Dissertation

Student ID: 11484265

2024-07-04

rmdwc::rmdcount('Paper.rmd')

add that LCA also uses same motives, but 2 with diff wording add that social cognitive theory is outdated (if time)

Contents

Introduction

Physical inactivity remains widespread across different age groups and is considered a major public health issue (Mitchell, 2019). Nearly one third of global population suffer from insuffıcent physical activity, a 9% increase from 20 years ago (Strain et al, 2024). Understanding the factors that influence engagement in PA is critical, as regular activity supports cardiovascular health, metabolic function, mental well-being, and overall quality of life. Identifying the underlying determinants of PA can inform interventions, policies, and educational strategies aimed at improving health outcomes in both youths and adults. Motivation has consistently been identified as a central determinant of behaviour, including participation in physical activity (Daley & Duda, 2006; Deci & Ryan, 2008; Duncan et al., 2010; Falk et al., 2015; Flannery, 2017). However, most existing studies investigate age-related differences within youths or within adults, but rarely compare the two populations directly. This leaves unclear whether the same motives operate similarly across life stages, or whether different developmental contexts shape the salience of specific motives. To address this, the present study explores how different motives for PA manifest across youths and adults, using both observed differences and latent class analysis to identify distinct motivational profiles.

Research Questions

1. Do perceived exercise motives influence physical activity differently in youths and adults?
2. How do age differences shape dominant exercise motives within youth and adult groups?

To test whether motives predict PA differently across youths and adults, multigroup structural equation modelling was used to examine associations between key motives and self-reported PA. Latent class analysis was used to identify motivational profiles within each group, enabling age-based comparisons.

Literature Review

While other frameworks, such as Social Cognitive Theory, are widely cited, this dissertation focuses on Self-Determination Theory and the Theory of Planned Behaviour, which provide useful conceptual perspectives for understanding and interpreting motivational patterns.

Self-Determination Theory (SDT)

Self-Determination Theory (SDT) is a popular framework that has gained popularity in the past two decades. Developed by Deci & Ryan, it has shed light on the understanding of human motivation, emphasizing the degree to which behavior is self-determined versus controlled. Motivation is distinguished in terms of autonomy, with autonomous forms associated with volition and self-endorsement of behavior, and controlled forms reflecting pressure from external contingencies or internalized demands such as guilt or ego involvement. Within SDT, extrinsic motivation is subdivided into external regulation, introjected regulation, identified regulation, and integrated regulation, representing increasing internalization of instrumental behaviors. Intrinsic motivation, by contrast, arises from interest, enjoyment, or inherent satisfaction in the activity itself. These motivational types have implications for physical activity, as more autonomous motives are linked to sustained engagement, positive psychological outcomes, and long-term health (CITE).

In physical activity and health behaviour research, this framework is especially influential because it links the quality of motivation, rather than its quantity, to long-term adherence and wellbeing. Empirical studies confirm that need-supportive interventions can improve exercise uptake, reduce dropout, and enhance both physical and psychological wellbeing (Teixeira et al., 2012; Ng et al., 2012). Consequently, SDT provides not just a descriptive model of motivation but also a prescriptive guide for designing effective physical activity and health promotion programmes.

Theory of Planned Behaviour (TPB)

The Theory of Planned Behavior (TPB), proposed by Ajzen in 1991, is a widely used framework for predicting and understanding human behavior, emphasizing the role of intention as the proximal determinant of action. According to TPB, behavioral intentions are influenced by three key factors: attitudes toward the behavior, subjective norms, and perceived behavioral control. Attitudes reflect an individual's positive or negative evaluation of performing the behavior, subjective norms capture perceived social pressure from significant others to engage or not engage, and perceived behavioral control represents the perceived ease or difficulty of performing the behavior, akin to self-efficacy. These factors interact to shape intention, which in turn predicts behavior, although perceived behavioral control can also have a direct effect on behavior. In the context of physical activity, TPB has been used to explain variations in exercise participation across age groups, demonstrating that stronger intentions—supported by favorable attitudes, positive social norms, and higher perceived control—are associated with higher levels of activity. However, the framework also recognizes that intentions do not always translate into behavior, highlighting the importance of situational constraints and individual capabilities.

Age Differences

Across multiple studies, autonomous motivations, such as intrinsic motivation and identified regulation, are consistently associated with higher levels of physical activity. However, it was also consistently found that intrinsic motivation decreases with age (Brunet & Sabiston, 2011; Dishman et al., 2018; Nascimento et al., 2023).

Longitudinal evidence from adolescents indicates that intrinsic motivation and integrated regulation can mitigate age-related declines in physical activity, but their effects are conditional on goal contents. Specifically, Dishman et al. (2018) found that adolescents who maintained higher intrinsic motivation or integrated regulation remained more active when they also maintained higher enjoyment goals. Declines in appearance, social, or competence goals could either weaken or strengthen the protective influence of autonomous motivation depending on the interaction. These findings suggest that enjoyment- and social-oriented motives play a critical role in sustaining physical activity in youths, and that interventions targeting these motives should consider the interaction with specific goal contents.

Among adults, Brunet and Sabiston (2011) observed that identified regulation remained high and external regulation low across age groups, suggesting that adults engage in activity because it aligns with personal values rather than external incentives. However, middle-age adults exhibited lower intrinsic motivation, introjected regulation, and physical activity than younger adults. Nascimento et al. (2023) similarly found that adolescents and younger adults were motivated more by extrinsic factors such as appearance and social approval, whereas older adults emphasized intrinsic motives related to psychological well-being. These studies collectively indicate that fitness- and health-related motives gain prominence with age.

In adolescents, intrinsic motivation, self-efficacy, and enjoyment of physical activity are strong predictors of habitual engagement, with social support from parents, teachers, and peers playing a facilitative role (Shao & Zhou, 2023). Boys consistently report higher overall physical activity, whereas girls tend to engage more in moderate-to-vigorous activities, and a gradual decline in activity is observed across adolescence. These findings align with longitudinal evidence suggesting that intrinsic motivation in youths helps buffer against age-related declines in physical activity, particularly when coupled with the maintenance of personally meaningful goals such as enjoyment or competence (Dishman et al., 2018). Social and goal-oriented factors may act as effect modifiers, amplifying or mitigating the influence of autonomous motivation on behavior during this critical developmental period.

In contrast, studies focusing on adults over 70 highlight a shift in motivational determinants, with barriers increasingly dominated by concerns about health, functional ability, and fear of injury, including fear of falling (Kilgour et al., 2024). Social motivators, such as support from family or the opportunity for interaction during activity, emerge as key facilitators, while personal fulfillment and accessible facilities are also cited, albeit less consistently. Notably, intrinsic or autonomous forms of motivation are less frequently reported as primary drivers in this population, suggesting that the relative importance of autonomous versus controlled motivation may decline with advanced age. The evidence also indicates gendered differences in motivators and barriers, with women valuing social aspects more strongly and men more often citing lack of interest as a barrier, highlighting the need for tailored interventions.

Evidence also suggests that certain contextual factors, such as opportunity or access to facilities, may influence physical activity similarly across age groups. While the specific motives driving activity vary with age, supportive environmental contexts appear to facilitate activity regardless of life stage, consistent with SDT's emphasis on autonomy-supportive environments (Brunet & Sabiston, 2011; Dishman et al., 2018).

While the body of literature provides valuable insights into the determinants of physical activity across different age groups, several limitations temper the conclusions. Many studies of adolescents, including Dishman et al., Shao and Zhou, rely on self-report measures for both activity and motivation, which introduces potential biases such as social desirability or inaccurate recall. Even in longitudinal designs, as in Dishman et al., causality cannot be firmly established; observed associations between intrinsic motivation, goal content, and physical activity may be confounded by unmeasured factors such as peer influences, family context, or school policies. Most data are derived from high-income countries with predominantly white participants, limiting cross-cultural generalizability. Functional impairments and comorbidities are inconsistently measured, making it difficult to disentangle the effects of physical limitations from motivational factors. Across the lifespan, gender differences are inconsistently explored, with some studies noting divergent patterns but few systematically testing interactions between sex, age, and motivation type.

Hypothesis 1: The influence of exercise motives on physical activity differs between youths and adults.

- H1a: Enjoyment, social, and guilt motives are more influential in youths than in adults.
- H1b: Fitness related motives are more influential in adults than in youths, particularly in middle-aged and older groups.
- H1c: Environmental opportunities influence physical activity similarly across age groups.

Motivational Profiles

In adults, Ostendorf et al. (2021) identified three primary motivational profiles in individuals with overweight or obesity: high autonomous, high combined, and moderate combined. The high autonomous profile was characterized by strong intrinsic and identified motivations, with minimal influence of external or introjected regulation. The high combined profile reflected elevated levels across all regulatory types, while the moderate combined profile exhibited intermediate levels on all regulations. Notably, baseline device-measured moderate-to-vigorous physical activity (MVPA) did not differ significantly across profiles, in contrast to prior studies relying on self-reported PA that suggested higher autonomous profiles correlate with greater activity (Friederichs et al., 2015; Gourlan et al., 2016). Longitudinally, the high autonomous profile demonstrated the least decline in MVPA during transitions from supervised to unsupervised exercise, suggesting that intrinsic and identified motivations support sustained behavior, whereas moderate-to-high external regulation may necessitate continued support for adherence. Limitations of this study include a small, predominantly female and motivated sample, the absence of amotivation as a profile dimension, reliance on baseline-only motivation measurement, and potential dropout-related attenuation of observed associations, limiting generalizability to broader adult populations.

In adolescents, Moreno-Murcia et al. (2011) found two primary motivational profiles in physical education students: a self-determined profile, with high scores on intrinsic and identified regulation, and a non-self-determined profile, characterized by elevated external, introjected, and amotivated scores. The self-determined profile was positively associated with Theory of Planned Behavior constructs such as intention, subjective norm, perceived behavioral control, and attitude, suggesting that autonomous motivation supports both favorable cognitions and participation behaviors. However, the study relied on self-reported measures, was conducted in a non-representative sample of Spanish children, and integrated regulatory forms that may not fully emerge in younger populations, limiting developmental generalizability.

Cross-sectional studies of secondary school and PE students provide converging evidence. Manzano-Sánchez et al. (2019) identified distinct motivational profiles among adolescents aged 12–16, showing that higher motivation profiles were associated with adaptive outcomes such as greater responsibility, resilience, and perceived social support. Similarly, Heredia-León et al. (2021) found that students with high-quality and high-quantity motivational profiles demonstrated greater autonomous motivation, intention to be physically active, and enjoyment in PE classes, whereas low-quality or low-quantity profiles corresponded to higher boredom and lower engagement. Both studies support the idea that motivation is multifaceted and that enjoyment- and autonomy-oriented motives are particularly salient for younger populations, which aligns with H1a.

Tapia-Serrano et al. (2022) further extended the understanding of motivational profiles in children, identifying five profiles: highly amotivated, moderately amotivated, averagely motivated, controlled motivated, and autonomously motivated. While autonomously motivated students maintained higher PA intentions long-term, controlled motivated students also demonstrated elevated PA behavior in the short term, highlighting that controlled motivation may, in specific contexts, promote adaptive engagement. Nonetheless, the study relied exclusively on self-reported PA and did not capture objective long-term behavior, limiting confidence in the persistence of these effects.

Nuss et al. (2023) identified four motivational profiles in a sample of Canadian adults: high controlled and high autonomous (HCHA), high autonomous and introjected (HAI), high amotivation and external (HAE), and low overall motivation (LOM). Contrary to expectations, HCHA participants reported the highest MVPA levels across three time points, including during COVID-19 stay-at-home orders, suggesting that combinations of controlled and autonomous motivation may synergistically support activity in some contexts. Limitations include the cross-sectional assessment of motivation, reliance on retrospective self-report for multiple time points, and potential instability of profiles over time.

It's likely that neither youths nor adults form a single homogeneous motivational category. Rather, multiple profiles coexist within each population, differing in the prevalence of motives. However, methodological limitations—including reliance on self-report, limited longitudinal tracking, small or non-representative samples hinder analysis.

Hypothesis 2: Distinct motivational profiles exist within youths and within adults.

- H2a: Middle-aged and older adults are more likely than younger adults to belong to fitness-dominant profiles.
- H2b: Younger youths are more likely than older youths to belong to enjoyment-dominant profiles.

Data and Methods

- Discuss what variables you will use, how they are coded, the amount of missing data they have and present descriptive statistics.
- Discuss the statistical models that you will use. Explain how the models will answer your research questions. Discuss what sequence of models you will run.

Data

The study uses survey responses from datasets collected by Ipsos on behalf of Sport England (2024, 2025). These datasets were selected because the youth and adult surveys share a parallel structure, and several items are worded identically, providing a strong baseline for direct comparison between age groups. All motivational measures were captured using single-item survey questions. A list of relevant survey questions is provided in Appendix A. For motive variables, descriptive statistics, bivariate correlations, and variance inflation factors (VIF) were calculated to assess distributional properties, relationships among variables. Challenge and relaxation are only included in the LCA models due to different wording in the survey questions for adults and youths. A total of 117,247 adult and 30,670 youth observations were used in the SEM analyses, whereas 110,378 adult and 28,886 youth observations were included in the LCA. Only cisgender adults without disabilities were included in the analyses to maintain comparability and avoid skewed results due to small subgroup sizes. As ethnicity, education, and gender were included as control variables, these factors were retained alongside the analyzed items. Participants with missing responses on any relevant items were also excluded.

- Enjoyment – whether the individual finds exercise satisfying.
- Social engagement – exercising for fun with friends.
- Health and fitness – exercising to maintain physical well-being.
- Opportunity – having the chance to exercise.
- Guilt – sense of personal obligation to exercise.
- Challenge – exercising to push oneself or compete with others.
- Relaxation – exercising to reduce stress and worry.
- Minutes Exercised - weekly minutes of moderate-to-vigorous PA; see appendix for specific activities.

Adult Dataset

The adult dataset analysed in this dissertation is derived from the Active Lives Adult Survey (Year 8: 2022–2023), conducted by Ipsos on behalf of Sport England with additional funding from the Office for Health Improvement and Disparities (OHID). The survey is designed as a large-scale, nationally representative measure of adult participation in sport and physical activity in England. It follows a push-to-web methodology, whereby sampled households received up to four postal mailings inviting participation. The first two mailings provided instructions and passcodes to complete the survey online, while the third included a paper self-completion questionnaire to maximise response rates. A final reminder letter encouraged online participation. Data were collected between November 2022 and November 2023, with responses obtained from 172,968 adults aged 16 and above across England.

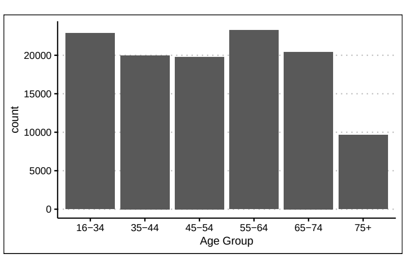
The questionnaire focused on a wide range of topics related to sport and physical activity, including frequency, duration, and location of participation, alongside attendance at sporting events, volunteering, club membership, and motivational and attitudinal measures. Standard demographic questions were also included, covering age, gender, socio-economic status, education, household composition, sexual identity, and religion. To balance questionnaire length and data breadth, some modules were administered only to subsamples of respondents, particularly in the online version.

