Do Teachers' Expectations Influence Students' Educational Aspirations? A study Using the Chinese Education Panel Survey (CEPS)

Yu Ao

MSc in Education (Research Design and Methodology), 2025

DECLARATION BY THE CANDIDATE AS AUTHOR OF THE DISSERTATION

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.
2. I allow the Department to deposit on my behalf a copy of this dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].
3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals (unless permission has been obtained from the individuals) and that such material should be removed before this dissertation is deposited in ORA.
4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [https://ora.ox.ac.uk/deposit_agreements] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.
5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]:	Yu Ao
Date:	27th August 2025

Abstract:

Using the 2013 to 2014 baseline of the China Education Panel Survey, this study examines whether students' perceptions of teacher expectations are associated with their educational aspirations. To address this question, the research analyses the nationally representative Grade 9 cohort (N = 9,208; analytic N = 7,971), estimating sequential linear regressions with controls for gender, ethnicity, socioeconomic background, cognitive ability, and family factors including parental expectations and parental involvement. A logistic specification is estimated as a robustness check. The results indicate a clear ordered association: students who perceive higher teacher expectations report higher aspirations, and those perceiving no clear expectation fall between the high and explicitly low groups. In models that adjust for these covariates, parental educational expectations have the strongest association with aspirations, whereas routine parental involvement shows little impact. Cognitive ability is positively related to aspirations, although the estimated effect is modest. Regarding demographic patterns, girls report higher aspirations than boys, and minority students slightly exceed Han peers after adjustment. Given the cross-sectional data structure, temporal ordering cannot be established, nor can unobserved confounding be ruled out. Therefore, the estimates should be interpreted as associations rather than causal effects. Taken together, the evidence underscores the importance of expectation signals from teachers and parents in shaping students' educational goals in China. It also points to the value of teacher-awareness initiatives, parent-focused guidance on communicating educational visions, and system-level reforms that secure equitable access to aspirational support for all students.

Acknowledgment:

I would like to express my deepest gratitude to my supervisor, Professor Steve Strand, for his unwavering support, insightful guidance, and constructive feedback throughout the course of this research. His rigorous academic standards and encouragement have been instrumental in shaping my work and pushing me toward greater intellectual growth.

I am immensely grateful to the University of Oxford and the Department of Education for providing me with the opportunity to study at one of the world's most prestigious institutions. This past year has been transformative, both intellectually and personally. I was fortunate to engage deeply with educational equity issues, and to continue developing my interest in the intersection of data, policy, and social justice.

To my family, especially my parents and grandparents, thank you for your unconditional support, patience, and belief in me throughout this journey. Your love has been the solid foundation that carried me through every moment of doubt and challenge.

A heartfelt thank-you to my 2025 RDM cohort and closest friends Usman, Xinyi, Hang, Ellen, whose encouragement, humour and companionship turned difficult days into manageable ones. I am especially thankful for the countless late-night chats, academic discussions, and emotional support.

Lastly, I would like to take a moment to acknowledge my own journey. From struggling to understand statistical models to independently analysing national survey data and constructing a complete dissertation, I have grown in ways I could never have anticipated. Whether it was writing from quiet corners of libraries or from hotel rooms across continents, I stayed committed to my goal. As I prepare to step into the next phase of my life, I hope to remain true to my educational ideals and continue working toward making learning more equitable and meaningful for all.

Table of Contents

Chapter 1: Introduction

Educational aspirations, defined as the highest level of education students aim to attain, play a pivotal role in shaping their academic trajectories and long-term life outcomes (Khattab, 2015). They influence not only students' motivation and academic effort, but also critical decisions regarding school continuation, track selection and career planning (Schoon & Parsons, 2002). In today's increasingly competitive global economy, fostering ambitious educational goals is widely regarded by educators and policymakers as a cornerstone of social mobility and human capital development (Marginson, 2016).

In China, students' academic ambitions hold significant implications for their future trajectories, as access to educational opportunities is largely determined by high-stakes examinations and constrained by rigid institutional pathways (Liu, 2016). However, these ambitions are often constrained by systemic inequalities, including unequal access to resources and regional disparities in school quality (Zhang et al., 2015; Hannum & Park, 2007). In this context, understanding what shapes students' educational aspirations a theoretical concern is not merely but also a matter of practical urgency.

In the Chinese school system, teachers operate on the front lines of students' educational experiences, directly influencing how learners perceive their academic potential (De Boer et al., 2010). At the same time, they act as gatekeepers of opportunity, with their judgments and expectations potentially fostering or constraining students' aspirations (Wang et al., 2018; Wang et al., 2025).

While prior research in Western contexts has consistently underscored the pivotal role of teacher expectations in shaping student outcomes, empirical evidence on this dynamic within China remains scarce. Furthermore, few studies have explored how teacher expectations intersect with other influential factors, including gender, ethnicity, family background and cognitive ability (Fan & Chen, 2001; Wang et al., 2018). This study, therefore, aim to address these gaps by drawing on nationally representative data from the China Education Panel Survey (CEPS) to conduct a nuanced and robust analysis of aspiration formation during a critical transitional stage in Grade 9, when students, typically aged 15 to 16, complete compulsory education and decide whether to continue to high school.

Using a sequential modelling strategy, this study constructs four linear regression models to examine the relationship between teacher expectations and student aspirations, while progressively controlling for demographic information, cognitive, and family-level variables. A final logistic regression model is employed as a robustness check, treating aspiration as a binary outcome to evaluate whether the results remain consistent across different modelling approaches. Through this design, the study provides empirical insight into how teacher expectations influence student aspirations, along with practical implications for policy and practice aimed at fostering equitable and supportive learning environments.

