128(Pt 10), 2430–2441. https://doi.org/10.1093/brain/awh561

Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 65(4), 229–235. https://doi.org/10.1037/a0025406

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. Journal of experimental psychology. Learning, memory, and cognition, 34(4), 859–873. https://doi.org/10.1037/0278-7393.34.4.859

Bird, G., & Cook, R. (2013). Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism. Translational psychiatry, 3(7), e285. https://doi.org/10.1038/tp.2013.61

BlaiBlair C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. American Psychologist, 57, 111-127 https://dx.doi.org/10.1037/0003-066X.57.2.111

Bölte, S., Poustka, F., Constantino, J.N. (2008). Assessing autistic traits: cross-cultural validation of the social responsiveness scale (SRS). Autism Research, 1(6), 354-363. https://doi.org/10.1002/aur.49.

Brown, R. H., Murray, A., Stewart, M. E., & Auyeung, B. (2021, April 22). Psychometric Validation of a Parent- Reported Measure of Childhood Alexithymia: The Alexithymia Questionnaire for Children – Parent (AQC-P). European Journal of Psychological Assessment. Advance online publication. http://dx.doi.org/10.1027/1015-5759/a000640

Butterworth, T., Hodge, A., Sofronoff, K., Beaumont, R., Gray, K. M., Roberts, J., Horstead, S. K., Clarke, K. S., Howlin, P., Taffe, J. R., & Einfeld, S. L. (2014). Validation of the Emotion Regulation and Social Skills Questionnaire for Young People with Autism Spectrum Disorders. Journal of Autism and Developmental Disorders, 44(7),1535–1545.https://doi.org/10.1007/s10803-013- 2014-5

Cai, R.Y., Richdale, A. L., Uljarević, M., Dissanayake, C., & Samson, A. C. (2018). Emotion regulation in autism spectrum disorder: Where we are and where we need to go. Autism research: official journal of the International Society for Autism Research, 11(7), 962–978. https://doi.org/10.1002/aur.1968

Connelly, M., & Denney, D. R. (2007). Regulation of emotions during experimental stress in alexithymia. Journal of Psychosomatic Research, 62(6), 649–656. https://doi.org/10.1016/j.jpsychores.2006.12.008

Conner, C. M., Golt, J., Shaffer, R., Righi, G., Siegel, M., & Mazefsky, C. A. (2021). Emotion Dysregulation is Substantially Elevated in Autism Compared to the General Population: Impact onPsychiatric Services. Autism research : official journal of the International Society for Autism Research, 14(1), 169–181. https://doi.org/10.1002/aur.2450

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, not autism, predicts poor recognition of emotional facial expressions. Psychological science, 24(5), 723–732. https://doi.org/10.1177/0956797612463582

Costa, A. P., Steffgen, G., & Samson, A.C. (2017). Expressive incoherence and alexithymia in autism spectrum disorder. Journal of Autism and Developmental Disorders, 47(6), 1659–1672. https://doi.org/10.1007/s10803-017-3073-9

Dakin, S., & Frith, U. (2005). Vagaries of visual perception in autism. Neuron, 48(3), 497–507. https://doi.org/10.1016/j.neuron.2005.10.018

Davis, Elizabeth & Levine, L.J. (2013). Emotion Regulation Strategies That Promote Learning: Reappraisal Enhances Child-

ren's Memory for Educational Information. Child development. 84. https://doi.org/10.1111/j.1467-8624.2012.01836.x.

Diamond, A. (2013). Executive functions. Annual Revision Psychology. 64:135-68. https://doi.org/ 10.1146/annurev- psych-113011-143750

Dijkhuis R, De Sonneville, L., Ziermans, T., Staal, W., Swaab, H. (2020). Autism Symptoms, Executive Functioning and Academic Progress in Higher Education Students. J Autism Dev Disord. 50(4):1353-1363. https://doi.org/10.1007/s10803

Eaves, L. C., & Ho, H. H. (1997). School placement and academic achievement in children with autistic spectrum disorders. Journal of Developmental and Physical Disabilities, 9, 277–291.