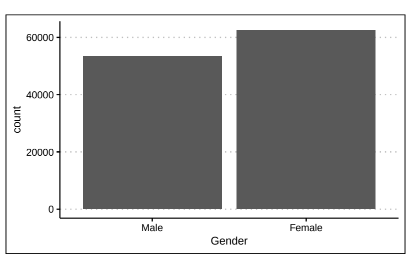
The sample was drawn using random probability sampling from the Postcode Address File (PAF). A total of 562,644 addresses were invited, with up to two adults per household eligible to respond. Multiple reminders ensured greater response coverage, helping mitigate nonresponse bias. The combination of online and postal modes, alongside stratified sampling and weighting, was intended to enhance representativeness of the adult population in England.

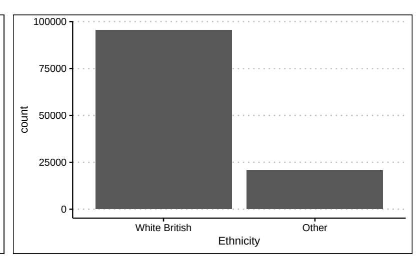
The adult dataset was drawn from households sampled via the Postcode Address File (PAF), with up to two residents aged 16 or older invited to participate through either an online survey or a paper questionnaire. Data were gathered in successive waves, aiming for approximately 500 responses per local authority, and were distributed as evenly as possible across the period from November 2022 to November 2023 to reduce seasonal bias. In total, 173,950 surveys were completed.

Each item was rated on a five-point Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree). All motive items had responses 4 (disagree) and 5 (strongly disagree) collapsed into a single category due to the low proportion of responses to minimize distortion.

Variable	Mean	Median	SD	PercentNA
Enjoyment	2.125675	2.0	1.0248819	4.344232
Social	2.886690	3.0	1.1603390	6.383262
Fitness	1.862708	2.0	0.8631442	3.942348
Guilt	2.553219	2.0	1.1044993	5.035701
Opportunity	2.010306	2.0	0.9913168	4.017506
Importance	1.979520	2.0	0.9118422	4.214147
Challenge	2.757458	3.0	1.1511982	6.142755
Relaxation	2.262504	2.0	1.0124889	5.331622
Minutes.Exercised	493.496059	337.5	475.1089002	0.000000







Females constitute 53.89% of the sample while males constitute 46.11%. 82.25% are white British. The youngest and oldest two age groups are collapsed together due to the samll size of most extreme groups. The decision to collapse the oldest adult categories (64–75 and 76+) was made because motivational profiles in these two groups did not differ meaningfully in preliminary analyses, and because the very small number of 76+ respondents would have produced unstable estimates if analyzed independently. Combining them ensured more robust model performance without obscuring theoretically important distinctions, which were not expected to be as pronounced at that stage of life.

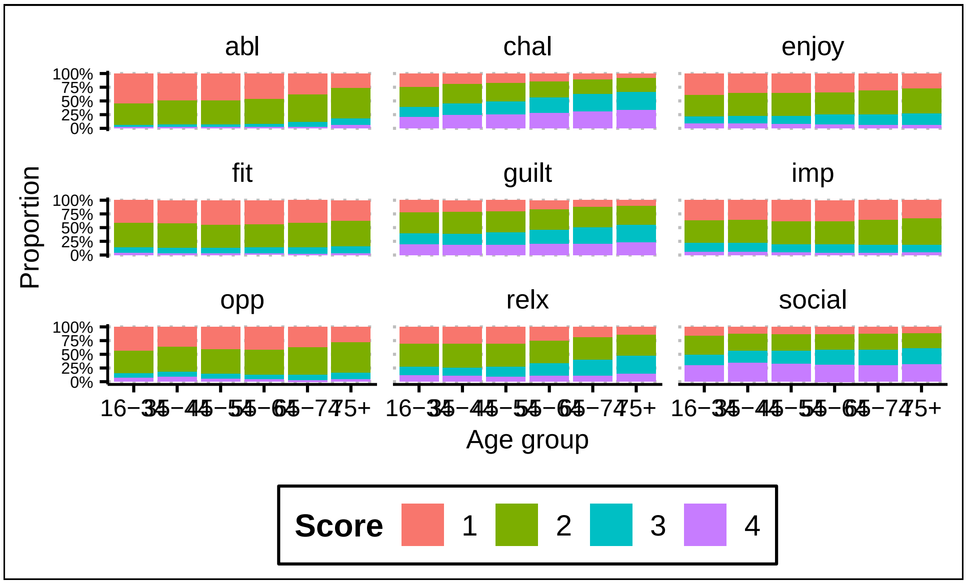


Figure 1: Distribution of adult responses.

Across most motive items, the proportion of adults expressing strong positive agreement declines as age increases. In practical terms, older adults are less likely than their younger counterparts to strongly endorse beliefs such as feeling capable of engaging in physical activity, valuing it for the challenge, or pursuing it for enjoyment and fitness. They are also less likely to report strong feelings of guilt if they do not exercise, to believe they have sufficient opportunities to be active, or to view physical activity as an important way to relax.

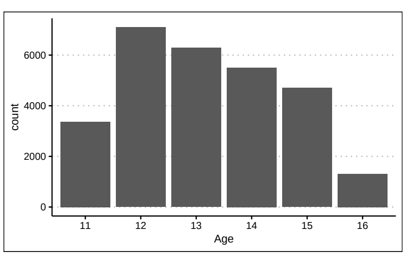
Youths Dataset

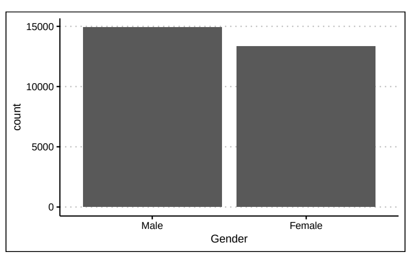
The youth dataset used in this dissertation is drawn from the Active Lives Children and Young People Survey (Year 6: 2022–2023), conducted by Ipsos on behalf of Sport England. It is a large-scale, school-based online survey administered to pupils in Years 1–11, their parents (for Years 1–2 pupils), and teachers. Only pupil responses from year 6-11 (age 11-16) were used in this dissertation. The pupil questionnaire focused on participation in sport and physical activity over the previous week, alongside items on swimming, cycling, volunteering, wellbeing, and attitudes towards physical activity. It also included classification questions such as gender, disability, and long-term health conditions.

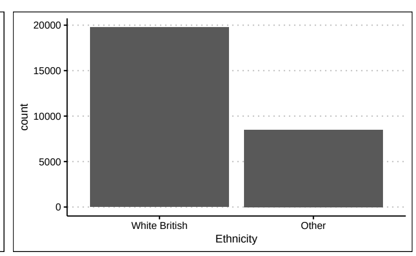
The sampling strategy was designed to permit analysis at both national and local authority levels. A stratified three-stage sampling process was used: schools were first sampled from the January 2021 school census, then three year groups were randomly selected within each participating school, and finally, one mixed-ability class was chosen per selected year group. Fieldwork was carried out in three phases aligned with the academic terms (September 2022–July 2023). Pupils typically completed the survey at school under teacher supervision, although in some secondary schools it could also be set as homework. To encourage participation, schools received credits for sports equipment and, if response thresholds were met, school-level feedback reports.

Variable	Mean	Median	SD	PercentNA
Enjoyment	1.653111	2	0.7188537	9.817977
Social	2.185073	2	0.8672623	42.211906
Fitness	1.780182	2	0.7032159	40.949921
Opportunity	1.612999	2	0.6290499	39.725535
Guilt	2.520805	3	0.9114734	42.623031
Importance	1.413660	1	0.5802182	6.787253
Challenge	1.870674	2	0.7527904	16.347765
Relaxation	2.223833	2	0.9051345	42.034541
Minutes.Exercised	426.587188	290	427.9877225	1.043753

Each item was rated on a four-point Likert scale ranging from 1 (strongly agree) to 4 (strongly disagree). Similar to the adults dataset, the 3 (disagree) and 4 (stronlgy disagree) responses were collapsed into a single cateogry due to low proportions of those resopness. Missing responses and "I don't know" were removed from the dataset.

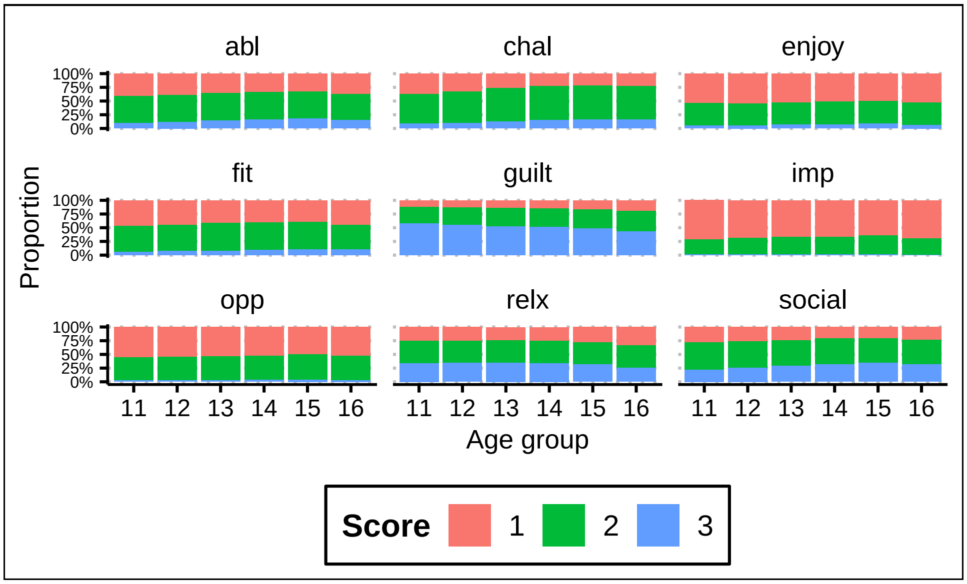






52.79% are male, while 47.21% are female. 69.92% are white British. Most youths are between 12-15 years old. There are only 4.63% of 16 year olds. Although 16-year-olds make up only about 5% of the youth sample, they were not collapsed with the 15-year-olds. This decision reflects a substantive rather than purely numerical consideration: the transition from 15 to 16 often marks a distinct developmental and social stage (e.g., legal thresholds, end of compulsory schooling, increased autonomy), which may correspond to meaningful differences in PA motives. Retaining this separation allows us to examine whether these turning points are reflected in motivational patterns, even with a relatively small subgroup.

The response patterns among youths are less consistent than those observed in adults. With increasing age, youths become less likely to report exercising for challenge or social reasons. A greater proportion report



exercising because they would feel guilty if they did not, or because it helps them relax. Among the oldest youth group, there is a slight but noticeable reversal: a sudden uptick in the proportion who "strongly agree" that ability and fitness are motivating factors. It is also notable that only a small minority of youths, regardless of age, indicate a lack of understanding about the importance of PA.

Multigroup Structural Equation Modeling (SEM)

Differences in the relationships between self-reported motives and PA levels across youths and adults were examined while controlling for demographic factors (see appendix for code and model specs).

Motivation variables included enjoyment, social, fitness, guilt, and opportunity. To account for differences in Likert scales between adults and youths, all motivation variables were dichotomized into "strongly agree" and "not strongly agree." Demographic covariates included gender, age, and ethnicity. Gender was limited to female and male due to small sample sizes of other categories. Ethnicity was collapsed into White British and Non-White British for similar reasons. Youth participants included only those aged 11 and older who responded to the relevant items. Adult participants were grouped by age ranges (16–34, 35–44, 45–54, 55–64, 65–74, 75+) because exact ages were unavailable. The youngest and oldest two groups were further collapsed to reduce skew and ensure balanced distributions. A cap of 1680 minutes per week was applied to reported PA to minimize the impact of potential data entry errors and extreme values.

Multigroup SEM was used to assess how each motive predicts PA levels, allowing direct comparison of pathway strengths between youths and adults. A freely estimated model was compared to constrained models in which individual or all motive pathways were fixed to equality, enabling evaluation of whether the effects of motives differ across age groups. Differences in the predictive strength of each motive on physical activity minutes were also calculated.

Latent Profile Analysis (LCA)

Latent class analysis (LCA) was conducted separately within the youths and adults groups to explore age-related differences in motivational profiles.

The original Likert-scale responses were retained. Additional predictors capturing similar motivational constructs but worded differently were included (see Appendix B). Motives served as predictors, while ethnicity, gender, age, and education (for adults only) were included as covariates. Ten random starts were used per class model to ensure stable solutions.

The optimal number of classes was determined by evaluating BIC elbow plots, relative entropy, bootstrap Vuong-Lo-Mendell-Rubin likelihood ratio tests (BLRT), class proportions, and substantive interpretability. Class-specific statistics were calculated, and multinomial logistic regression was performed with age predicting class membership. Odds ratios and 95% confidence intervals were derived by exponentiating the estimated coefficients and their standard errors (

$$\text{OR} = \exp(\hat{\beta}), \quad 95\%\, CI = \exp\left(\hat{\beta} \pm 1.96 \times SE\right)$$

).

This procedure allows assessment of both the magnitude and statistical significance of age effects on class membership and, consequently, on PA-related motivational profiles.

Residual diagnostics for conditional independence, such as standardized bivariate chi-square or bivariate residuals, were computed for the PA-related items. Some cells in the cross-tabulations of item responses were extremely sparse (proportions <1%), which mathematically inflates these statistics. As a result, the standardized bivariate chi-square values were unrealistically large and cannot serve as reliable indicators of local dependence. Given this, and that the motives (ability, importance, fitness, enjoyment, and opportunity) are substantively distinct, local dependence statistics are not reported. Instead, model evaluation relied on standard LCA fit indices (BIC, aBIC, entropy) and the interpretability and stability of the latent classes, which were robust.

Results

• Present the results of the analysis. • Try to focus on how the results answer your research questions and hypotheses. • Try to focus on substantive interpretation of the results (and not just if something is significant or not). Are the effects large? Are they substantively important?

SEM

There are minor yet significant differences in the impact of every motive.

var	est.youth	est.adult	diff
enjoyb	139.26873	115.33890	23.92983
guiltb	28.74306	11.88358	16.85948
oppb	33.19179	96.44417	-63.25238
fitb	67.36262	92.55927	-25.19665
socialb	32.70297	56.63315	-23.93018
relxb	-59.28077	43.19372	-102.47449

Summary and spec of model in appendix.