The remainder of this dissertation is structured as follows. Chapter 2 outlines the theoretical framework and reviews existing literature on educational aspirations, teacher expectations, and their key social determinants. Chapter 3 presents the research philosophical stances and positionality, followed by a description of the data source, variable construction and analytical strategies. Chapter 4 reports the summary statistics and the results of the sequential regression models, followed by robustness checks. Chapter 5 interprets the findings in relation to prior research, while also addressing the study's limitations and identifying directions for future research. Finally, Chapter 6 concludes the dissertation.

Chapter 2: Literature Review

2.1 Teacher Expectation

Teacher expectations have emerged as a pivotal area of educational research due to their significant impact on student learning outcomes (Danişman, 2017; Friedrich et al., 2015). Educators' perceptions of students' abilities often serve as a lens through which they shape instructional strategies, provide academic feedback and foster a conducive classroom environment (De Boer et al., 2018; Rubie-Davies, 2010). Decades earlier, Rosenthal and Jacobson (1968) demonstrated that when teachers hold high expectations, they are more likely to establish a classroom climate that promotes student engagement and academic success.

A growing body of literature further underscores that teacher expectations not only influence academic achievement but also shape students' self-concept, motivation and aspirations (Caplin, 1968; Hornstra et al., 2018; Hynds et al., 2017; Walkey et al., 2013). These expectations can either promote or inhibit students' beliefs about their own potential, leading to long-term effects on their educational trajectories (De Boer et al., 2010).

However, teacher expectations are not always objectively grounded in students' actual academic performance. Emerging research indicates that such expectations often vary systematically based on students' demographic characteristics, including socioeconomic status, ethnicity and gender (Bognar, 1983; Tenenbaum & Ruck, 2007; Wang et al., 2018). For instance, students from disadvantaged or minority backgrounds are frequently perceived less favourably, leading to lowered expectations and inequitable instructional support from teachers (Wang et al., 2021; Yiu & Adams, 2013). These biased expectations can evolve into self-fulfilling prophecies, limiting students' engagement, confidence and access to challenging learning opportunities (Brophy, 1983). Over time, such dynamics may exacerbate existing educational inequalities and further marginalise already vulnerable student populations (Cooper, 1979).

In response, recent scholarship has turned to strategies aimed at mitigating expectation-related biases through teacher professional development and reflective practice. For example, Rubie-Davies (2014) advocates for the cultivation of "high-expectation teachers" who consciously foster equitable expectations and work to promote positive academic identities among all learners, regardless of background.

2.2 Theoretical Framework

The theoretical foundation for understanding the impact of teacher expectations originates from the seminal field experiment by Rosenthal and Jacobson (1968) in a California elementary school. Teachers were told that certain randomly selected students were likely to exhibit exceptional intellectual growth, and by the end of the year these students showed significantly greater IQ gains than their peers. This phenomenon, later termed the Pygmalion Effect, demonstrates how teacher expectations shape teacher behaviour, which in turn influences students' motivation and achievement (Merton, 1948).

Building on the Pygmalion effect, Rosenthal (1973) developed the Four-Factor Model, explaining that high expectations prompt teachers to foster a warmer socio-emotional climate, deliver richer and more challenging instruction, expand opportunities for participation and provide frequent feedback. Collectively, these behaviours reshape the learning environment, enabling high-expectation students to thrive.

Brophy (1983) noted that teachers often form expectations based on prior knowledge of students, shaped by personal beliefs, assumptions and sometimes stereotypes. These expectations act as predictions of students' future behaviour or performance and once established, influence how teachers interact with them (Brophy & Good, 1970). They are conveyed through differences in teaching behaviours and the classroom climate (Brophy & Good, 1974; Turner et al., 2015), which in turn shape students' learning behaviours, academic achievement, and broader developmental outcomes (Archambault et al., 2012; Tyler & Boelter, 2008; Weinstein, 2002).

In 1985, Harris and Rosenthal's meta-analysis found that the climate and input components of the Four-Factor Model had the strongest expectancy effects, followed by output, while feedback had minimal impact. Based on these findings, Rosenthal (1989) proposed the Two-Factor Theory, in which teacher expectations are communicated through two dimensions: affect, reflecting the emotional tone and warmth of teacher–student interactions; and effort, combining input and output to capture the quantity and quality of academic attention and opportunities. This revision simplified the model while retaining its explanatory power for how expectations are transmitted and internalised in classrooms.

Three main theoretical frameworks explain how teacher expectations shape students' educational aspirations: the Self-Fulfilling Prophecy (SFP), the Theory of Planned Behaviour (TPB), and the Situated Expectancy-Value Theory (SEVT). All recognise teachers' influence on students' beliefs and choices, but each differs in conceptual focus and the mechanisms through which expectations operate.

Self-Fulfilling Prophecy (SFP)

The Self-Fulfilling Prophecy (SFP), first articulated by Merton (1948), explains how an initially unfounded belief can shape behaviour in ways that make the belief come true. In education, Rosenthal and Jacobson's (1968) experiment demonstrated that teachers' expectations, whether accurate or not, influence their behaviour toward students and shape how students see themselves. Over time, students may internalise these cues, adjusting their self-concept and performance to match the teacher's beliefs. SFP emphasises this interpersonal feedback loop, in which subtle patterns of differential treatment reinforce the original expectation.

Theory of Planned Behaviour (TPB)

While SFP focuses on the interpersonal transmission of expectations, Ajzen's Theory of Planned Behaviour (TPB) (1991) shifts the lens to the internal cognitive processes that precede action. It proposes that a student's intention to pursue an educational goal is shaped by three components: their attitude toward the behaviour, perceived social norms from significant others, perceived behavioural control, and belief in their ability to overcome obstacles. TPB thus bridges the external influence of teacher expectations with the internal decision-making process, explaining how students integrate personal evaluations, self-efficacy, and social pressures into a concrete intention to act.