Fitzgerald, M., & Bellgrove, M. A. (2006). THE OVERLAP BETWEEN ALEXITHYMIA AND ASPERGER'S SYNDROME. Journal of Autism and Developmental Disorders, 36(4), 573–576. https://doi.org/10.1007/s10803-006-0096-z

Gonzalez-Barrero, A. M., & Nadig, A. S. (2019). Can Bilingualism Mitigate Set-Shifting Difficulties in Children With Autism Spectrum Disorders?. Child development, 90(4), 1043–1060. https://doi.org/10.1111/cdev.12979 Graziano, P. & Reavis, Rachael & Keane, Susan & Calkins, Susan. (2007). The Role of Emotion Regulation and Children's Early Academic Success. Journal of school psychology. 45. 3-19. https://doi.org/10.1016/j.jsp.2006.09.002

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well- being. Journal of personality and social psychology, 85(2), 348–362. https://doi.org/10.1037/0022-3514.85.2.348

Gross, J. J., & Thompson, R. A. (2007). Emotion Regulation: Conceptual Foundations. In J. J. Gross (Ed.), Handbook of emotion regulation (pp. 3–24). The Guilford Press. https://doi.org/10.1037/0022-3514.79.3.410

Hess, P. (2020, 7. Mai). Autistic girls may experience more emotional challenges than autistic boys do. Spectrum | Autism Research News. https://www.spectrumnews.org/news/autistic-girls-may-experience-more- emotional-challenges-than-autistic- boys-do/

Hill, N. E., Castellino, D. R., Lansford, J. E., Nowlin, P., Dodge, K. A., Bates, J. E., & Pettit, G. S. (2004). Parent academic involvement as related to school behavior, achievement, and aspirations: demographic variations across adolescence. Child development, 75(5), 1491–1509. https://doi.org/10.1111/j.1467-8624.2004.00753.x

Joseph, R. M., Keehn, B., Connolly, C., Wolfe, J. M., & Horowitz, T. S. (2009). Why is visual search superior in autism spectrum disorder? Developmental science, 12(6),1083–1096. https://doi.org/10.1111/j.1467-7687.2009.00855.x

Kaldy, Z., Kraper, C., Carter, A. S., & Blaser, E. (2011). Toddlers with Autism Spectrum Disorder are more successful at visual search than typically developing toddlers. Developmental science, 14(5), 980–988. https://doi.org/10.1111/j.1467-7687.2011.01053.x

Karrass, J., Walden, T. A., Conture, E. G., Graham, C. G., Arnold, H. S., Hartfield, K. N. & Schwenk, K. A. (2006). Relation of emotional reactivity and regulation to childhood stuttering. Journal of Communication Disorders, 39(6), 402–423. https://doi.org/10.1016/j.jcomdis.2005.12.004

Keen, D., Webster, A., Ridley, G. (2016). How well are children with autism spectrum disorder doing academically at school? An overview of the literature. Autism: the international journal of research and practice. 20(3):276-94. https://doi.org/10.1177/1362361315580962.

Kopp, C. B. (1989). Regulation of distress and negative emotions: A developmental view. Developmental Psychology,

25(3), 343–354. https://doi.org/10.1037/0012-1649.25.3.343

Lamme, V. & Roelfsema, Pieter. (2000). The Distinct Modes of Vision Offered by Feedforward and Recurrent Processing. Trends in neurosciences.23. 571-9. https://doi.org/10.1016/S0166-2236(00)01657-X.

Langener, A.M., Kramer, A.W., van den Bos, W., Huizenga, H.M. (2022) A shortened version of Raven's standard progressive matrices for children and adolescents. The British journal of developmental psychology. Mar;40(1):35-45. https://doi.org/10.1111/bjdp.12381 .

Laushman, P. (2023). How to Overcome Social Fatigue. Thrive Autism Coaching. https://www.thriveautismcoaching.com/post/how-to-overcome-social- fatigue

Lecavalier, L. (2006). Behavioral and Emotional Problems in Young Peoplewith Pervasive Developmental Disorders: Relative Prevalence, Effects of Subject Characteristics, and Empirical Classification. Journal of Autism and Developmental Disorders, 36(8), 1101–1114. https://doi.org/10.1007/s10803-006-0147-5

Li, P., Zhang, F., Yu, A., & Zhao, X. (2019). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. Bilingualism: Language and Cognition, 23(5), 938–944. https://doi.org/10.1017/s1366728918001153