All motivational items in the model influenced physical activity in the same direction for youths and adults, with the exception of the relaxation motive, which was negatively associated with exercise in youths but positively in adults. Since the variables are binary (1 = strongly agree, 0 = not strongly agree), the coefficients represent the estimated difference in minutes of activity between participants who strongly endorse a motive and those who do not. While all differences between age groups were statistically significant, effect sizes were generally modest, except for opportunity (greater impact in adults) and relaxation (divergent effect), which showed substantively larger differences. Other motives, including enjoyment, guilt/obligation, fitness, and social factors, exhibited smaller differences, indicating broadly similar motivational mechanisms across age groups.

Youths were most strongly influenced by enjoyment and guilt/obligation, highlighting the role of intrinsic and internalized motivation in shaping adolescent activity. Adults, by contrast, responded more strongly to opportunity, fitness, and social motives, underscoring the importance of external facilitation, health awareness, and social engagement. The relaxation motive's opposite associations suggest that exercise competes with leisure priorities in youths but functions as a stress-relief mechanism for adults.

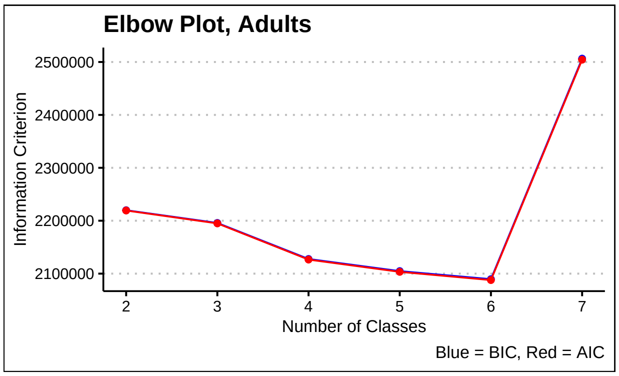
From an SDT perspective, these findings align with the framework's distinction between intrinsic, internalized, and external regulation. Youths' strong responses to enjoyment and guilt/obligation reflect intrinsic and internalized motivations, emphasizing autonomy and internalized norms as key drivers. Adults' stronger responsiveness to opportunity, fitness, and social motives reflects identified and external regulation, where structured opportunities, competence-related goals, and social reinforcement play a larger role. The relaxation motive's divergent effects illustrate a developmental shift in how intrinsic and extrinsic motives operate: adolescents may view exercise as competing with leisure, whereas adults use it to satisfy broader self-regulatory needs such as stress reduction.

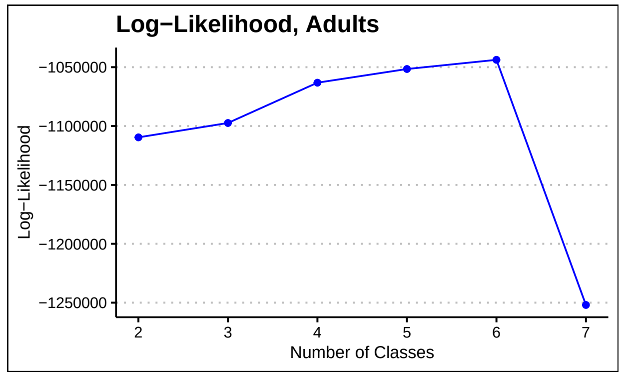
Within the TPB framework, behavior is determined by intention, which is influenced by attitudes, subjective norms, and perceived behavioral control (PBC). For youths, full endorsement of enjoyment and guilt/obligation likely reflects favorable attitudes and perceived social norms supporting activity. For adults, stronger effects of opportunity, fitness, and social motives suggest that PBC, shaped by access to resources and social support, is a key driver of activity. The relaxation motive's opposite effects indicate that perceived behavioral control may be constrained by competing priorities in youths, whereas it reinforces positive intentions in adults.

LCA

Median mins per class, rather than mean due to skew in distribution of minutes exercised (right skew), is calculated.

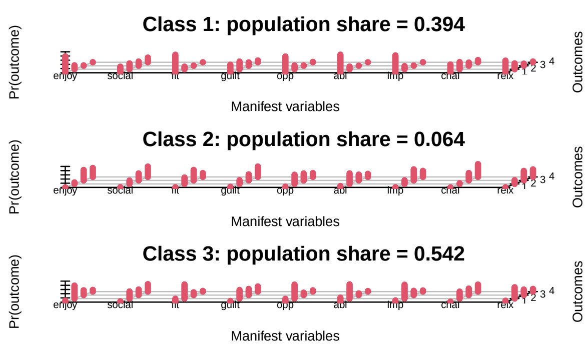
Adults





ADD PLOT? ad.lca.output? add plot(LCAE.ad, nclass = 3) and maybe 4? to appendix

3 to 6 classes all have relative high likelihood and low AIC and BIC, and average posterior entropy per class. From the elbow plot, 4 classes seem to be the number of classes that start displaying diminishing return. However, the 3-class model has significantly higher relative entropy, and the four-class solution produced two classes with highly similar endorsement patterns, effectively splitting the moderate group without adding meaningful differentiation. The three-class model provides a clearer, more parsimonious representation of motivational profiles while still capturing the small but conceptually important subgroup with generally negative attitudes.



Class 1

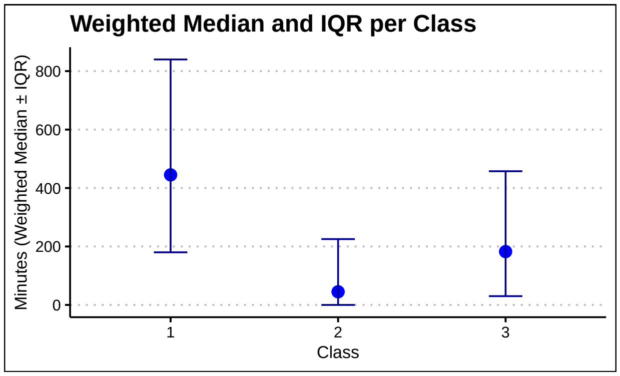
This class shows high probabilities of strongly agreeing with enjoyment, fitness, ability, opportunity, and importance. Members also moderately endorsed social, challenge, and guilt motives, and showed relatively high agreement with relaxation. Overall, this class represents adults who consistently report strong agreement with a broad range of motivations. This class is labelled the High Engagement class.

Class 2

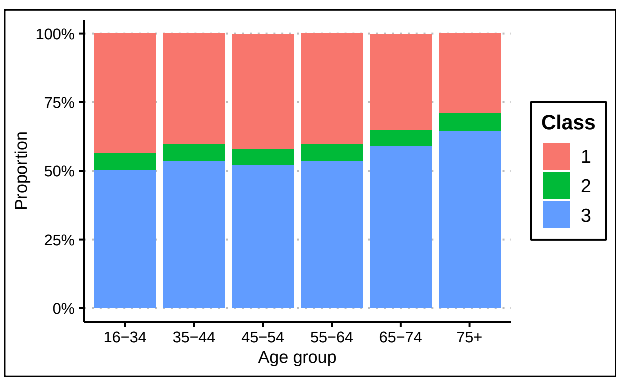
This class is characterized by very low probabilities of strongly agreeing with enjoyment and ability, and by a higher prevalence of neutral or negative responses across most items. Members of this class express the least favorable attitudes toward PA. This class is labelled the Low Engagement class. While this class represents a relatively small proportion of the sample (~6%), it aligns with the raw data, in which very few respondents selected negative responses.

Class 3

This class has a mixed attitude toward many motives, with moderate positive endorsement toward enjoyment, fitness, opportunity, ability and importance It represents adults with more ambivalent responses. This class is labelled the Mixed Motivation class.



Analysis of weekly minutes of physical activity across the adult latent classes highlights substantial differences in exercise behavior. The High Engagement class exhibits the highest levels, with a weighted median of 445 minutes per week and a wide interquartile range of 180 to 840 minutes, indicating considerable variability even among this highly active group. The Mixed Motivation class reports intermediate activity, with a median of 182.5 minutes per week and an IQR of 30 to 457.5 minutes, reflecting both moderate overall engagement and substantial individual differences. In contrast, the Low Engagement class shows markedly lower activity, with a median of only 45 minutes per week and an IQR from 0 to 225.25 minutes, highlighting that while most members are minimally active, some still achieve higher levels. Overall, the broad IQRs across all classes indicate that even within latent classes defined by motivation, actual exercise behavior varies considerably.



The distribution of age groups across the three latent classes highlights notable differences in motivational profiles. The High Engagement class is consistently the second most populous across age groups, though its representation declines gradually with increasing age. The Low Engagement class remains the smallest in all age groups, representing a relatively rare subgroup with low endorsement of motives, and its relative size is largely stable across ages. In contrast, the Mixed Motivation class contains the largest number of adults in every age group, with membership increasing steadily as age rises, suggesting that some individuals may shift from high engagement to mixed motivation over time.

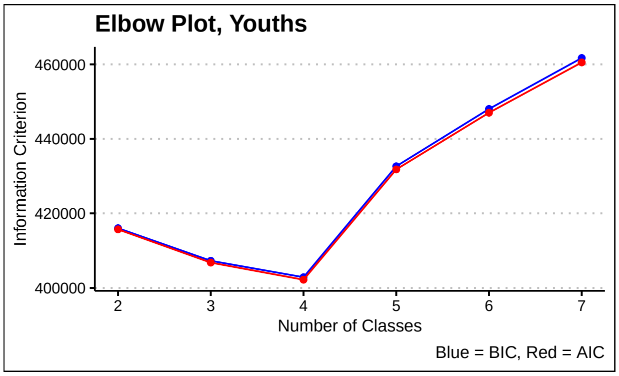
Multinomial logistic regression examined the association between age and motivation profile, with the High Engagement class as reference. Examination of the odds of latent class membership across adult age groups, using the youngest group (16–34 years) as the reference, reveals clear age-related patterns. Adults in the

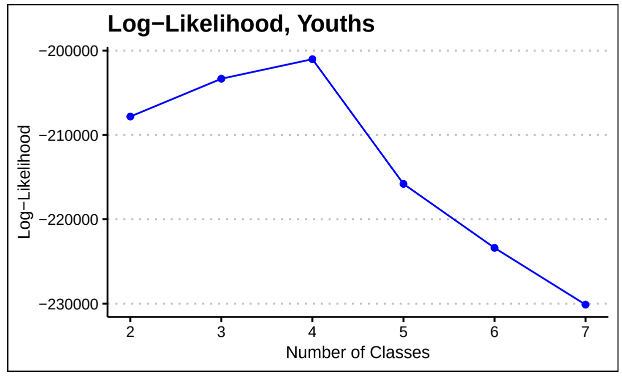
	(Intercept)	age2	age3	age4	age5	age6
2	0.1489715	1.022797	0.9117888	1.046000	1.142746	1.515260
3	1.1534150	1.158554	1.0695943	1.149048	1.453274	1.930015

youngest age group are over 6.6 times more likely to belong to the High Engagement class than to the Low Engagement class. Conversely, a slightly higher proportion of young adults belong to the Mixed Motivation class, which may partly reflect the overall larger size of this class. For the Low Engagement class, older age groups generally show modestly higher odds of membership compared with the reference, though the differences for age groups 35–44 and 55–64 are not statistically significant. The highest odds are observed in the oldest age group (75+), suggesting that adults in this group are 1.5 times more likely to belong to the Low Engagement class than younger adults. In the Mixed Motivation class, the odds of membership increase consistently with age, with adults aged 65–74 being 1.45 times, and those 75+ being 1.93 times more likely than 16–34-year-olds to belong to this class rather than High Engagement.

Youths

BIC plot indicates 3 or four 4 to reduce BIC the most. BLRT preferred 4 classes. Relative entropy values are 0.7998388 and 0.7473211 for 3 and 4 classes, respectively, which shows 3 to have slightly better separation between classes. The likelihood is similar between 3 and 4 classes. Average posterior probabilities are better in the 3-class model, as all classes have a >.80 pp.

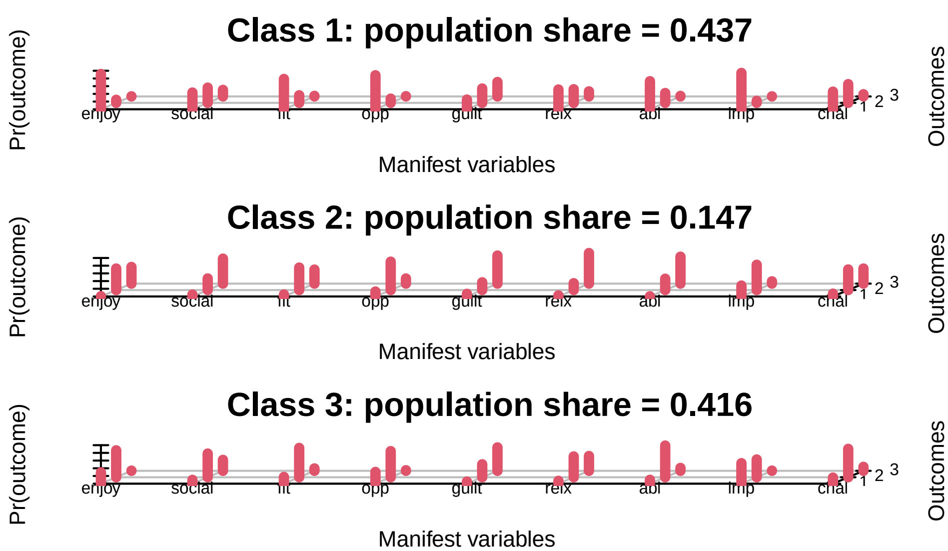




See appendix for elbow plot avereage pp etc

Both 3 and 4- class models show promise. BIC plot does not show a typical elbow shape, as adding more classes beyond 4 actually diminished the fit. This is possibly due to the log-likelihood not increasing significantly with more classes, and BIC's penalty for higher complexity outweighs the improvement

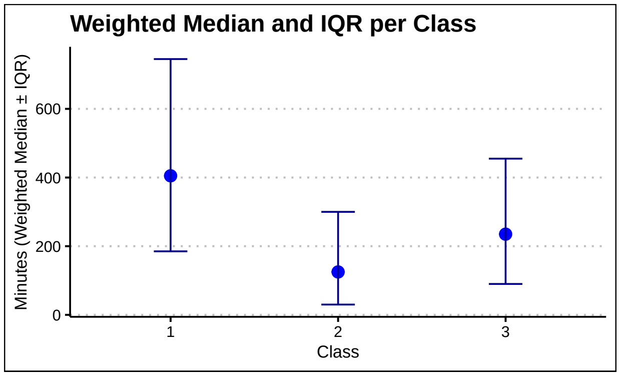
However, the 4-class model contains 2 very similar classes. Hence the 3-class model was chosen.