Expectancy-Value Theory (EVT) and Situated Expectancy-Value Theory (SEVT)
Where TPB explains the formation of intention, Expectancy-Value Theory (EVT) (Wigfield & Eccles, 2000) and its context-sensitive revision, the Situated Expectancy-Value Theory (SEVT) (Eccles & Wigfield, 2020), step further upstream to identify the beliefs that feed those TPB components. Both frameworks posit that students' choices, effort, and persistence are driven by their expectancy for success ("Can I do this?") and the value they attach to the task ("Is it worth doing?"), with SEVT adding that these beliefs, and perceived costs such as effort, opportunity, and emotional strain, are shaped by specific situations. Together, EVT and SEVT explain why attitudes, norms, and perceived control rise or fall, thereby clarifying how intentions can strengthen or weaken depending on the learning context.

Collectively, these models highlight the interplay between external social signals and internal belief systems, showing that teacher expectations function both as behavioural cues and as psychological stimuli that shape how aspirations are formed, adjusted and pursued (Eccles & Wigfield, 2020; Merton, 1948; Rosenthal & Jacobson, 1968). This integrated theoretical foundation offers a strong basis for examining aspiration development within the specific sociocultural and institutional context of the Chinese education system.

In the Chinese research context, the terms "teacher expectation" and "teacher expectation effect" began to receive increasing scholarly attention in the early twenty-first century. Recent studies have examined how teacher expectations, together with other factors, influence students' educational aspirations (Ling et al., 2014; Pi, 2013). The following sections review both international and Chinese literature on how gender, ethnicity, socioeconomic status, cognitive ability, parental involvement, and parental expectations shape students' educational aspirations.

2.3 Gender

Gender plays a consistent role in shaping educational aspirations. Internationally, girls often report higher academic ambitions than boys. Schoon and Polek (2011), analysing two nationally representative UK birth cohorts, found that adolescent girls were more likely than boys to aspire to professional or managerial careers and to remain longer in full-time education, even after controlling for socioeconomic background and cognitive ability. Similarly, Berrington et al. (2016) showed that white working-class boys in the UK expressed the lowest aspirations to enter higher education, reinforcing the view that gender intersects with social class and ethnicity in shaping young people's future-oriented attitudes.

These differences in aspirations are mirrored in broader patterns of school achievement and motivation. A meta-analysis by Voyer and Voyer (2014), drawing on over 500 effect sizes, found that girls consistently outperform boys in overall school grades, with the largest advantages observed in language-based subjects. This academic edge may contribute to girls' stronger self-concept and greater confidence in their educational abilities. Complementing this, cross-national evidence from the OECD's PISA analyses (Del Pero & Bytchkova, 2013) shows that girls report more positive learning attitudes, higher levels of school engagement, and fewer behavioural problems than boys. Yet, persistent gender stereotypes continue to shape subject choices, leaving girls underrepresented in STEM fields despite their strong overall academic profiles.

Beyond academic performance and engagement, more recent research highlights the role of social and cultural factors in shaping aspirations. Lundberg (2020) argues that gender gaps in aspiration are not solely a function of academic preparedness or socioeconomic status but are also shaped by gender identity norms and motivational beliefs. She finds that boys, particularly those from disadvantaged backgrounds, are less likely to value education or perceive it as compatible with masculine identity. Even after controlling for achievement and non-cognitive skills, boys exhibited significantly lower educational expectations than girls.

This pattern reflects not only individual attitudes but also the influence of socially embedded norms that discourage educational ambition among certain groups of boys.

In China, similar patterns emerge. Drawing on CEPS data, Wei and Ma (2017) found that female students not only exhibited greater educational ambition but were also more likely to be influenced by maternal education and parental expectations. Liu (2020) further confirmed this trend, noting that girls' higher aspirations were partly shaped by stronger parental and teacher encouragement. These findings resonate with international patterns but also reflect specific sociocultural dynamics within Chinese families and schools.

In addition to individual and family influences, gendered aspirations are shaped by both school-level and broader societal contexts. Using China Education Panel Survey data, Huang and Wu (2016) showed that school-based factors, including teacher expectations and peer interactions, were significant in shaping students' aspirations, with these influences differing by gender in ways that reinforced girls' academic engagement and future-oriented planning. At the macro level, Chen et al. (2022) demonstrated that regional gender inequality suppresses educational aspirations for both girls and boys. Their longitudinal analysis revealed that in provinces with higher levels of gender discrimination, adolescents of both sexes were less likely to maintain high educational goals, while household gender attitudes similarly shaped aspirations at the micro level.

In sum, gender differences in educational aspirations are shaped by intertwined academic, social and cultural factors. Evidence from both international and Chinese contexts shows that girls generally display higher ambitions than boys.

2.4 Ethnicity

International research presents mixed findings on ethnicity and aspirations. Some studies suggest that many ethnic minority adolescents set goals that match or even exceed those of their White peers, whereas others report persistent disparities that disadvantage minority youth. For instance, longitudinal evidence from Kao and Tienda (1998) shows that Black and Hispanic students often express high aspirations in early adolescence but are less likely to maintain them through high school, likely due to structural barriers and a lack of sustained institutional support. By contrast, Asian American students were more likely to sustain high expectations over time, a pattern attributed to strong family investment and culturally reinforced academic norms. However, other research points to enduring gaps. Nitardy et al. (2015), using over a decade of Minnesota Student Survey data, documented consistent differences in both academic performance and

postsecondary plans. Native American, Black, and Hispanic students reported lower GPAs and weaker aspirations than their White and Asian peers, with these differences partly explained by poverty and family structure, both of which disproportionately affect minority youth.