Lindquist, K. A. & Barrett, L. F. (2008). Constructing Emotion. Psychological Science, 19(9), 898–903. https://doi.org/10.1111/j.1467-9280.2008.02174.x

Lindquist, K. A., Satpute, A. B., & Gendron, M. (2015). Does Language Do More Than Communicate Emotion? Current Directions in Psychological Science, 24(2), 99–108. https://doi.org/10.1177/0963721414553440

Lumley, M. A., Neely, L. C., & Burger, A.J. (2007). The assessment of alexithymia in medical settings: implications for understanding and treating health problems. Journal of personality assessment, 89(3), 230–246. https://doi.org/10.1080/00223890701629698

Lupyan, G.(2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. Frontiers in Psychology, 3. https://doi.org/10.3389/fpsyg.2012.00054

Maisel, M. E., Stephenson, K. G., South, M., Rodgers, J., Freeston, M. H., & Gaigg, S. B. (2016). Modeling the cognitive mechanisms linking autism symptoms and anxiety in adults. Journal of abnormal psychology, 125(5), 692–703. https://doi.org/10.1037/abn000016 Mazefsky, C. A., Herrington, J., Siegel,

M., Scarpa, A., Maddox, B. B., Scahill, L., & White, S. W. (2013). The role of emotion regulation in autism spectrum disorder. Journal of the American Academy of Child and Adolescent Psychiatry, 52(7), 679–688. https://doi.org/10.1016/j.jaac.2013.05.006

Mazzoni, N., Ricciardelli, P., Actis- Grosso, R., & Venuti, P. (2022). Difficulties in Recognising Dynamic but not Static Emotional Body Movements in Autism Spectrum Disorder. Journal of autism and developmental disorders, 52(3), 1092–1105. https://doi.org/10.1007/s10803-021-05015-7

Nemiah, J. C., Freyberger, H., & Sifneos, P. E. (1976). Alexithymia: A view of the psychosomatic process. In O. W. Hill (Ed.), Modern trends in psychosomatic medicine (Vol. 3, pp. 430-439). London: Butterworths.

Niedenthal, P. M. (2007). Embodying emotion. Science (New York, N.Y.), 316(5827),1002–1005. https://doi.org/10.1126/science.1136 930

O'riordan, M. A. (2004). Superior visual search in adults with autism. Autism: the international journal of research and practice, 8(3), 229–248.

https://doi.org/10.1177/1362361304045219

Perlovsky, L. (2009). Language and cognition. Neural networks : the official journal of the International Neural Network Society, 22(3), 247–257. https://doi.org/10.1016/j.neunet.2009.03.007

Perlovsky, L. (2009). Language and emotions: Emotional Sapir-Whorf hypothesis. Neural Networks, 22(5- 6), 518–526. https://doi.org/10.1016/j.neunet.2009.06.034

Pessoa, L., & Adolphs, R, (2010). Emotion and the amygdala: from a 'low road 'to' many roads of evaluating biological significance, Nature review neuroscience, 11 (11), 773 -782

Plaisted, K., O'Riordan, M., & Baron- Cohen, S. (1998). Enhanced visual search for a conjunctive target in autism: a research note. Journal of child psychology and psychiatry, and allied disciplines, 39(5), 777–783.

Richards, J. M. & Gross, J. J. (2000). Emotion regulation and memory: The cognitive costs of keeping one's cool. Journal of Personality and Social Psychology, 79(3), 410–424. https://doi.org/10.1037/0022-3514.79.3.410

Rieffe, C., Ketelaar, L., & Wiefferink, C. (2010). Assessing empathy in young children: Construction and validation of an Empathy Questionnaire (EmQue). Personality and Individual Differences, 49(5), 362–367. https://doi.org/10.1016/j.paid.2010.03.046

Rieffe, C., Meerum Terwogt, M., & Kotronopoulou, K. (2007). Awareness of single and multiple emotions in high-functioning children with autism. Journal of autism and developmental disorders, 37(3), 455–465. https://doi.org/10.1007/s10803-006-0171-5

Rieffe, C., Oosterveld, P., & Terwogt, M. M. (2006). An Alexithymia Questionnaire for Children: Factorial and concurrent validation results. Personality and Individual Differences, 40(1), 123–133. https://doi.org/10.1016/j.paid.2005.05.013

Samson, A. C., Hardan, A. Y., Podell, R.W., Phillips, J. M., & Gross, J. J. (2015). Emotion regulation in children and adolescents with autism spectrum disorder. Autism research: official journal of the International Society for Autism Research, 8(1), 9–18. https://doi.org/10.1002/aur.1387 Samson, A. C., Huber, O., & Gross, J. J. (2012). Emotional reactivity and regulation in adults with autism spectrum disorders. Emotion, 12(4), 659-665.