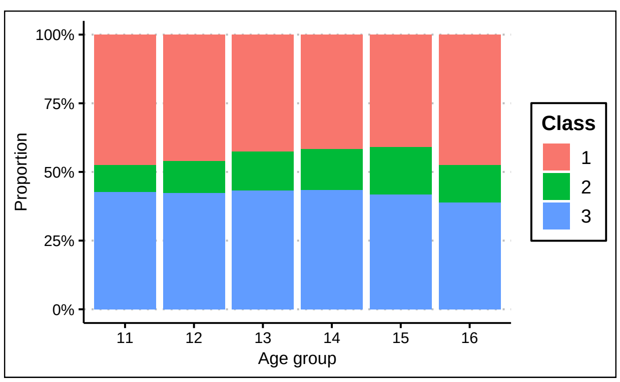
Class 1 This class exhibits very high probabilities of strongly agreeing with enjoyment, fitness, opportunity, importance, and ability. Endorsement of social, challenge, and relaxation motives is somewhat more moderate, though still leaning toward agreement, and guilt shows a more even split across response options. This group is characterized by consistently positive attitudes toward a wide range of motives. This class will be labelled the High Engagement class.

Class 2 This class shows very low probabilities of strongly agreeing with all motives, and negative attitudes toward social, guilt, relaxation, and ability motives. This class will be labelled the Low Engagement class.

Class 3 This class primarily endorsed "agree" across most motives, with the exception of guilt. Disagreement is relatively uncommon, but strong agreement is also less prevalent than in the High Engagement class. This class falls between the other two, representing generally positive but not strongly emphatic motivation toward PA. This class will be labelled the Moderate Engagement class.



Analysis of weekly minutes of physical activity across the youth latent classes reveals meaningful differences in exercise behavior. The High Engagement class reports the highest levels, with a weighted median of 405 minutes per week and an interquartile range of 185 to 745 minutes, indicating substantial variability even among the most motivated youths. The Mixed Motivation class shows intermediate activity, with a median of 235 minutes per week and an IQR from 90 to 455 minutes, reflecting moderate engagement with considerable individual differences. The Low Engagement class exhibits the lowest activity, with a median of 125 minutes per week and an interquartile range of 30 to 300 minutes, suggesting that while most members are less active, some still achieve moderate levels. These patterns imply that, similar to adults, class membership captures meaningful distinctions in motivation that are associated with actual physical activity.



The distribution of age groups across the three latent classes highlights notable differences in motivational profiles among youths. The High Engagement and Mixed Motivation classes are roughly similar in size, with the High Engagement class generally decreasing across age groups but showing a sudden increase at age 16. The Low Engagement class is smaller than the other two but not as rare as in the adult sample; it increases steadily with age, with a slight dip at age 16. These patterns suggest that while high engagement may decline during adolescence, a substantial portion of youths maintain strong or mixed motivation, and low engagement gradually becomes more common over time. The sudden increase in High Engagement at age 16 indicates that some youths may experience a late boost in motivation during this stage of adolescence.

Examination of the odds of latent class membership across youth age groups, using 11-year-olds as the reference, reveals distinct developmental patterns. At age 11, youths are approximately 5 times less likely

	(Intercept)	age2	age3	age4	age5	age6
2	0.2086491	1.197415	1.585630	1.701284	2.035560	1.3616760
3	0.8991138	1.021158	1.129209	1.157742	1.133571	0.9086866

to belong to the Low Engagement class compared with High Engagement, whereas they are about equal in odds to belong to the Mixed Motivation class. For the Low Engagement class, the odds of membership increase steadily with age, with 12-year-olds about 1.2 times more likely, 13-year-olds 1.6 times more likely, 14-year-olds 1.7 times more likely, and 15-year-olds 2 times more likely than 11-year-olds to belong to this class, before slightly declining at age 16 (1.4 times more likely). In the Mixed Motivation class, the odds of membership increase modestly from ages 12–15, ranging from 1.0 to 1.2 times more likely than 11-year-olds, before declining slightly at age 16 (0.9 times as likely). These patterns indicate that as youths age, they increasingly shift from High Engagement toward Low Engagement or Mixed Motivation, with a subtle rise in motivation at age 16 suggesting that some youths regain or consolidate engagement in physical activity.

Conclusions

- Summarise what you have found. Restate your questions and hypotheses and show how you answer them.
- Discuss possible limitations and implications they might have for the results.

Limitations

adult data Push-to-web design: Reliance on online completion may underrepresent groups with lower digital literacy or internet access (e.g., older adults, lower-income households), even though a paper option was provided at the third mailing. Response bias: Out of over 560,000 households invited, only ~173,000 valid responses were obtained. This implies a relatively low response rate, raising the possibility that respondents are more health-conscious or sport-engaged than non-respondents. Household-level sampling: Allowing up to two respondents per household could introduce intra-household clustering, potentially inflating similarities between responses. Mode effects: Differences between online and paper responses could influence reported activity levels or willingness to disclose sensitive demographic details. Split questionnaire design: Some questions were only asked to subsets of participants, which led to the exclusion of data from participants who did not answer key questions required for this analysis.

youths data Self-report data: Pupils' answers, especially younger ones, may be influenced by recall issues, comprehension, or social desirability (e.g., overreporting activity). Mode of administration: Conducted in schools, often under teacher supervision, which may affect how candid pupils are in reporting behaviours or wellbeing.

No factor analysis Wording slightly different between datasets for 2 vars Correlation moderate in sem interpreattion may be diff despite same skewed data (like most self-reports did not log transform minutes, rather just took the median, results are extremely similar and does not detract from comparison removed "does not know" from youths rather than using it as neutral answer only counted moderate to heavy exercises, but for older adults they may only be able to perform and benefit equally from light exercises

demogrpahics like gender, ethnicity, disability, relationship, not considered

motive items limited to only the ones that are worded the same or extremely similar data is cross sectinoal

- Discuss implications for theory and/or policy based on what you found.

References

Ostendorf, D.M., Schmiege, S.J., Conroy, D.E. et al. Motivational profiles and change in physical activity during a weight loss intervention: a secondary data analysis. Int J Behav Nutr Phys Act 18, 158 (2021). https://doi.org/10.1186/s12966-021-01225-5

Moreno-Murcia, J. A., Cervelló Gimeno, E., Hernández, E. H., Pedreño, N. B., & Rodríguez Marín, J. J. (2013). Motivational profiles in physical education and their relation to the theory of planned behavior. Journal of sports science & medicine, 12(3), 551–558.

Nuss, K., Sui, W., Rhodes, R., & Liu, S. (2023). Motivational Profiles and Associations With Physical Activity Before, During, and After the COVID-19 Pandemic: Retrospective Study. JMIR formative research, 7, e43411. https://doi.org/10.2196/43411

Ajzen, Icek (1991). "The theory of planned behavior". Organizational Behavior and Human Decision Processes. 50 (2): 179–211. doi:10.1016/0749-5978(91)90020-T. hdl:11323/1038. S2CID 260959149.

Dishman, R. K., McIver, K. L., Dowda, M., & Pate, R. R. (2018). Declining Physical Activity and Motivation from Middle School to High School. Medicine and science in sports and exercise, 50(6), 1206–1215. https://doi.org/10.1249/MSS.0000000000001542

Varma, V. R., Dey, D., Leroux, A., Di, J., Urbanek, J., Xiao, L., & Zipunnikov, V. (2017). Re-evaluating the effect of age on physical activity over the lifespan. Preventive medicine, 101, 102–108. https://doi.org/10.1016/j.ypmed.2017.05.030

Salvy, S. J., Roemmich, J. N., Bowker, J. C., Romero, N. D., Stadler, P. J., & Epstein, L. H. (2009). Effect of peers and friends on youth physical activity and motivation to be physically active. Journal of pediatric psychology, 34(2), 217–225. https://doi.org/10.1093/jpepsy/jsn071

Biddle, Stuart J. H. , Atkin, Andrew J. , Cavill, Nick and Foster, Charlie(2011) 'Correlates of physical activity in youth: a review of quantitative systematic reviews', International Review of Sport and Exercise Psychology, 4:1, 25 — 49. URL:http://dx.doi.org/10.1080/1750984X.2010.548528

Shao, T., & Zhou, X. (2023). Correlates of physical activity habits in adolescents: A systematic review. Frontiers in physiology, 14, 1131195. https://doi.org/10.3389/fphys.2023.1131195

Choi, J., Lee, M., Lee, Jk. et al. Correlates associated with participation in physical activity among adults: a systematic review of reviews and update. BMC Public Health 17, 356 (2017). https://doi.org/10.1186/s12889-017-4255-2

Alixe H M Kilgour, Matthew Rutherford, Joanna Higson, Samantha J Meredith, Jessica McNiff, Stephanie Mitchell, Anusan Wijayendran, Stephen E R Lim, Susan D Shenkin, Barriers and motivators to undertaking physical activity in adults over 70—a systematic review of the quantitative literature, Age and Ageing, Volume 53, Issue 4, April 2024, afae080, https://doi.org/10.1093/ageing/afae080

Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet (London, England). 2012, 380(9838):219–229.

Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet (London, England). 2012, 380(9838):247–257.

Tapia-Serrano, M. Á., López-Gajardo, M. Á., Sánchez-Miguel, P. A., Llanos-Muñoz, R., & Burgueño, R. (2024). Analysis of motivational profiles of physical activity behavior in primary school students: A self-determination theory-based perspective. Personality and Individual Differences, 231, 112837. https://doi.org/10.1016/j.paid.2024.112837

Sebire, S.J., Jago, R., Fox, K.R. et al. Testing a self-determination theory model of children's physical activity motivation: a cross-sectional study. Int J Behav Nutr Phys Act 10, 111 (2013). https://doi.org/10.1186/1479-5868-10-111

Heredia-León, D. A., Valero-Valenzuela, A., Gómez-Mármol, A., & Manzano-Sánchez, D. (2023). Motivational Profiles in Physical Education: Differences at the Psychosocial, Gender, Age and Extracurricular Sports Practice Levels. Children (Basel, Switzerland), 10(1), 112. https://doi.org/10.3390/children10010112

Geller, K., Renneke, K., Custer, S., & Tigue, G. (2018). Intrinsic and Extrinsic Motives Support Adults' Regular Physical Activity Maintenance. Sports medicine international open, 2(3), E62–E66. https://doi.org/10.1055/a-0620-9137

Brunet, J., & Sabiston, C. M. (2011). Exploring motivation for physical activity across the adult lifespan. Psychology of Sport and Exercise, 12(2), 99–105. https://doi.org/10.1016/j.psychsport.2010.09.006

Oyibo, K., Adaji, I., & Vassileva, J. (2018). Social cognitive determinants of exercise behavior in the context of behavior modeling: a mixed method approach. Digital health, 4, 2055207618811555. https://doi.org/10.1177/2055207618811555

Netz, Yael & Raviv, Shulamith. (2004). Age Differences in Motivational Orientation Toward Physical Activity: An Application of Social—Cognitive Theory. The Journal of psychology. 138. 35-48. 10.3200/JRLP.138.1.35-48.

Beauchamp, M. R., Crawford, K. L., & Jackson, B. (2019). Social cognitive theory and physical activity: Mechanisms of behavior change, critique, and legacy. Psychology of Sport and Exercise, 42, 110–117. https://doi.org/10.1016/j.psychsport.2018.11.009

de Maio Nascimento, M., Gouveia, É. R., Gouveia, B. R., Marques, A., França, C., Campos, P., Martins, F., García-Mayor, J., & Ihle, A. (2023). Differential Patterns in Motivations for Practicing Sport and Their Effects on Physical Activity Engagement across the Lifespan. Healthcare (Basel, Switzerland), 11(2), 274. https://doi.org/10.3390/healthcare11020274

National, regional, and global trends in insufficient physical activity among adults from 2000 to 2022: a pooled analysis of 507 population-based surveys with 5 · 7 million participants. Strain, TessaAbdul Raheem, Raheema et al. The Lancet Global Health, Volume 12, Issue 8, e1232 - e1243

Neipp, Mª & Quiles, María & Leon, Eva & Rodríguez-Marín, Jesús. (2013). Theory of Planned Behavior and physical exercise: Differences between people who do regular physical exercise and those who do not. Wulfenia. 20. 324-335.

Wang, L., & Wang, L. (2015). Using Theory of Planned Behavior to Predict the Physical Activity of Children: Probing Gender Differences. BioMed research international, 2015, 536904. https://doi.org/10.1155/2015/536904