Beyond these trends, Qian and Blair (1999) examined the mechanisms underlying ethnic differences in aspirations. They found that although academic performance strongly predicts aspirations across all racial and ethnic groups, human and financial capital exert a stronger influence on White students, whereas social capital, particularly parental involvement in school activities, plays a greater role for Black and Hispanic youth. This suggests that different groups draw on distinct resources to support their educational goals.

Cultural meaning also shapes the pursuit of higher education. Turcios-Cotto and Milan (2013) found that Latino students were less likely than their White or Black counterparts to include college in their future goals, instead prioritising family-related aspirations such as forming a household. Notably, Latino youth who did aspire to college reported higher levels of psychological distress, indicating that, for some, higher education may be experienced as a source of cultural tension rather than empowerment.

While much of the existing evidence comes from Western contexts, emerging research in China demonstrates both similarities and context-specific differences in ethnic patterns of educational aspiration. In the Chinese context, research has increasingly examined how ethnic background shapes students' educational aspirations, particularly in rural and minority-dominated regions. Pan (2012), studying Bai families in rural Yunnan, found that while many parents expressed a willingness to support both sons and daughters in their schooling, their actual aspirations remained modest. Most families aimed for high school completion, and only a minority hoped for university enrolment. Decisions about continuing education were strongly shaped by students' academic performance and perceived career prospects, rather than long-term intellectual goals.

While Pan's findings emphasise the role of academic performance and labour market considerations, other research identifies a broader pattern in which minority students often report lower aspirations than their Han counterparts, not necessarily because of cognitive ability or access to schooling, but due to entrenched cultural norms and differing perceptions of the value of education. Hu (2004), in her study of Hui Muslim families in rural Xinjiang, found that both students and parents held significantly lower educational expectations than nearby Han families. Only slightly more than half of Hui parents hoped their children would attend university, and many expressed stronger supports for religious education over secular schooling. Similarly, Liu (2012) documented that in the poverty-stricken and multi-ethnic Zhemi Township

of Yunnan Province, families showed limited belief in the transformative power of education. Despite government investments and tuition subsidies, both parents and students often regarded schooling as a low-return, high-risk endeavour, especially when immediate income from labour was readily accessible.

Other studies highlight the role of household circumstances and local opportunity structures in mediating educational aspirations. Tian (2022), focusing on left-behind children in Enshi Prefecture, Hubei Province, found that students in ethnic regions were more likely to internalise cultural narratives framing education as unnecessary or unattainable, particularly when older generations had little formal schooling themselves. In the case of Tibetan communities, Meng (1998) found that families in agricultural areas expressed stronger educational demand, underpinned by lower costs and higher perceived returns to schooling. By contrast, pastoral areas showed weaker demand, constrained by high opportunity costs and limited local returns to education.

Together, these findings underscore that in China, as elsewhere, ethnicity shapes educational aspirations through the interplay of cultural norms, economic realities, and the perceived availability of mobility pathways.

2.5 Social Economic Status (SES)

A large international literature shows that socioeconomic status shapes how high young people aim educationally. Analyses of PISA 2003 across 28 OECD countries indicate that students from more advantaged families are more likely to aspire to university, and this association persists after accounting for achievement (Dupriez et al., 2012). Consistent with this pattern, evidence from Luxembourg and the Swiss Canton of Bern shows that students from lower SES backgrounds are less likely to aspire to higher education (Hadjar et al., 2021). In the United Kingdom, family background remains a strong predictor of teenagers' aspirations even once prior attainment is considered (Berrington et al., 2016). Hanson (1994) further shows that among high ability United States youths, socioeconomic status is the most consistent predictor of lost talent, with lower SES students more likely both to hold expectations that fall short of aspirations and to reduce or fail to realise expectations over time.

Building on broader understandings of socioeconomic status, recent research has examined how its components influence aspirations in distinct ways. Drawing on Slovak data, Madarasova Geckova et al. (2010) report patterns that vary by track: father's education matters in grammar schools, whereas mother's education is more influential in specialised and vocational tracks, and perceived support from fathers is

positively associated with intending to continue study. Classic status attainment research likewise shows that parental education, income, and occupational prestige influence adolescents' aspirations primarily indirectly through encouragement from significant others, school achievement, and prior aspirations (Sewell et al., 1969).

In the Chinese context, family socioeconomic status likewise emerges as a central determinant of students' educational aspirations and attainment. Using a large PISA-based sample, Pang et al. (2013) show that mothers' education and learning-related home resources such as books, a computer and broadband are the most salient SES components for academic achievement, whereas general wealth items show weaker or inconsistent links. They also report that Chinese parents' educational expectations are high overall and rise with SES. Complementing this evidence, An (2005) shows that in poor rural contexts, students' educational engagement and aspirations are limited less by a lack of motivation than by deficits in social and cultural capital within the home and the school. Strengthening these forms of capital through resources, networks and culturally valued practices is associated with higher levels of engagement.

In this study, I use parental occupation as the sole proxy for socioeconomic status (SES), a choice that is both methodological and substantive grounded in prior research (Hauser & Warren, 1997). First, occupation reflects not only a household's economic resources but also the social networks and forms of cultural capital (Hopper et al., 1968). Second, compared to income or wealth, occupational information is reported more reliably in large-scale Chinese surveys. Income data often suffer from high nonresponse rates and year-to-year volatility, whereas parental occupation tends to be more stable and offers a consistent reference point throughout an adolescent's schooling years (Sirin, 2005). Third, occupation-based classification systems such as ISCO codes are internationally standardised, enabling cross-national comparisons. Finally, using a single, well-measured SES indicator helps avoid multicollinearity and model instability, which can occur when income, education, and occupation are included simultaneously but differ in measurement precision (Campbell & Parker, 1983). Considering its conceptual coverage, data reliability and cross-study comparability, this study adopts parental occupation as the preferred measure of SES in analysing how family background shapes educational aspirations.