Samson, A. C., Phillips, J. M., Parker, K. J., Shah, S., Gross, J. J., & Hardan, A. Y. (2014). Emotion dysregulation and the core features of autism spectrum disorder. Journal of autism and developmental disorders, 44(7), 1766–1772. https://doi.org/10.1007/s10803-013-2022-5

Scheibe, S., Blanchard-Fields, F. (2009) Effects of regulating emotions on cognitive performance: what is costly for young adults is not so costly for older adults. Psychol Aging. 24(1):217-23. https://doi.org/10.1037/a0013807.

Shah, A., & Wing, L. (2006). Psychological approaches to chronic catatonia-like deterioration in autism spectrum disorders. International review of neurobiology, 72, 245–264. https://doi.org/10.1016/S0074-7742(05)72015-8

Shapero, B.G., Dale, C.F., Archibald, A.R., Pedrelli, P. (2016). Emotional Reactivity. In: Levesque, R. (eds) Encyclopedia of Adolescence. Springer, Cham. https://doi.org/10.1007/978-3-319-32132-5_786-1

Sifneos, P. E. (1973). The prevalence of 'alexithymic' characteristics in psychosomatic patients. Psychotherapy and psychosomatics, 22(2), 255–262. https://doi.org/10.1159/000286529

Thompson, R. A. (1994). Emotion regulation: a theme in search of definition.

Monographs of the Society for Research in Child Development, 59(2-3), 25–52.

Trevisan, D. A., Hoskyn, M. & Birmingham, E. (2018). Facial Expression Production in Autism: A Meta-Analysis. Autism research: official journal of the International Society for Autism Research, 11(12), 1586–1601. https://doi.org/10.1002/aur.2037

Uddin, L.Q. (2022). Exceptional abilities in autism: Theories and open questions. Curr Dir Psychol Sci. 31(6):509-517. https://doi.org/10.1177/09637214221113760.

Westdeutscher Rundfunk (2022, March 6). Wasser in der Wüste [Video]. Die Sendung mit der Maus-WDR. https://www.wdrmaus.de/filme/sachg eschichten/wasser_in_der_wueste.p hp5

Williams, K. D., & Jarvis, B. (2006). Cyberball: a program for use in research on interpersonal ostracism and acceptance. Behaviour Research Methods, 38(1), 174–180. https://doi.org/10.3758/bf03192765

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. Cognition, 85(3), 223–250. https://doi.org/10.1016/s0010-0277(02)00109-9

Zentella, A. C. (1997). Growing up Bilingual: Puerto Rican Children in New York. CiNii Books. http://ci.nii.ac.jp/ncid/BA30469460

Funding: The present study was supported by the Luxembourg National Research Fund (FNR) (13651499).

Impressum

Luxemburger Experimentalpraktikum Journal
Band 17, Jahrgang 2023

Herausgeber

Dr. Andreia Costa
Koordinatorin des Practical Training in Empiricial Research
Université du Luxembourg
Maison Sciences Humaines
11, Porte des Sciences
L-4366 Esch-sur-Alzette

Redaktion

Clémentine Offner, Studentin des BAP

Mitherausgeber

Nina Buntić
Elisabeth Holl
Gary Wagener
Dr. Mila MARINOVA
Styliani POLITI
Dr. André Melzer
Dr. Andreia Costa
Maïte Franco

Manuskriptrichtlinien

Die Beiträge richten sich nach dem „Publication Manual of the American Psychological Association" (2010) bzw. den „Zur Gestaltung von Haus- und Abschlussarbeiten" (2004) der Deutschen Gesellschaft für Psychologie.

Erscheinungsweise

Ein- bis zweimal jährlich

Bezug

Ass.-Prof. André Melzer, Université du Luxembourg, Maison Sciences Humaines 11, Porte des Sciences, L-4366 Esch-sur-Alzette

Online unter

https://bap.uni.lu

ISSN 3093-1045