Appendix A - Survey Questions

Adults

Youths

Appendix B - Exercise Types

Adults

Youths

Appendix B - R Code

```
> # Library -------------------------------------------------------------
> set.seed(2025)

> library(tidyverse)

> library(car)

> # Read Data -----------------------------------------------------------
> #
> # data.child <- read.csv('data/child_main.tab', header=T, sep='\t')
> # data.adult <- read.csv('data/adult.tab', header=T, sep='\t')
>
> # Read relevant fields
> # child.var <- data.child %>% select(# likert predictors
>                                     # 'PL_Enjoy_bc_ans', 'PL_Conf_bc_ans',
>                                     # 'PL_Easy_bc_ans', 'PL_GdMe_bc_ans',
>                                     # 'PL_Know_c_ans', 'MO_Opp_c',
>                                     # 'MO_Fit_c', 'MO_Relax_c', 'MO_Fun_c',
>                                     # 'MO_Guilt_c', 'MO_Haveto_b_36',
>                                     # 'MO_Haveto_c_711', 'PR_Fam_c', 'PR_Oth_c',
>                                     # 'Try_bc', 'outdoor_bcd_Overall',
>                                     # 'Exeramt_bc', 'ExeramtMore_bc1_2',
>                                     # 'ExeramtMore_bc2_2', 'ExeramtMore_bc3_2',
>                                     # 'mins_modplus_outschool_Week_ALL',
>                                     #
>                                     # # demographic
>                                     # 'age_11', 'eth2', 'gend3', 'eth6',
>                                     # 'Disab_All_POP',
>                                     #
>                                     # # binary predictors
>                                     # 'PL_Enjoy_bc_SA_gr2', 'MO_Fun_c_SA',
>                                     # 'MO_Fit_c_SA',
>                                     # 'MO_Guilt_c_SA', 'MO_Opp_c_SA',
>                                     # 'MO_Relax_c_SA'
> # )
>
> # Save to save computation time
> # save(child.var, file = "child.var.RData")
> #
>
> # Same process for adults, different variables
> # adult.var <- data.adult %>% dplyr::select('Motiva_POP','motivb_POP',
> #                                           'motivc_POP','motivd_POP',
> #                                           'motive_POP','READYAB1_POP',
> #                                           'READYOP1_POP','motivex2a',
> #                                           'motivex2b','motivex2c',
> #                                           'motivex2d','inclus_a',
> #                                           'inclus_b','inclus_c',
> #                                           'indev','indevtry',
> #                                           'workactlvl',
> #                                           'DUR_HVY_CAPPED_SPORTCOUNT_A01',
```

```
> #                                           'DUR_MOD_CAPPED_SPORTCOUNT_A01',
> #
> #                                           # demographic
> #                                           'Age17','Age3','AgeTGC',
> #                                           'Age4','Age5','Age5_2',
> #                                           'Age9','Disab2_POP',
> #                                           'Gend3','Eth2','Eth7',
> #                                           'Educ6',
> #
> #                                           # binary predictors
> #                                           'Motiva_POP_GR2', 'motivex2c_GR2',
> #                                           'motivex2a_GR2', 'motivc_POP_GR2',
> #                                           'READYOP1_POP_GR2','motivex2b_GR2')
> #
> # save(adult.var, file = "adult.var.RData")
>
> # Basic Distributions and Stats -------------------------------------------
>
> load("child.var.RData")

> load("adult.var.RData")

> glimpse(child.var)
Rows: 122,347
Columns: 32
$ PL_Enjoy_bc_ans                 <int> 4, 1, 2, 2, 1, 5, 1, 4, 2, 1, 2, 1, 1,~
$ PL_Conf_bc_ans                  <int> 4, 1, 2, 3, 1, 2, 1, 2, 1, 1, 2, 2, 2,~
$ PL_Easy_bc_ans                  <int> 4, 2, 2, 3, 2, 3, 2, 2, 2, 1, 5, 3, 3,~
$ PL_GdMe_bc_ans                  <int> 1, 1, 2, 2, 1, 1, 1, 2, 5, 1, 2, 1, 2,~
$ PL_Know_c_ans                   <int> 2, 2, 2, 2, 1, 2, 1, -98, -98, -98, -9~
$ MO_Opp_c                        <int> 1, 2, 2, 2, 1, 2, 1, -98, -98, -98, -9~
$ MO_Fit_c                        <int> 99, 1, 2, 3, 2, 2, 1, -98, -98, -98, -~
$ MO_Relax_c                      <int> 3, 1, 3, 3, 2, 3, 1, -98, -98, -98, -9~
$ MO_Fun_c                        <int> 4, 2, 3, 2, 3, 3, 3, -98, -98, -98, -9~
$ MO_Guilt_c                      <int> 4, 1, 2, 3, 1, 4, 2, -98, -98, -98, -9~
$ MO_Haveto_b_36                  <int> -98, -98, -98, -98, -98, -98, -98, 1, ~
$ MO_Haveto_c_711                 <int> 2, 4, 3, 3, 3, 2, 4, -98, -98, -98, -9~
$ PR_Fam_c                        <int> 4, 3, 2, 3, 3, 2, 3, -91, -91, -91, -9~
$ PR_Oth_c                        <int> 2, 5, 2, 2, 3, 2, 3, -91, -91, -91, -9~
$ Try_bc                          <int> 5, 1, 2, 3, 2, 1, 1, 2, 2, 2, 2, 1, 2,~
$ outdoor_bcd_Overall             <int> 3, 3, 3, 2, 3, 3, 3, -98, -98, -98, -9~
$ Exeramt_bc                      <int> 1, 2, 1, 1, 1, 1, 1, 3, 1, 1, 3, 1, 1,~
$ ExeramtMore_bc1_2               <int> 1, -98, 0, 1, 0, 0, 0, -98, 1, 1, -98,~
$ ExeramtMore_bc2_2               <int> 0, -98, 0, 0, 0, 1, 1, -98, 1, 1, -98,~
$ ExeramtMore_bc3_2               <int> 0, -98, 1, 0, 1, 0, 0, -98, 0, 0, -98,~
$ mins_modplus_outschool_Week_ALL <int> 330, -96, 90, 60, 0, 95, 490, 0, 840, ~
$ age_11                          <int> 12, 12, 12, 13, 12, 13, 13, 10, 10, 9,~
$ eth2                            <int> 2, 2, 2, 1, 2, 3, 1, 2, 2, 2, 1, 3, 3,~
$ gend3                           <int> 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 3, 1, 2,~
$ eth6                            <int> 3, 3, 3, 1, 2, 7, 1, 5, 3, 4, 1, 7, 7,~
$ Disab_All_POP                   <int> 2, 3, 3, 2, 2, 2, 2, 1, 1, 2, 4, 2, 2,~
$ PL_Enjoy_bc_SA_gr2              <int> 2, 1, 2, 2, 1, 99, 1, 2, 2, 1, 2, 1, 1~
$ MO_Fun_c_SA                     <int> 2, 2, 2, 2, 2, 2, 2, -98, -98, -98, -9~
$ MO_Fit_c_SA                     <int> 99, 1, 2, 2, 2, 2, 1, -98, -98, -98, -~
```

```
$ MO_Guilt_c_SA                <int> 2, 1, 2, 2, 1, 2, 2, -98, -98, -98, -9~
$ MO_Opp_c_SA                  <int> 1, 2, 2, 2, 1, 2, 1, -98, -98, -98, -9~
$ MO_Relax_c_SA                <int> 2, 1, 2, 2, 2, 2, 1, -98, -98, -98, -9~

> glimpse(adult.var)
Rows: 172,968
Columns: 37
$ Motiva_POP                  <int> 1, 3, 2, 1, -95, -98, 2, 5, 2, 2, 1, 2, ~
$ motivb_POP                  <int> 1, 2, 2, 2, 3, 2, 2, 3, 2, 3, 2, 3, 1, 1~
$ motivc_POP                  <int> 2, -95, -98, 2, 3, 2, 2, -99, 3, 4, 3, 3~
$ motivd_POP                  <int> 3, 5, 4, 2, 3, -98, 5, -99, 3, 3, 5, 3, ~
$ motive_POP                  <int> -98, -99, -98, -98, -99, -98, -99, -99, ~
$ READYAB1_POP                <int> 1, -95, 2, 2, 3, -95, 2, 2, 1, 2, 1, 2, ~
$ READYOP1_POP                <int> 1, 5, 2, 2, 3, -95, 2, 2, 2, 2, 1, 2, 1,~
$ motivex2a                   <int> 1, 2, 2, 2, 3, 1, 2, 2, 3, 2, 1, 3, 1, 1~
$ motivex2b                   <int> 1, 3, 2, 2, 3, 2, 2, 2, 3, 3, 2, 3, 1, 2~
$ motivex2c                   <int> 2, 3, -95, 2, 3, 4, 2, 3, 3, 2, 1, 2, 3,~
$ motivex2d                   <int> 2, 3, 2, 2, 3, -95, 4, 2, 3, 3, 3, 3, 2,~
$ inclus_a                    <int> 1, -98, -95, 2, -98, 4, -98, -98, 3, 2, ~
$ inclus_b                    <int> 2, -98, 2, 2, -98, -98, -98, -98, 4, 2, ~
$ inclus_c                    <int> 2, -98, -95, 2, -98, -95, -98, -98, 4, 2~
$ indev                       <int> 5, -98, 4, 1, -98, 4, -98, -98, -98, -98~
$ indevtry                    <int> 4, -98, 3, 4, -98, 4, -98, -98, -98, -98~
$ workactlvl                  <int> -98, -98, 1, 2, -98, 2, -98, -98, -98, 2~
$ DUR_HVY_CAPPED_SPORTCOUNT_A01 <dbl> 0, 0, 0, 0, 0, 210, 0, 0, 0, 0, 0, 180, ~
$ DUR_MOD_CAPPED_SPORTCOUNT_A01 <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00~
$ Age17                       <int> 10, 11, 2, 3, 9, 6, 10, 15, 12, 10, 7, 4~
$ Age3                        <int> 3, 3, 1, 1, 3, 2, 3, 3, 3, 3, 2, 1, 3, 2~
$ AgeTGC                      <int> 3, 3, 1, 1, 2, 2, 3, 3, 3, 2, 2, 1, 2, 2~
$ Age4                        <int> 3, 3, 1, 1, 3, 2, 3, 4, 3, 3, 2, 1, 3, 2~
$ Age5                        <int> 4, 5, 2, 3, 4, 3, 4, 5, 5, 4, 4, 3, 4, 4~
$ Age5_2                      <int> 5, 5, 1, 2, 5, 3, 5, 5, 5, 5, 4, 2, 5, 4~
$ Age9                        <int> 6, 7, 2, 3, 6, 4, 6, 9, 7, 6, 5, 3, 6, 5~
$ Disab2_POP                  <int> 2, 1, 2, 2, 1, -94, 2, 1, 2, 2, 2, 2, 2,~
$ Gend3                       <int> 1, 1, 2, 1, 2, 2, 1, 1, 2, 1, 2, 2, 1, 1~
$ Eth2                        <int> 2, 1, 2, -94, 1, 2, 1, 2, 2, 1, 1, 1, 1,~
$ Eth7                        <int> 2, 1, 3, -94, 1, 2, 1, 4, 3, 1, 1, 1, 1,~
$ Educ6                       <int> 1, 6, 3, 3, 6, 1, 1, 6, 6, 1, 1, 2, 1, 2~
$ Motiva_POP_GR2              <int> 1, 0, 0, 1, -95, -98, 0, 0, 0, 0, 1, 0, ~
$ motivex2c_GR2               <int> 0, 0, -95, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,~
$ motivex2a_GR2               <int> 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1~
$ motivc_POP_GR2              <int> 0, -95, -98, 0, 0, 0, 0, -99, 0, 0, 0, 0~
$ READYOP1_POP_GR2            <int> 1, 0, 0, 0, 0, -95, 0, 0, 0, 0, 1, 0, 1,~
$ motivex2b_GR2               <int> 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0~

> # ethnicity
> prop.table(table(adult.var$Eth7))

> prop.table(table(child.var$eth6))

> # 2 is no disa
> table(child.var$Disab_All_POP)

> table(adult.var$Disab2_POP)
```

```
> # adult in bands of 5 years, child just in years
> table(child.var$age_11)

> table(adult.var$Age19plus)

> # too few transgendered adults, filter out
> table(adult.var$gend2_GR6)

> table(adult.var$indevtry)

> table(adult.var$motive_POP)

> # Clean Data for SEM -------------------------------------------------------
>
>
> child.bi <- child.var %>%
+   filter(Disab_All_POP == 2, # remove disabled and no answer
+          gend3 %in% c(1,2),
+          eth2 %in% c(1,2),
+
+          if_all(c(age_11, mins_modplus_outschool_Week_ALL), ~ .x > -1),
+
+          if_all(c(PL_Enjoy_bc_SA_gr2, MO_Fun_c_SA, MO_Fit_c_SA,
+                   MO_Guilt_c_SA, MO_Opp_c_SA, MO_Relax_c_SA), ~ .x > -1 & .x < 3)) %>%
+
+   dplyr::select(enjoyb=PL_Enjoy_bc_SA_gr2,
+                 socialb=MO_Fun_c_SA,
+                 fitb=MO_Fit_c_SA,
+                 guiltb=MO_Guilt_c_SA,
+                 oppb=MO_Opp_c_SA,
+                 relxb=MO_Relax_c_SA,
+
+                 gender=gend3,
+                 age=age_11,
+                 eth=eth2,
+                 mins=mins_modplus_outschool_Week_ALL
+
+   ) %>%
+
+
+   # change 2 (not strongly agree) to 0, consistent with adult
+   mutate(across(c(enjoyb,socialb,fitb,guiltb,oppb), ~ ifelse(.x==2, 0, .x)),
+          gender = gender-1,
+          eth = eth-1,
+          age = age-11)

> adult.bi <- adult.var %>% filter(Disab2_POP==2,
+                                  Gend3 %in% c(1,2),
+                                  Eth2 %in% c(1,2),
+                                  if_all(c(AgeTGC,
+                                           DUR_MOD_CAPPED_SPORTCOUNT_A01,
+                                           DUR_HVY_CAPPED_SPORTCOUNT_A01),
+                                         ~ .x > -1),
```

```
+                                   if_all(c(Motiva_POP_GR2, motivex2c_GR2,
+                                            motivex2a_GR2, motivc_POP_GR2,
+                                            READYOP1_POP_GR2, motivex2b_GR2),
+                                          ~ .x %in% c(0,1))) %>%
+
+
+   mutate(mins=DUR_MOD_CAPPED_SPORTCOUNT_A01 +
+            DUR_HVY_CAPPED_SPORTCOUNT_A01,
+          Gend3 = Gend3-1,
+          Eth2 = Eth2-1,
+          age = case_when(Age9==2~3L,
+                          Age9==9~8L,
+                          TRUE~as.integer(Age9)),
+          age=as.integer(age-3)
+   ) %>%
+
+
+   dplyr::select(enjoyb=Motiva_POP_GR2,
+                 socialb=motivex2c_GR2,
+                 fitb=motivex2a_GR2,
+                 guiltb=motivc_POP_GR2,
+                 oppb=READYOP1_POP_GR2,
+                 relxb=motivex2b_GR2,
+                 gender=Gend3,
+                 age,
+                 eth=Eth2,
+                 mins
+   )

> dallb <- bind_rows(
+   adult.bi %>% mutate(group = "adult"),
+   child.bi %>% mutate(group = "youth")
+ ) %>%
+   mutate(mins = ifelse(mins > 1680, 1680, mins))

> dallb$gender <- relevel(factor(dallb$gender), ref = "0")

> dallb$eth <- relevel(factor(dallb$eth), ref = "0")

> # Clean Data for LCA ------------------------------------------------------
>
> # # Check if collapsing is necessary
> # child.lik %>% dplyr::select(-max_post,-mins,-age,-eth) %>%
> #   pivot_longer(
> #     cols = everything(),   # or specify your Likert vars if df has other columns
> #     names_to = "Variable",
> #     values_to = "Response"
> #   ) %>%
> #   group_by(Variable, Response) %>%
> #   summarise(n = n(), .groups = "drop_last") %>%
> #   #"drop_last" drops the response variable,
> #   #so that the next part (proportion) does not calculate within each response
> #
> #   mutate(prop = n / sum(n)) %>%
```

```
> #   arrange(Variable, Response) %>% filter(prop < 0.05)
>
> # Check which motive responses need to be collapsed
> prop.table(table(child.var$PL_Enjoy_bc_ans))

> prop.table(table(child.var$MO_Fun_c))

> prop.table(table(child.var$MO_Fit_c))

> prop.table(table(child.var$MO_Opp_c))

> prop.table(table(child.var$MO_Guilt_c))

> prop.table(table(child.var$MO_Relax_c))

> prop.table(table(child.var$PL_Conf_bc_ans))

> prop.table(table(child.var$PL_GdMe_bc_ans))

> prop.table(table(child.var$Try_bc))

> prop.table(table(adult.var$Motiva_POP))

> prop.table(table(adult.var$motivex2c))

> prop.table(table(adult.var$motivex2a))

> prop.table(table(adult.var$motivc_POP))

> prop.table(table(adult.var$READYOP1_POP))

> prop.table(table(adult.var$READYAB1_POP))

> prop.table(table(adult.var$motivb_POP))

> prop.table(table(adult.var$motivex2d))

> prop.table(table(adult.var$motivex2b))

> child.lik <- child.var %>%
+
+   # 1-4, 1=strong agree, 4=strong disagree, 5=can't say
+   dplyr::select(enjoy=PL_Enjoy_bc_ans,
+                 social=MO_Fun_c,
+                 fit=MO_Fit_c,
+                 opp=MO_Opp_c,
+                 guilt=MO_Guilt_c, #99 instead of 5 for "can't say"
+                 relx=MO_Relax_c,
+
+                 abl=PL_Conf_bc_ans,
+                 imp=PL_GdMe_bc_ans,
+                 chal=Try_bc,
+
+                 dsbl=Disab_All_POP,
```

```
+               gender=gend3,
+               age=age_11,
+               eth=eth2,
+               mins=mins_modplus_outschool_Week_ALL
+   ) %>%
+
+   filter(dsbl == 2,
+          gender %in% c(1,2),
+          eth %in% c(1,2),
+          mins > -1,
+          if_all(c(enjoy,social,fit,guilt,opp,imp,chal,relx,abl),
+                 ~ .x > -1 & .x < 5)) %>%
+   mutate(
+     mins = ifelse(mins > 1680, 1680, mins),
+     across(c(enjoy,social,fit,guilt,imp,chal,opp,relx,abl),
+            ~ case_when(.x==4~3L, TRUE ~ as.integer(.x))),
+     age=age-10
+
+
+   ) %>%
+   dplyr::select(-dsbl)

> child.lik.back0 <- child.lik

> adult.lik <- adult.var %>%
+   mutate(mins=DUR_HVY_CAPPED_SPORTCOUNT_A01+DUR_MOD_CAPPED_SPORTCOUNT_A01) %>%
+
+   # 1=strong agree, 5=strong disagree
+   dplyr::select(enjoy=Motiva_POP,
+               social=motivex2c,
+               fit=motivex2a,
+               guilt=motivc_POP,
+               opp=READYOP1_POP,
+
+               abl=READYAB1_POP,
+               imp=motivb_POP,
+               chal=motivex2d,
+               relx=motivex2b,
+
+               dsbl=Disab2_POP,
+               gender=Gend3,
+               age=Age9,
+               eth=Eth2,
+               # edu=Educ6,
+               mins
+   ) %>%
+
+   filter(dsbl==2,
+          if_all(c(gender,eth), ~ .x %in% c(1,2)),
+          if_all(everything(), ~ .x > -1)
+          # edu != 5
+   ) %>%
+
+   mutate(across(c(enjoy,social,fit,guilt,opp,imp,chal,relx,abl),
```

```
+                 ~ case_when(.x==5~4L, TRUE ~ as.integer(.x))),
+           # edu = case_when(edu==6~5L, TRUE~edu),
+           age = as.integer(case_when(age==2~3L,