To sum up, students from higher socioeconomic backgrounds consistently report higher educational aspirations in both international and Chinese contexts. Parental occupation, reflecting economic resources and social capital, serves as a reliable and comparable proxy for SES in examining how family background shapes aspirations.

2.6 Cognitive Ability

International research consistently finds that cognitive ability is an important determinant of educational aspirations. Defined as an individual's capacity for reasoning, problem solving, and information processing, it is typically measured using standardised verbal, quantitative, and spatial assessments (Ones et al., 2012; Ree et al., 2001).

In Western contexts, empirical studies show that cognitive ability is associated with more ambitious adolescent aspirations and predicts aspirations indirectly via academic achievement. For example, using UK longitudinal data, Schoon and Polek (2011) report that higher adolescent cognitive test scores are linked to more ambitious teenage job aspirations and to higher adult social status, net of social class and gender. Beyond these associations, ability can also function as a mediator of family background. Ganzach (2000) shows that cognitive ability partly mediates the association between parental education and youths' educational expectations, consistent with a pathway from SES through ability to expectation.

Taken together, these findings accord with Marks' (2013) comparative evidence that the influence of social background on attainment is moderate, whereas the impact of cognitive ability remains strong and distinct from socioeconomic status. This challenges views that aspirations are shaped primarily by class-based constraints and supports a model in which ability plays a leading role.

Recent psychological frameworks extend this understanding by examining how ability interacts with self-perception and planning. Sternberg (2011) highlights that aspirations are not determined by measured intelligence alone, but also by students' metacognitive skills, goal setting and strategic thinking, which help explain why some high ability students develop strong aspirations while others do not.

In the Chinese context, evidence indicates that parental educational expectations, alongside concrete investments and learning time, are proximate mechanisms shaping children's cognitive development. Zhou et al. (2016), using panel data from the China Family Panel Studies, show that rural children whose parents hold higher educational expectations score significantly higher on cognitive tests; this association operates largely through greater parental educational investment, with effects often stronger for girls. Similarly, Liang et al. (2018) analyse migrant children in CEPS, disaggregating parental involvement into eight behavioural

dimensions, and find that study-related practices, such as homework supervision, discussing school matters, and home–school communication, are positively associated with children's cognitive competence.

Building on this evidence, Liu et al. (2024) use CEPS to show that parental schooling is positively related to children's cognitive outcomes, particularly for girls, and they identify students' educational expectations, tutoring expenditures, and parent–child academic interactions as key mediators. Finally, Fang and Hou (2019), using structural equation modelling, show that family socioeconomic status shapes students' cognitive competence via parental daily caring and students' self-educational expectations, with a chain pathway from SES to caring to expectations to cognition.

Cognitive ability, therefore, affects aspirations both directly, by shaping perceived capability, and indirectly, by mediating family background effects.

2.7 Parental Involvement

Evidence from multiple countries indicates that how parents engage with their children's schooling can significantly shape both academic performance and future aspirations (Castro et al., 2015). Using data on Japanese adolescents, Otani (2019) shows that both maternal and paternal involvement predict students' educational goals when measured through discussions between parents and children and through parents' educational expectations. The study also reveals nuanced gender dynamics: discussions with fathers are more predictive of boys' aspirations and grades, whereas discussions with mothers relate more clearly to girls' academic performance. Students' own aspirations mediate part of the association from parental involvement to achievement.

Turning to specific forms of involvement, Jung and Zhang (2016) analyse new immigrant families in the United States and show that school based parental involvement is linked to higher academic achievement, in part because it fosters children's educational aspirations. By contrast, controlling involvement such as strict monitoring is negatively associated with cognitive development and shows little connection to aspiration. Consistent evidence emerges in Germany: Karbach et al. (2013) find that, once general cognitive ability is controlled, autonomy supportive practices are significantly associated with grades, whereas result focused parental pressure and rigid, tightly regulated home study routines are negatively associated with performance in both mathematics and language.

In line with international evidence, Chinese studies converge on two points: not all forms of parental involvement are equally effective, and the most consequential forms are aspirational and academically communicative rather than directive or controlling. Using CEPS baseline data, Huang (2016) shows that parents' educational participation and sense of educational responsibility, together with school facing actions such as attending parent teacher meetings, are positively associated with junior secondary students' cognitive test scores. Extending the focus from what parents do to how such involvement works, Duan et al. (2018) find that academic socialisation, defined as parents' achievement-oriented engagement and confidence in children's learning, predicts both achievement and positive school conduct, whereas general home-based involvement is linked primarily to behavioural outcomes and shows little association with achievement.

Classifying parental engagement into profiles, Gan and Bilige (2019) identify four home-based patterns and show that the supportive profile, which combines frequent parent child communication, encouragement, homework support and high expectations, is most strongly associated with achievement even after adjusting for SES. Finally, using CEPS data, Xie and Zhao (2022) distinguish three dimensions of parental engagement and report that parental expectations are the strongest positive predictor, followed by school related communication. Routine homework checking yields little benefit, and intrusive homework help is negatively associated with performance, with girls gaining more from expectations and communication than boys.

Across contexts, involvement that communicates aspirations and supports learning positively influences achievement and aspirations, whereas controlling or intrusive practices tend to undermine them.

2.8 Parental Expectation

A consistent body of research demonstrates that parental expectations are a powerful predictor of adolescents' educational aspirations, with particularly strong effects among socioeconomically disadvantaged groups. Longitudinal evidence from Englund et al. (2004) found that expectations set in early childhood predict subsequent parental involvement and children's academic achievement, while children's academic performance also feeds back into parental expectations. Among low-income families in their sample, higher expectations led to more active academic engagement from parents, which then improved achievement, suggesting a reciprocal developmental process. In a survey of low-income families in the U.S. Midwest, Kirk et al. (2011) further demonstrate that parental expectations explained 16% of the variance in students' own aspirations, and that adolescents whose parents anticipated higher educational attainment tended to hold more positive academic self-perceptions and achieve higher GPAs.

Similar results are reported in Finland, where Lazarides et al. (2016) found that both parental expectations and students' motivational profiles significantly contributed to adolescents' aspirations; notably, expectations retained a positive effect even after controlling for intrinsic motivation and prior performance, indicating an additive influence. In Australia, Koshy et al. (2019) examined expectations for higher education participation and found that parents' perceptions of their child's academic ability were the most influential determinant of their expectations, which in turn closely matched the student's aspirations.

In the Chinese context, research identifies parental expectations as a central influence on students' educational aspirations, while also highlighting how such expectations are shaped and transmitted. Using CEPS data, Wei and Ma (2017) found that parental expectations were among the strongest predictors of students' self-reported aspirations, with stronger effects for girls than boys. Expectations were in turn influenced by parents' own education and by frequent parent–child communication, especially with mothers. Building on these findings, Ma and Wei (2017) developed a cyclical model of the Rosenthal effect in the family context. Drawing on expectancy theory, they propose that expectations originate from family structure and prevailing social norms, and are conveyed through cognitive, emotional and material investment. Students internalise these expectations, which shape their motivation, self-perceptions and behaviours; as academic outcomes align with expectations, parental beliefs are reinforced, and the cycle continues.

The social embeddedness of parental expectations is further illustrated in Sheng's (2014) study of urban Beijing, which, using both survey and interview data, showed that middle-class parents expressed higher and more specific expectations than working-class parents. These class differences were closely tied to cultural capital indicators such as parental education, home book collections, and participation in cultural activities. While working-class parents often voiced high aspirations, they lacked the informational and cultural resources to translate them into effective guidance.

Collectively, existing evidence shows that students' educational aspirations are shaped by a combination of individual characteristics, family resources, and school influences. However, key gaps remain. Few domestic studies investigate how teacher expectations interact with other influential factors such as gender (Richardson et al., 2020; Gil-Flores et al., 2011; Ojeda & Flores, 2008), ethnicity (Gil-Hernández & Gracia, 2018; Khattab, 2018), socioeconomic status (Berrington et al., 2016; Madarasova Geckova et al., 2010), cognitive ability (Ones et al., 2012; Ree et al., 2001), parental involvement (Castro et al., 2015; Karbach et al., 2013; Otani, 2019), and parental expectations (Kirk et al., 2011; Koshy et al., 2019; Lazarides et al.,

2016). Even when such variables are considered, they are often analysed in isolation. Consequently, the field lacks an integrated analytical framework capable of assessing how these factors combine and interact to shape students' educational aspirations.

Moreover, most Chinese studies rely on regional or school-based samples, limiting the generalisability of findings and raising concerns about whether observed patterns reflect national trends or localised conditions (Liu, 2012; Meng et al., 1998; Pan, 2012; Tian, 2022). Addressing these gaps calls for a nationally representative study that situates teacher expectations within the broader sociocultural ecosystem influencing student aspirations (Liang et al., 2018; Liu et al., 2024; Wei & Ma, 2017; Xie & Zhao, 2022).

2.9 Summary and Research Questions

Building on the insights reviewed above, this study is guided by the following research questions:

RQ1. To what extent are students' educational aspirations associated with their perceived teacher expectations?

RQ2. Do perceived teacher expectations remain a significant association of aspirations after adjusting for gender, ethnicity, socioeconomic status, cognitive ability, parental involvement, and parental educational expectations?

RQ3. How does the relative influence of perceived teacher expectations compare with that of parental educational expectations in predicting students' educational aspirations?

Chapter 3: Methodology

3.1 Philosophical Stance and Positionality

3.1.1 Ontology, Epistemology and Axiology

This research is grounded in a realist ontological and positivist epistemological stance (Bryman, 2016; Crotty, 1998). It assumes that educational aspirations and teacher expectations are social phenomena that exist independently of the observer and can be objectively measured through questionnaire instruments (Krosnick & Presser, 2010). A positivist epistemology supports the use of quantitative data and statistical modelling to test hypotheses and infer generalisable patterns across large populations (Cohen, 2002). This realist-positivist stance also reflects an axiological commitment to generating evidence that can inform data-driven educational policymaking (Pawson, 2006; Slavin, 2002). In doing so, the research aims to contribute to broader discussions on how teacher expectations may shape students' educational aspirations, with the goal of informing more inclusive and equity-oriented educational practices (Rosenthal & Jacobson, 1968; Rubie-Davies, 2010).

3.1.2 Positionality

As a Chinese student currently undertaking postgraduate study in education at the University of Oxford, I occupy a unique position in relation to this research. My interest in the relationship between teacher expectations and student aspirations is shaped by my academic training and by first-hand experiences within the Chinese education system. My teaching experience in rural China exposed me to marked variation in how teacher expectations shaped students' confidence and aspirations. This experience deepened both my scholarly interest and my commitment to understanding how educational inequality may be perpetuated or challenged through expectation dynamics in schools.

These experiences inform my interpretive lens (Berger, 2015). While I aim to analyse the CEPS data with methodological rigour, I acknowledge that my interpretations may be shaped by background knowledge of Chinese schooling. By making my positionality explicit, I maintain a critical awareness of how my background may influence the research process, including the framing of questions, the selection of variables, and the interpretation of results (England, 1994). To mitigate bias, I document analytic decisions, report robustness checks and prioritise transparency in data processing and modelling.