+                           age==9~8L,
+                           TRUE~as.integer(age)))-2
+
+   ) %>%
+
+
+   dplyr::select(-dsbl)

> adult.lik.back0 <- adult.lik

> # Checks -------------------------------------------------------------------
> # Collinearity
> dallb1 <- dallb %>% dplyr::select(-gender,-eth,-group)

> cor(dallb1, method = "pearson")

> # opp, fit and enjoy have mod corr with each other, others ok
>
> # Check adult lik corr
> cor(child.lik.back0 %>% dplyr::select(-gender,-eth, -age), method = "pearson")

> cor(adult.lik.back0 %>% dplyr::select(-gender,-eth,-age), method = "pearson")

> # Check sparsity of highly correlated (>.05) items
> prop.table(table(child.lik$abl, child.lik$enjoy))

> prop.table(table(adult.lik$fit, adult.lik$enjoy))

> prop.table(table(adult.lik$imp, adult.lik$enjoy))

> prop.table(table(adult.lik$fit, adult.lik$imp))

> prop.table(table(adult.lik$abl, adult.lik$opp))

> child.lik.back <- child.lik

> adult.lik.back <- adult.lik

> # VIF
> vif_model <- lm(mins ~ enjoyb + socialb + fitb + guiltb + oppb + relxb, data = dallb1)

> vif(vif_model)
```

```
> # Libraries ------------------------------------------------------------
> set.seed(2025)

> library(tidyverse)

> library(lavaan)

> # SEM ------------------------------------------------------------------
>
> # Free model
> m0 <- '
+   # Mediators: controlling for age, gender, and ethnicity (group-specific coefficients)
+   enjoyb ~ c(a1_adult, a1_youth)*age + c(g1_adult, g1_youth)*gender + c(e1_adult, e1_youth)*eth
+   guiltb ~ c(a2_adult, a2_youth)*age + c(g2_adult, g2_youth)*gender + c(e2_adult, e2_youth)*eth
+   oppb   ~ c(a3_adult, a3_youth)*age + c(g3_adult, g3_youth)*gender + c(e3_adult, e3_youth)*eth
+   fitb   ~ c(a4_adult, a4_youth)*age + c(g4_adult, g4_youth)*gender + c(e4_adult, e4_youth)*eth
+   socialb~ c(a5_adult, a5_youth)*age + c(g5_adult, g5_youth)*gender + c(e5_adult, e5_youth)*eth
+   relxb~ c(a6_adult, a6_youth)*age + c(g6_adult, g6_youth)*gender + c(e6_adult, e6_youth)*eth
+
+   # Main outcome: motives predicting mins, controlling for demographics (group-specific coefficients)
+   mins ~ c(b1_adult, b1_youth)*enjoyb + c(b2_adult, b2_youth)*guiltb + c(b3_adult, b3_youth)*oppb +
+           c(b4_adult, b4_youth)*fitb + c(b5_adult, b5_youth)*socialb + c(b6_adult, b6_youth)*relxb
+           + c(c_adult, c_youth)*age +
+           c(g7_adult, g7_youth)*gender + c(e7_adult, e7_youth)*eth
+
+ # For Adults
+ indirect_age_enjoyb_adult  := a1_adult * b1_adult
+ indirect_age_guiltb_adult  := a2_adult * b2_adult
+ indirect_age_oppb_adult    := a3_adult * b3_adult
+ indirect_age_fitb_adult    := a4_adult * b4_adult
+ indirect_age_socialb_adult := a5_adult * b5_adult
+ indirect_age_relxb_adult := a6_adult * b6_adult
+ total_age_adult := c_adult + indirect_age_enjoyb_adult + indirect_age_guiltb_adult +
+                    indirect_age_oppb_adult + indirect_age_fitb_adult +
+                    indirect_age_socialb_adult + indirect_age_relxb_adult
+
+ # For Youth
+ indirect_age_enjoyb_youth  := a1_youth * b1_youth
+ indirect_age_guiltb_youth  := a2_youth * b2_youth
+ indirect_age_oppb_youth    := a3_youth * b3_youth
+ indirect_age_fitb_youth    := a4_youth * b4_youth
+ indirect_age_socialb_youth := a5_youth * b5_youth
+ indirect_age_relxb_youth := a6_youth * b6_youth
+ total_age_youth := c_youth + indirect_age_enjoyb_youth + indirect_age_guiltb_youth +
+                   indirect_age_oppb_youth + indirect_age_fitb_youth +
+                   indirect_age_socialb_youth + indirect_age_relxb_youth
+ '

> f0 <- sem(m0, data = dallb, group = "group")


> sem.free <- summary(f0, fit.measures = TRUE, standardized = TRUE)
```

```
> # Constrain all to be equal
> f.con <- sem(m0, dallb, group = "group",
+              group.equal = c("intercepts", "regressions"))


> # Check if significantly different
> f0fcon <- anova(f0, f.con)

> f0fcon

> # Spec one constraint at a time
> m1 <- '
+   # Mediators
+   enjoyb ~ age + gender + eth
+   guiltb ~ age + gender + eth
+   oppb ~ age + gender + eth
+   fitb ~ age + gender + eth
+   socialb ~ age + gender + eth
+   relxb ~ age + gender + eth
+
+   # Main outcome
+   mins ~ c("a1","a1")*enjoyb + guiltb + oppb + fitb + socialb + age + gender + eth + relxb
+ '

> m2 <- '
+   # Mediators
+   enjoyb ~ age + gender + eth
+   guiltb ~ age + gender + eth
+   oppb ~ age + gender + eth
+   fitb ~ age + gender + eth
+   socialb ~ age + gender + eth
+   relxb ~ age + gender + eth
+
+   # Main outcome
+   mins ~ enjoyb + c(a,a)*guiltb + oppb + fitb + socialb + age + gender + eth + relxb
+ '

> m3 <- '
+   # Mediators
+   enjoyb ~ age + gender + eth
+   guiltb ~ age + gender + eth
+   oppb ~ age + gender + eth
+   fitb ~ age + gender + eth
+   socialb ~ age + gender + eth
+   relxb ~ age + gender + eth
+
+   # Main outcome
+   mins ~ enjoyb + guiltb + c(a,a)*oppb + fitb + socialb + age + gender + eth + relxb
+ '

> m4 <- '
+   # Mediators
+   enjoyb ~ age + gender + eth
+   guiltb ~ age + gender + eth
```

```
+   oppb ~ age + gender + eth
+   fitb ~ age + gender + eth
+   socialb ~ age + gender + eth
+   relxb ~ age + gender + eth
+
+   # Main outcome
+   mins ~ enjoyb + guiltb + oppb + c(a,a)*fitb + socialb + age + gender + eth + relxb
+ '

> m5 <- '
+   # Mediators
+   enjoyb ~ age + gender + eth
+   guiltb ~ age + gender + eth
+   oppb ~ age + gender + eth
+   fitb ~ age + gender + eth
+   socialb ~ age + gender + eth
+   relxb ~ age + gender + eth
+
+   # Main outcome
+   mins ~ enjoyb + guiltb + oppb + fitb + c(a,a)*socialb + age + gender + eth + relxb
+ '

> m6 <- '
+   # Mediators
+   enjoyb ~ age + gender + eth
+   guiltb ~ age + gender + eth
+   oppb ~ age + gender + eth
+   fitb ~ age + gender + eth
+   socialb ~ age + gender + eth
+   relxb ~ age + gender + eth
+
+   # Main outcome
+   mins ~ enjoyb + guiltb + oppb + fitb + c(a,a)*relxb + age + gender + eth + socialb
+ '

> # Small eigenvalue close to 0, does not matter
> f1 <- sem(m1, data = dallb, group = "group", meanstructure = TRUE)


> f2 <- sem(m2, data = dallb, group = "group", meanstructure = TRUE)


> f3 <- sem(m3, data = dallb, group = "group", meanstructure = TRUE)


> f4 <- sem(m4, data = dallb, group = "group", meanstructure = TRUE)


> f5 <- sem(m5, data = dallb, group = "group", meanstructure = TRUE)


> f6 <- sem(m6, data = dallb, group = "group", meanstructure = TRUE)
```

```
> # Check all models are significantly different from m0
> anova(f0, f1)

> anova(f0, f2)

> anova(f0, f3)

> anova(f0, f4)

> anova(f0, f5)

> anova(f0, f6)

> # Put slope diff. in a table
> params <- parameterEstimates(f0, standardized = T)

> # filter
> slopes <- params %>%
+   filter(lhs == "mins", op == "~") %>%
+   dplyr::select(var=rhs, group, est, se)

> # filtre more
> slopes.ad <- slopes %>% filter(group == 1) %>%
+   dplyr::select(var, est.adult = est, se.adult = se)

> slopes.ch <- slopes %>% filter(group == 2) %>%
+   dplyr::select(var, est.youth = est, se.youth = se)

> # join!
> slopes.diff <- data.frame()

> slopes.diff <- left_join(slopes.ch, slopes.ad, by = "var")

> # calculate
> slopes.diff <- slopes.diff %>%
+   mutate(
+     diff = est.youth - est.adult
+   ) %>%
+   filter(!var %in% c("gender","eth","age")) %>%
+   dplyr::select(-se.youth, -se.adult)

> slopes.diff
```

```
> # Libraries ---------------------------------------------------------------
> set.seed(2025)

> library(tidyverse)

> library(Hmisc)

> library(ggplot2)

> library(nnet)

> library(tidyLPA)

> library(poLCA)

> library(poLCAExtra)

> # LCA, Youths -------------------------------------------------------------
> child.lik <- child.lik.back

> # Predictors (motives)
> child.lik.y <- (child.lik %>%
+                            dplyr::select(-mins,-age,-gender,-eth))

> child.lik.y <- as.matrix(child.lik.y %>% mutate(across(everything(), as.integer)))

> # Spec formula for LCA
> lca.f.child <- child.lik.y ~ gender + eth

> # Run LCA with 2-7 classes
> # LCAE.ch <- poLCA(lca.f.child, data = child.lik, nclass = 2:7)
> # save(LCAE.ch, file="LCAE.ch.RData")
> load("LCAE.ch.RData")

> # bootstrapped Vuong-Lo-Mendell-Rubin likelihood ratio test
> # blrt.ch <- poLCA.blrt(LCAE.ch,quick = T, nrep=10)
> # save(blrt.ch,file="blrt.ch.RData")
> # load("blrt.ch.RData")
>
>
> # Output
> ch.lca.output <- LCAE.ch$output %>% dplyr::select(nclass,llike,AIC,BIC,
+                                                      Rel.Entropy,LMR,p)

> ch.lca.output

> # check max posterior
> # for(k in 2:4){
> #
> #   child.lik$post <- apply(LCAE.ch$LCA[[k]]$posterior, 1, max)
> #
> #   child.lik$class <- LCAE.ch$LCA[[k]]$predclass
> #
```

```
> #   print(
> #     ggplot(child.lik, aes(x = post, fill = factor(class))) +
> #     geom_histogram(binwidth = 0.05, alpha = 0.7, position = "identity") +
> #     labs(x = "Max Posterior Probability", y = "Count", fill = "Class",
> #          title = paste0(k+1," Classes, Youths")) +
> #     theme_minimal()
> #   )
> #
> #   print(ggplot(child.lik, aes(x = factor(class), y = post)) +
> #     geom_boxplot(fill = "skyblue") +
> #     labs(x = "Class", y = "Max Posterior Probability",
> #          title = paste0(k+1," Classes, Youths")) +
> #     theme_minimal()
> #   )
> # }
>
> # Compare 3 and 4 class average posterior and class prop
> post4.ch <- LCAE.ch$LCA[[3]]$posterior

> class4.ch <- apply(post4.ch, 1, which.max)

> class.size4.ch <- prop.table(table(class4.ch))

> ave.pp4.ch <- sapply(1:ncol(post4.ch), function(k) {
+   inds <- which(class4.ch == k)
+   mean(post4.ch[inds, k])
+ })

> post3.ch <- LCAE.ch$LCA[[2]]$posterior

> class3.ch <- apply(post3.ch, 1, which.max)

> class.size3.ch <- prop.table(table(class3.ch))

> ave.pp3.ch <- sapply(1:ncol(post3.ch), function(k) {
+   inds <- which(class3.ch == k)
+   mean(post3.ch[inds, k])
+ })

> # BEST CLASS decided
> # 3 classes is best
> lca.best.ch <- LCAE.ch$LCA[[2]]

> child.lik$class <- lca.best.ch$predclass

> # child.lik$post <- apply(lca.best.ch$posterior, 1, max)
>
> # Calculate median minutes
> n.classes <- 3

> wmed.ch <- numeric(n.classes)

> wq25.ch <- numeric(n.classes)
```

```
> wq75.ch <- numeric(n.classes)

> for (k in 1:n.classes) {
+
+   q <- wtd.quantile(child.lik$mins,
+                     weights = lca.best.ch$posterior[,k],
+                     probs = c(0.25, 0.5, 0.75))
+   wq25.ch[k] <- q[1]
+   wmed.ch[k] <- q[2]
+   wq75.ch[k] <- q[3]
+ }

> # Regressions
> child.lik$age <- child.lik.back$age

> child.lik$class <- relevel(factor(child.lik$class), ref = "1")

> child.lik$age <- relevel(factor(child.lik$age), ref = "1")

> fit.ch <- multinom(class ~ age,
+                    data = child.lik)
# weights:  21 (12 variable)
initial  value 31056.670788
iter  10 value 28676.812139
final  value 28091.735008
converged

> # odds ratio
> or.ch <- exp(coef(fit.ch))

> or.ch

> sum.fit.ch <- summary(fit.ch)

> se <- sum.fit.ch$standard.errors

> # Coefficients
> coefs.ch <- coef(fit.ch)

> # 95% CI for odds ratios
> ci.l.ch <- exp(coefs.ch - 1.96 * se)

> ci.u.ch <- exp(coefs.ch + 1.96 * se)

> # Odds ratios themselves
> or <- exp(coefs.ch)

> # Combine into a table
> or.ci.ch <- data.frame(
+   CI.lower = round(ci.l.ch, 3),
+   CI.upper = round(ci.u.ch, 3)
+ )

> colnames(or.ci.ch) <- c("Intercept.L", "Age2.L", "Age3.L", "Age4.L",
```

```
+                         "Age5.L","Age6.L","Intercept.U","Age2.U", "Age3.U", "Age4.U",
+                         "Age5.U","Age6.U")

> # Check class distribution per age
>
> tb.byage.ch <- child.lik %>%
+ count(age, class) %>%
+   pivot_wider(names_from = class, values_from = n, values_fill = 0)

> # LCA, Adults ------------------------------------------------------------
>
> adult.lik <- adult.lik.back

> # Predictors (motives)
> adult.lik.y <- as.matrix(adult.lik %>%
+                            dplyr::select(-mins,-age,-gender,-eth))

> # Spec formula for LCA
> lca.f.adult <- adult.lik.y ~ gender + eth

> # LCAE.ad <- poLCA(lca.f.adult, data = adult.lik, nclass = 2:7)
> # save(LCAE.ad, file="LCAE.ad.RData")
> load(file="LCAE.ad.RData")

> # bootstrapped Vuong-Lo-Mendell-Rubin likelihood ratio test
> # blrt.ad <- poLCA.blrt(LCAE.ad, quick = T,nreps = 10)
> # save(blrt.ad,file="blrt.ad.RData")
> # load(file="blrt.ad.RData")
>
>
> # Take relevant stats
> ad.lca.output <- LCAE.ad$output %>% dplyr::select(nclass,llike,AIC,BIC,
+                                                 Rel.Entropy,LMR,p)

> ad.lca.output

> # adeck posterior and boxplots
> # for(k in 2:5){
> #
> #   adult.lik$post <- apply(LCAE.ad$LCA[[k]]$posterior, 1, max)
> #   adult.lik$class <- LCAE.ad$LCA[[k]]$predclass
> #
> #   print(
> #     ggplot(adult.lik, aes(x = post, fill = factor(class))) +
> #       geom_histogram(binwidth = 0.05, alpha = 0.7, position = "identity") +
> #       labs(x = "Max Posterior Probability", y = "Count", fill = "Class",
> #            title = paste0(k+1," Classes, Adults")) +
> #       theme_minimal()
> #   )
> #
> #   print(ggplot(adult.lik, aes(x = factor(class), y = post)) +
> #           geom_boxplot(fill = "skyblue") +
> #           labs(x = "Class", y = "Max Posterior Probability",
> #                title = paste0(k+1," Classes, Adults")) +
```

```
> #           theme_minimal()
> #   )
> # }
>
> # Compare class average posteriors and class prop
>
> post6.ad <- LCAE.ad$LCA[[5]]$posterior

> class6.ad <- apply(post6.ad, 1, which.max)

> class.size6.ad <- prop.table(table(class6.ad))

> ave.pp6.ad <- sapply(1:ncol(post6.ad), function(k) {
+   inds <- which(class6.ad == k)
+   mean(post6.ad[inds, k])
+ })

> ave.pp6.ad

> post5.ad <- LCAE.ad$LCA[[4]]$posterior

> class5.ad <- apply(post5.ad, 1, which.max)

> class.size5.ad <- prop.table(table(class5.ad))

> ave.pp5.ad <- sapply(1:ncol(post5.ad), function(k) {
+   inds <- which(class5.ad == k)
+   mean(post5.ad[inds, k])
+ })

> ave.pp5.ad

> post4.ad <- LCAE.ad$LCA[[3]]$posterior

> class4.ad <- apply(post4.ad, 1, which.max)

> class.size4.ad <- prop.table(table(class4.ad))

> ave.pp4.ad <- sapply(1:ncol(post4.ad), function(k) {
+   inds <- which(class4.ad == k)
+   mean(post4.ad[inds, k])
+ })

> ave.pp4.ad

> post3.ad <- LCAE.ad$LCA[[2]]$posterior

> class3.ad <- apply(post3.ad, 1, which.max)

> class.size3.ad <- prop.table(table(class3.ad))

> ave.pp3.ad <- sapply(1:ncol(post3.ad), function(k) {
+   inds <- which(class3.ad == k)
+   mean(post3.ad[inds, k])
```

```
+ })

> ave.pp3.ad

> # BEST CLASS decided
> # 3 classes is best
> lca.best.ad <- LCAE.ad$LCA[[2]]

> adult.lik$class <- lca.best.ad$predclass

> adult.lik$post <- apply(lca.best.ad$posterior, 1, max)

> # Calculate median minutes
> n.classes <- 3

> wmed.ad <- numeric(n.classes)

> wq25.ad <- numeric(n.classes)

> wq75.ad <- numeric(n.classes)

> for (k in 1:n.classes) {
+
+   q <- wtd.quantile(adult.lik$mins,
+                     weights = lca.best.ad$posterior[,k],
+                     probs = c(0.25, 0.5, 0.75))
+   wq25.ad[k] <- q[1]
+   wmed.ad[k] <- q[2]
+   wq75.ad[k] <- q[3]
+ }


> # Regressions
> adult.lik$age <- adult.lik.back$age

> adult.lik$class <- relevel(factor(adult.lik$class), ref = "1")

> adult.lik$age <- relevel(factor(adult.lik$age), ref = "1")

> fit.ad <- multinom(class ~ age,
+                    data = adult.lik)
# weights:  21 (12 variable)
initial  value 127458.800507
iter  10 value 108640.273108
iter  20 value 100457.045377
iter  20 value 100457.044488
iter  20 value 100457.044435
final  value 100457.044435
converged

> # odds ratio
> or.ad <- exp(coef(fit.ad))

> or.ad
```

```
> sum.fit.ad <- summary(fit.ad)

> se.ad <- sum.fit.ad$standard.errors

> # Coefficients
> coefs.ad <- coef(fit.ad)

> # 95% CI for odds ratios
> ci.l.ad <- exp(coefs.ad - 1.96 * se.ad)

> ci.u.ad <- exp(coefs.ad + 1.96 * se.ad)

> # Combine into a table
> or.ci.ad <- data.frame(
+   CI.lower = round(ci.l.ad, 3),
+   CI.upper = round(ci.u.ad, 3)
+ )

> colnames(or.ci.ad) <- c("Intercept.L", "Age2.L", "Age3.L", "Age4.L",
+                         "Age5.L","Age6.L","Intercept.U","Age2.U", "Age3.U", "Age4.U",
+                         "Age5.U","Age6.U")

> # adeck class distribution per age
>
> tb.byage.ad <- adult.lik %>%
+   count(age, class) %>%
+   pivot_wider(names_from = class, values_from = n, values_fill = 0)
```