3.2 Data Source

The data used in this study are drawn from the China Education Panel Survey (CEPS), conducted by the National Survey Research Centre (NSRC) at Renmin University of China. Although CEPS is structured as a longitudinal panel survey, the current analysis is based exclusively on data from the 2013–2014 baseline wave. This study focuses specifically on the Grade 9 cohort, as the Grade 7 sample does not include measures of students' perceptions of teacher expectations, which are central to the research questions. Furthermore, the publicly available longitudinal data only follow the Grade 7 cohort for one additional year, and do not contain several key variables required for this analysis. For these reasons, the study adopts a cross-sectional design based on a single wave of data.

The baseline sample includes approximately 20,000 G7 & G9 students across 438 classrooms in 112 schools, covering 28 county-level units across mainland China. This sampling strategy ensures a high level of representativeness across geographic regions, school types and socioeconomic groups, making CEPS particularly suitable for examining issues of educational inequality and development. Five types of questionnaires were administered in the baseline wave: to students, parents, homeroom teachers, subject teachers, and school administrators. My research draws primarily on the student questionnaire, which provides rich information on students' background characteristics, academic attitudes, school experiences, and family environment.

The CEPS dataset is publicly available and widely used in academic research related to education and social stratification. All data used in this study were fully anonymised by the original data providers. Since the analysis is based exclusively on secondary data and does not involve direct contact with human participants, no formal ethical approval was required. Data were stored securely in encrypted drives and used solely for academic purposes. This study complies with the ethical guidelines of the University of Oxford's CUREC framework for the use of secondary data.

3.3 Variable Construction

This section presents the key variables used in the analysis, including the dependent variable, the focal independent variable, and the control variables employed in both sequential linear regression and logistic regression models. All variables are derived from the 2013–2014 CEPS Grade 9 student questionnaire. Although the baseline wave of CEPS includes two cohorts, Grade 7 and Grade 9 students, this study focuses exclusively on the Grade 9 cohort with an analytical sample size of N = 9,208. This decision is

driven by data availability: the key independent variable, students' perceptions of their teachers' expectations, is only available for Grade 9 students in the baseline dataset. Given the centrality of this variable to the research question, the analytical sample is accordingly restricted to this group.

3.3.1 Missing Data

Although no formal statistical test was conducted, the observed patterns of missingness do not suggest Missing Completely at Random (MCAR), which would imply an implausibly uniform distribution of missing values across unrelated variables. Nor do we find direct evidence for Missing Not at Random. Instead, missingness appears to correlate with observed characteristics such as students' demographic traits and family background. This pattern is consistent with the Missing at Random assumption, meaning that, conditional on the observed data, the probability of missingness does not depend on the unobserved values themselves (Baraldi & Enders, 2010; Little, 1988)

To assess the representativeness of the analytic sample (N = 7,971), I compared included and excluded cases (N = 1,237) on key variables, applying chi square tests to categorical indicators and t tests to continuous measures. The comparisons covered the share expecting a bachelor's degree or above, the share of girls, the share of ethnic minority students, mean socioeconomic status, mean cognitive ability, and mean parental expectations and involvement. Subtle differences were observed across most indicators. Relative to excluded cases, the analytic sample showed a higher share of students aspiring to a bachelor's degree or above, a larger share of girls, higher cognitive ability, and higher parental educational expectations, alongside lower socioeconomic status, lower parental involvement, and lower perceived teacher expectations, including lower proportions reporting expectations for both key senior high school and ordinary senior high school tracks (see Appendix A.1).

Comparisons between included and excluded cases indicate that missingness is not completely at random and that some selection bias is possible. To preserve a constant sample across the sequential regression models, I use listwise deletion so that coefficients are not confounded by changes in sample composition across specifications (Millsap & Maydeu-Olivares, 2009). This choice prioritises internal comparability over external representativeness. Under a Missing at Random assumption, where missingness depends only on observed covariates included in the models, complete case estimates are consistent for the regression parameters, although generalisability may be reduced (Carpenter & Smuk, 2021; Little & Rubin, 2019). Therefore, results should be interpreted with caution regarding extrapolation beyond the analytic sample.

3.3.2 Measurement

Dependent Variable: Educational Aspirations (c22)

Students were asked the following question:

C22. What is the highest level of education you expect yourself to receive?

1. Drop out now
2. Graduate from junior high school
3. Go to technical secondary school or technical school
4. Go to vocational high school
5. Go to senior high school
6. Graduate from junior college
7. Get a bachelor degree
8. Get a Master degree
9. Get a Doctor degree
10. I don't care

To enhance interpretability, students who selected either "drop out now" (code 1) or "I don't care" (code 10) were combined into a single category labelled *low aspiration*. This decision was based on the conceptual similarity between the two responses, as both indicate a lack of concrete educational goals or disengagement from formal schooling. To empirically support this classification, students' mean cognitive ability scores were compared across the two categories. While those selecting "I don't care" (code 10) demonstrated slightly higher mean scores than those selecting "drop out now" (code 1), both groups exhibited relatively lower cognitive ability relative to students expressing higher educational aspirations (e.g., M = 7.59 and 8.95 for codes 1 and 10, respectively, versus M = 9.45 - 10.80 for codes 7–9, representing aspirations for a bachelor's, master's, or doctoral degree). This pattern reinforces the rationale that both responses reflect academic disengagement, thereby justifying their combination into a unified low-aspiration group in subsequent analyses.