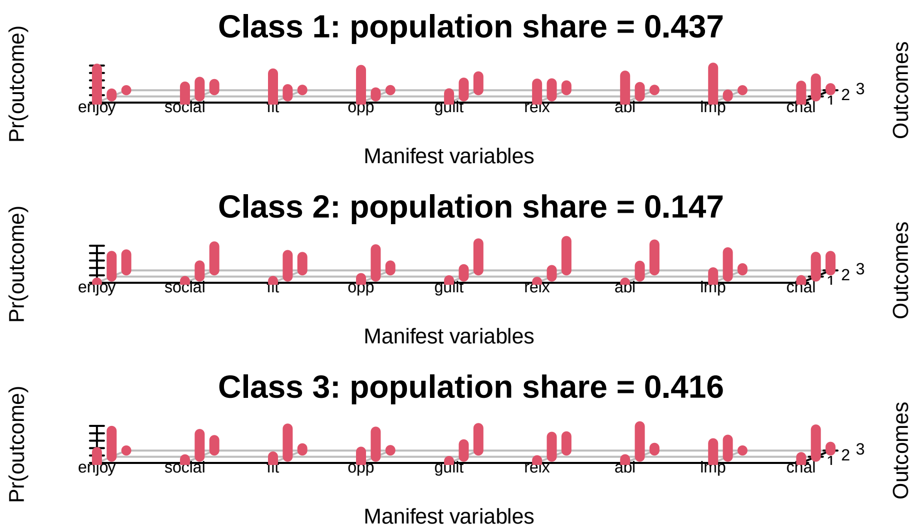
```
> set.seed(2025)

> library(tidyverse)

> library(ggplot2)

> library(poLCA)

> library(poLCAExtra)

> library(scales)

> library(ggthemes)

> # Descriptive ------------------------------------------------------------
> #
> cor.ie <- cor(adult.lik.back0 %>% dplyr::select(-gender,-eth), method = "pearson")[6,1]

> cor.if <- cor(adult.lik.back0 %>% dplyr::select(-gender,-eth), method = "pearson")[6,3]

> cor.imp <- data.frame("Imp,Enjoy"=cor.ie, "Imp,Fit"=cor.if)

> # get summary of all motives
> adult.summary <- adult.var %>%
+   mutate(mins = DUR_HVY_CAPPED_SPORTCOUNT_A01+
+                   DUR_MOD_CAPPED_SPORTCOUNT_A01) %>%
+   dplyr::select(
+     Enjoyment = Motiva_POP,
+     Social = motivex2c,
+     Fitness = motivex2a,
+     Guilt = motivc_POP,
+     Opportunity = READYOP1_POP,
+     Importance = motivb_POP,
+     Challenge = motivex2d,
+     Relaxation = motivex2b,
+     Minutes.Exercised = mins
+   ) %>%
+   summarise(
+     across(everything(),
+            list(
+              Mean = ~mean(.x[.x > 0], na.rm = TRUE),
+              Median = ~median(.x[.x > 0], na.rm = TRUE),
+              SD = ~sd(.x[.x > 0], na.rm = TRUE),
+              PercentNA = ~mean(.x < 0, na.rm = TRUE) * 100
+            ),
+            .names = "{.col}_{.fn}"
+     )
+   ) %>%
+   pivot_longer(everything(), names_to = c("Variable", "Stat"), names_sep = "_") %>%
+   pivot_wider(names_from = Stat, values_from = value)

> child.summary <- child.var %>%
+   dplyr::select(
```

```
+     Enjoyment = PL_Enjoy_bc_ans,
+     Social = MO_Fun_c,
+     Fitness = MO_Fit_c,
+     Opportunity = MO_Opp_c,
+     Guilt = MO_Guilt_c,
+     Importance = PL_GdMe_bc_ans,
+     Challenge = Try_bc,
+     Relaxation = MO_Relax_c
+   ) %>%
+   summarise(
+     across(everything(),
+            list(
+              Mean = ~mean(.x[.x > 0 & .x <= 4], na.rm = TRUE),
+              Median = ~median(.x[.x > 0 & .x <= 4], na.rm = TRUE),
+              SD = ~sd(.x[.x > 0 & .x <= 4], na.rm = TRUE),
+              PercentNA = ~mean(.x < 0 | .x > 4, na.rm = TRUE) * 100
+            ),
+            .names = "{.col}_{.fn}"
+     )
+   ) %>%
+   pivot_longer(everything(), names_to = c("Variable", "Stat"), names_sep = "_") %>%
+   pivot_wider(names_from = Stat, values_from = value)

> c.mins <- child.var %>%
+   summarise(Variable = "Minutes.Exercised",
+             Mean = mean(mins_modplus_outschool_Week_ALL[mins_modplus_outschool_Week_ALL > 0 ], na.rm =
+             Median = median(mins_modplus_outschool_Week_ALL[mins_modplus_outschool_Week_ALL > 0], na.r
+             SD = sd(mins_modplus_outschool_Week_ALL[mins_modplus_outschool_Week_ALL > 0], na.rm = TRUE
+             PercentNA = mean(mins_modplus_outschool_Week_ALL < 0, na.rm = TRUE) * 100)

> child.summary <- rbind(child.summary, c.mins)

> # get demographic overview (gender, edu, eth, mins)
> # adult
> #
> # # Disability
> # gg.ad.dsbl <- ggplot(adult.var, aes(x = as.factor(Disab2_POP))) +
> #   geom_bar() +
> #   labs(x = "Disability") +
> #   theme_clean()
>
> # Gender
> adult.lik$gender <- factor(adult.lik$gender, levels = c(1, 2),
+                            labels = c("Male", "Female"))

> gg.ad.gend <- ggplot(adult.lik, aes(x = as.factor(gender))) +
+   geom_bar() +
+   labs(x = "Gender") +
+   theme_clean()

> # Age
> adult.lik$age <- factor(adult.lik$age, levels = c(1,2,3,4,5,6),
+                         labels = c("16-34", "35-44", "45-54",
+                                    "55-64", "65-74", "75+"))
```

```
> gg.ad.age <-ggplot(adult.lik, aes(x = as.factor(age))) +
+   geom_bar() +
+   labs(x = "Age Group") +
+   theme_clean()

> # Ethnicity
> adult.lik$eth <- factor(adult.lik$eth, levels = c(1, 2),
+                         labels = c("White British", "Other"))

> gg.ad.eth <- ggplot(adult.lik, aes(x = as.factor(eth))) +
+   geom_bar() +
+   labs(x = "Ethnicity") +
+   theme_clean()

> # Education
> # gg.ad.edu <- ggplot(adult.lik, aes(x = as.factor(edu))) +
> #   geom_bar() +
> #   labs(x = "Education") +
> #   theme_clean()
>
> # YOuths
> #
> # Disability
> # gg.ch.dsbl <- ggplot(child.var, aes(x = as.factor(Disab_All_POP))) +
> #   geom_bar() +
> #   labs(x = "Disability") +
> #   theme_clean()
>
> # Gender
> child.lik$gender <- factor(child.lik$gender, levels = c(1, 2),
+                            labels = c("Male", "Female"))

> gg.ch.gend <- ggplot(child.lik, aes(x = as.factor(gender))) +
+   geom_bar() +
+   labs(x = "Gender") +
+   theme_clean()

> # Age
> child.lik$age <- factor(child.lik$age, levels = c(1,2,3,4,5,6),
+                         labels = c(11,12,13,14,15,16))

> gg.ch.age <- ggplot(child.lik, aes(x = as.factor(age))) +
+   geom_bar() +
+   labs(x = "Age") +
+   theme_clean()

> # Ethnicity
> child.lik$eth <- factor(child.lik$eth, levels = c(1, 2),
+                         labels = c("White British", "Other"))

> gg.ch.eth <- ggplot(child.lik, aes(x = as.factor(eth))) +
+   geom_bar() +
+   labs(x = "Ethnicity") +
```

```
+   theme_clean()

> # SEM ----------------------------------------------------------------------
> # # slope_youth - slope_adult, pooled sd
> # cohen <- rbind(cohen.enj, cohen.soc, cohen.fit,cohen.glt,cohen.opp)
> # rownames(cohen) <- c("Enjoy", "Social", "Fit","Guilt","Opp")
> # colnames(cohen) <- c("Std Eff", "Min")
> # cohen
> # LCA Youths----------------------------------------------------------------
>
> # elbow plot
> gg.elbow.ch <- ggplot(ch.lca.output, aes(x = nclass)) +
+   geom_line(aes(y = BIC), color = "blue") +
+   geom_point(aes(y = BIC), color = "blue") +
+   geom_line(aes(y = AIC), color = "red") +
+   geom_point(aes(y = AIC), color = "red") +
+   labs(y = "Information Criterion", x = "Number of Classes",
+        title = "Elbow Plot, Youths",
+        caption = "Blue = BIC, Red = AIC") +
+   theme_clean()

> gg.elbow.ch

> gg.llik.ch <- ggplot(ch.lca.output, aes(x = nclass)) +
+   geom_line(aes(y = llike), color = "blue") +
+   geom_point(aes(y = llike), color = "blue") +
+   labs(y = "Log-Likelihood", x = "Number of Classes",
+        title = "Log-Likelihood, Youths") +
+   theme_clean()

> gg.llik.ch

> #
> # # Max posterior
> # gg.post.his.ch <- ggplot(child.lik, aes(x = post, fill = factor(class))) +
> #   geom_histogram(binwidth = 0.05, alpha = 0.7, position = "identity") +
> #   labs(x = "Max Posterior Probability", y = "Count", fill = "Class",
> #        title = paste0(k," Classes, Youths")) +
> #   theme_clean()
> # gg.post.his.ch
> #
> # # Boxplot
> # gg.post.box.ch <- ggplot(child.lik, aes(x = factor(class), y = post)) +
> #   geom_boxplot(fill = "skyblue") +
> #   labs(x = "Class", y = "Max Posterior Probability",
> #        title = paste0(k," Classes, Youths")) +
> #   theme_clean()
>
>
> # class,size/proportion, average pp,entropy
>
> tb.class3.ch <- data.frame(
+   Class = 1:ncol(post3.ch),
+   Proportion = as.numeric(class.size3.ch),
```

```
+   Avg_Posterior = round(ave.pp3.ch, 3)
+ )

> tb.class3.ch

> tb.class4.ch <- data.frame(
+   Class = 1:ncol(post4.ch),
+   Proportion = as.numeric(class.size4.ch),
+   Avg_Posterior = round(ave.pp4.ch, 3)
+ )

> # Weighted minutes, youths
> mins.child <- data.frame(
+   Class = 1:n.classes,
+   Weighted.Median = wmed.ch,
+   Weighted.Q25 = wq25.ch,
+   Weighted.Q75 = wq75.ch
+ )

> mins.child

> gg.mins.ch <- ggplot(mins.child, aes(x = factor(Class), y = Weighted.Median)) +
+   geom_point(size = 3, color = "blue") +                 # median as a point
+   geom_errorbar(aes(ymin = Weighted.Q25, ymax = Weighted.Q75),
+                 width = 0.2, color = "darkblue") +      # IQR as error bars
+   labs(x = "Class", y = "Minutes (Weighted Median ± IQR)",
+        title = "Weighted Median and IQR per Class") +
+   theme_clean()

> gg.mins.ch

> gg.med.ch <- ggplot(mins.child, aes(x = Class, y = Weighted.Median)) +
+   geom_col() +
+   labs(x = "Latent Class", y = "Probability-Weighted Median Minutes")

> gg.med.ch

> # Predictor plot
> plot(LCAE.ch, nclass = 2)
```

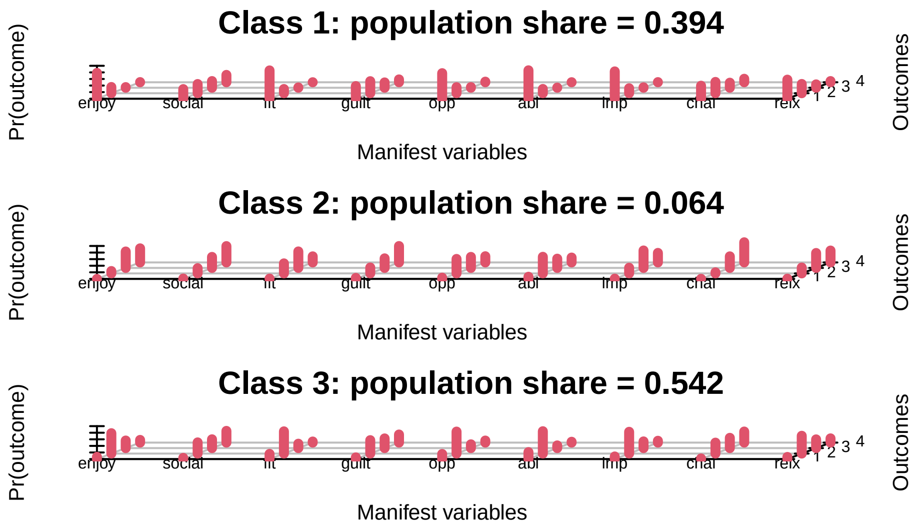



```
> # Bootstrap Vuong-Lo-Mendell-Rubin Likelihood Ratio Test
> or.ch

> # Appendix
> or.ci.ch

> # Include actual coeffs in appendix
> lca.best.ch$probs

> tb.byage.ch

> gg.byage.ch <- child.lik %>%
+   dplyr::count(age, class) %>%
+   group_by(age) %>%
+   mutate(prop = n / sum(n)) %>%
+   ggplot(aes(x = factor(age), y = prop, fill = factor(class))) +
+   geom_col() +
+   labs(x = "Age group", y = "Proportion", fill = "Class") +
+   scale_y_continuous(labels = scales::percent_format()) +
+   theme_clean()

> gg.byage.ch

> vars.ch <- setdiff(names(child.lik), c("age","mins","post","class",
+                                        "gender","eth","edu"))

> child.lik_long <- child.lik %>%
+   pivot_longer(cols = all_of(vars.ch), names_to = "variable", values_to = "score") %>%
+   count(age, variable, score) %>%
+   group_by(age, variable) %>%
+   mutate(prop = n / sum(n))

> gg.vars.ch <- ggplot(child.lik_long, aes(x = factor(age), y = prop, fill = factor(score))) +
+   geom_col() +
```

```
+   facet_wrap(~variable, nrow = 3, ncol = 3) +
+   labs(x = "Age group", y = "Proportion", fill = "Score") +
+   scale_y_continuous(labels = percent_format()) +
+   theme_clean() +
+   theme(legend.position = "bottom", axis.text.y = element_text(size = 6))

> # LCA Adults --------------------------------------------------------------
>
>
> # elbow plot
> gg.elbow.ad <- ggplot(ad.lca.output, aes(x = nclass)) +
+   geom_line(aes(y = BIC), color = "blue") +
+   geom_point(aes(y = BIC), color = "blue") +
+   geom_line(aes(y = AIC), color = "red") +
+   geom_point(aes(y = AIC), color = "red") +
+   labs(y = "Information Criterion", x = "Number of Classes",
+        title = "Elbow Plot, Adults",
+        caption = "Blue = BIC, Red = AIC") +
+   theme_clean()

> gg.elbow.ad

> gg.llik.ad <- ggplot(ad.lca.output, aes(x = nclass)) +
+   geom_line(aes(y = llike), color = "blue") +
+   geom_point(aes(y = llike), color = "blue") +
+   labs(y = "Log-Likelihood", x = "Number of Classes",
+        title = "Log-Likelihood, Adults") +
+   theme_clean()

> gg.llik.ad

> # # Max posterior
> # gg.post.his.ad <- ggplot(adult.lik, aes(x = post, fill = factor(class))) +
> #   geom_histogram(binwidth = 0.05, alpha = 0.7, position = "identity") +
> #   labs(x = "Max Posterior Probability", y = "Count", fill = "Class",
> #        title = paste0(k," Classes, Adults")) +
> #   theme_clean()
> # gg.post.his.ad
> #
> # # Boxplot
> # gg.post.box.ad <- ggplot(adult.lik, aes(x = factor(class), y = post)) +
> #   geom_boxplot(fill = "skyblue") +
> #   labs(x = "Class", y = "Max Posterior Probability",
> #        title = paste0(k," Classes, Adults")) +
> #   theme_clean()
>
>
> # class,size/proportion, average pp,entropy
>
> tb.class3.ad <- data.frame(
+   Class = 1:ncol(post3.ad),
+   Proportion = as.numeric(class.size3.ad),
+   Avg_Posterior = round(ave.pp3.ad, 3)
+ )
```

```
> tb.class3.ad

> tb.class4.ad <- data.frame(
+   Class = 1:ncol(post4.ad),
+   Proportion = as.numeric(class.size4.ad),
+   Avg_Posterior = round(ave.pp4.ad, 4)
+ )

> tb.class4.ad

> tb.class5.ad <- data.frame(
+   Class = 1:ncol(post5.ad),
+   Proportion = as.numeric(class.size5.ad),
+   Avg_Posterior = round(ave.pp5.ad, 5)
+ )

> tb.class5.ad

> tb.class6.ad <- data.frame(
+   Class = 1:ncol(post6.ad),
+   Proportion = as.numeric(class.size6.ad),
+   Avg_Posterior = round(ave.pp6.ad, 6)
+ )

> tb.class6.ad

> mins.adult <- data.frame(
+   Class = 1:n.classes,
+   Weighted.Median = wmed.ad,
+   Weighted.Q25 = wq25.ad,
+   Weighted.Q75 = wq75.ad
+ )

> mins.adult

> gg.mins.ad <- ggplot(mins.adult, aes(x = factor(Class), y = Weighted.Median)) +
+   geom_point(size = 3, color = "blue") +                 # median as a point
+   geom_errorbar(aes(ymin = Weighted.Q25, ymax = Weighted.Q75),
+                 width = 0.2, color = "darkblue") +      # IQR as error bars
+   labs(x = "Class", y = "Minutes (Weighted Median ± IQR)",
+        title = "Weighted Median and IQR per Class") +
+   theme_clean()

> gg.mins.ad

> #
> # # Weighted minutes, youths
> # gg.med.ad <- ggplot(mins.adult, aes(x = Class, y = Weighted.Median)) +
> #   geom_col() +
> #   labs(x = "Latent Class", y = "Probability-Weighted Median Minutes")
>
>
> # Predictor plot
```

```
> plot(LCAE.ad, nclass = 2)
```




```
> # plot(LCAE.ad, nclass = 3)
>
> # Bootstrap Vuong-Lo-Mendell-Rubin Likelihood Ratio Test
> # 100 reps
> # blrt.ad
> or.ad

> or.ci.ad

> # Include actual coeffs in appendix
> lca.best.ad$probs

> tb.byage.ad

> gg.byage.ad <- adult.lik %>%
+   dplyr::count(age, class) %>%
+   group_by(age) %>%
+   mutate(prop = n / sum(n)) %>%
+   ggplot(aes(x = factor(age), y = prop, fill = factor(class))) +
+   geom_col() +
+   labs(x = "Age group", y = "Proportion", fill = "Class") +
+   scale_y_continuous(labels = scales::percent_format()) +
+   theme_clean()

> gg.byage.ad

> vars.ad <- setdiff(names(adult.lik), c("age","mins","post","class",
+                                          "gender","eth","edu"))

> adult.lik_long <- adult.lik %>%
+   pivot_longer(cols = all_of(vars.ad), names_to = "variable", values_to = "score") %>%
```

```
+   count(age, variable, score) %>%
+   group_by(age, variable) %>%
+   mutate(prop = n / sum(n))

> gg.vars.ad <- ggplot(adult.lik_long, aes(x = factor(age), y = prop, fill = factor(score))) +
+   geom_col() +
+   facet_wrap(~variable, nrow = 3, ncol = 3) +
+   labs(x = "Age group", y = "Proportion", fill = "Score") +
+   scale_y_continuous(labels = percent_format()) +
+   theme_clean() +
+   theme(legend.position = "bottom", axis.text.y = element_text(size = 6))

> gg.vars.ad
```

Appendix C - R code: rmarkdown chunks

```
source(file="data.R")
source(file="SEM.R")
source(file="LCA.R")
source(file="Visualization.R")
opts <- options(knitr.kable.NA = '')
kable(list(adult.summary),align='c',booktabs = T) %>%
  kable_styling(position = 'center', font_size = 10,
                latex_options = c('striped',
                                  'hold_position'))
gg.ad.age
gg.ad.gend
gg.ad.eth

gg.vars.ad

opts <- options(knitr.kable.NA = '')
kable(list(child.summary),align='c',booktabs = T) %>%
  kable_styling(position = 'center', font_size = 10,
                latex_options = c('striped',
                                  'hold_position'))

gg.ch.age
gg.ch.gend
gg.ch.eth


gg.vars.ch

opts <- options(knitr.kable.NA = '')
kable(list(slopes.diff),align='c',booktabs = T) %>%
  kable_styling(position = 'center', font_size = 10,
                latex_options = c('striped',
                                  'hold_position'))

gg.elbow.ad
gg.llik.ad


plot(LCAE.ad, nclass = 2)


gg.mins.ad


gg.byage.ad

opts <- options(knitr.kable.NA = '')
kable(list(or.ad),align='c',booktabs = T) %>%
  kable_styling(position = 'center', font_size = 10,
                latex_options = c('striped',
                                  'hold_position'))
```

```r
gg.elbow.ch
gg.llik.ch


plot(LCAE.ch, nclass = 2)


gg.mins.ch


gg.byage.ch

opts <- options(knitr.kable.NA = '')
kable(list(or.ch),align='c',booktabs = T) %>%
  kable_styling(position = 'center', font_size = 10,
                latex_options = c('striped',
                                  'hold_position'))
source("data.R", echo = T, print.eval = F,
       max.deparse.length=Inf, keep.source=T)
source("SEM.R", echo = T, print.eval = F,
       max.deparse.length=Inf, keep.source=T)
source("LCA.R", echo = T, print.eval = F,
       max.deparse.length=Inf, keep.source=T)
source("Visualization.R", echo = T, print.eval = F,
       max.deparse.length=Inf, keep.source=T)
```