The dependent variable, educational aspirations, captures students' self-reported expectations regarding the highest level of education they hope to attain. In the linear regression models, this ordinal variable is treated as continuous and coded from 1 (lowest aspiration: "drop out / don't care") to 9 (highest aspiration: "doctoral degree"). This treatment follows established practice in education research, where ordinal variables with sufficiently granular categories (normally greater than 5 levels) are considered suitable for continuous modelling (Field, 2024; Zumbo & Zimmerman, 1993)

In addition to the linear specification, the variable is also used in a binary form for logistic regression as a robustness check. In this model, students aspiring to a bachelor's degree or higher are coded as 1, while those with lower aspirations are coded as 0. This threshold reflects a substantively meaningful divide in educational ambition, as aiming for tertiary academic education typically signals greater access to high-status

occupational trajectories in China (Wen, 2022). The bachelor-level cutoff is commonly adopted in education research, particularly in the context of China's competitive and stratified higher education system, where university admission is socially and economically significant (Sirin, 2005).

Independent Variable: Perceived Teacher Expectations (c29)

Students were asked the following question:

C29. What kind of advice do you think your homeroom teacher will give to you about your education after junior high school?

1. Going to a key senior high school
2. Going to an ordinary senior high school
3. Going to a vocational high school
4. Going to a technical secondary school/technical school
5. Not going any further after graduating from junior high school
6. None

The key independent variable in this study is students' perceptions of their teachers' expectations. As the CEPS dataset does not include a direct measure of teacher expectations, this variable provides the closest available proxy. It reflects students' beliefs about the highest level of education their homeroom teachers expect them to achieve and is used in this study as an indirect but meaningful indicator of teacher expectations.

Although the variable has more than five levels and could theoretically be treated as continuous (Zumbo & Zimmerman, 1993), it was dummy coded to enhance the interpretability of group-wise comparisons. This modelling decision is theoretically and empirically motivated (Long, 2006). First, the levels represent qualitatively distinct educational tracks rather than equally spaced intervals. Second, the distribution of students across these tracks is highly uneven, reflecting systemic stratification within the Chinese education system (Zhao & Glewwe, 2010). Dummy coding therefore facilitates the interpretation of specific contrasts between each category and a meaningful reference group, without assuming a linear relationship across levels (Field, 2024).

In the Chinese education system, "key senior high schools" are elite academic institutions that receive preferential resources and admit students with the highest entrance examination scores, serving as the main pathway to top-tier universities (Hannum & Park, 2007). "Ordinary senior high schools" admit students with moderate exam performance and offer fewer academic resources, though they still prepare students for university entrance (Li et al., 2015). "Vocational high schools" provide broad skill-oriented training in fields such as mechanics, hospitality, or accounting, typically leading to employment or enrolment in tertiary vocational colleges rather than academic universities (Shi, 2013). "Technical secondary schools" also deliver

occupational training but focus on more specialised and industry-specific programmes (e.g., nursing, electronics, applied arts) (Wu & Ye, 2018). These distinct tracks exemplify the systemic stratification that characterises upper secondary education in China.

Accordingly, five dummy variables were created to represent each category, with "going to a key senior high school" designated as the reference group. This track reflects the most academically prestigious form of perceived teacher expectation, making it a theoretically benchmark (Hannum & Park, 2007). Each dummy variable captures the perceived divergence from this highest expectation. For example, the dummy for "vocational high school" indicates students who believe their teacher expects them to follow a vocational rather. Apart from that, the inclusion of a "none" category, comprising students who reported perceiving no expectations from their teachers, enables the analysis to explore how disengagement or perceived neglect may be associated with lower educational aspirations (Wang & Eccles, 2013).

Control Variables

Demographic Controls

Gender

The gender variable was recoded as a binary indicator, with male coded as 0 (reference group) and female coded as 1. This coding scheme allows the regression coefficients to be interpreted as the difference in educational aspirations between male and female students.

Ethnicity

Ethnicity is recorded across multiple categories, including Han, Mongol, Manchu, Hui, Tibetan, Zhuang, and other minority groups. For analytical purposes, the ethnicity variable was recoded as a binary indicator, with Han coded as 0 (reference group) and all other ethnicities coded as 1. This dichotomous classification is both theoretically and empirically motivated. First, over 90% of the sample identifies as Han (please refer to Appendix A.2), meaning other ethnic categories are too small for separate analysis. Second, the binary coding aligns with common practice in education research using Chinese datasets, where the Han/non-Han distinction is analytically meaningful for exploring disparities in educational outcomes.

Socioeconomic Status (SES)

Socioeconomic status (SES) is a multidimensional construct typically encompassing parental education, income and occupation (Sirin, 2005). In CEPS, measures of household income or "perceived" financial conditions (e.g., SB9) rely on students' subjective assessments and may not accurately reflect actual

economic resources. Similarly, while parental education is frequently used in SES measurement, it may not fully capture the family's current socioeconomic positioning, especially in contexts like China, where educational attainment does not always correspond with occupational prestige or income (Li et al., 2017).

In contrast, occupational status has been widely accepted as a robust and pragmatic proxy for SES in stratification and educational research. It tends to be more directly linked to access to economic and social capital (Ganzeboom & Treiman, 1996; Hauser & Warren, 1997). Accordingly, this study adopts parental occupation as the operational indicator of SES.

Therefore, the SES variable is constructed based on students' reports of both parents' current and most recent job (B8). Following the dominance approach, the more prestigious of the two occupations, indicated by the lower numerical code in CEPS (see Appendix B), is used to represent household SES. This approach assumes that the highest occupational status within a household serves as the most accurate reflection of the family's overall access to resources and social positioning (Erikson, 1984).

In the regression models, the SES variable is entered using dummy coding. The reference group is "peasant", which represents the largest occupational category in the sample (23.90%) (see Appendix A.3). Selecting the most frequent category as the reference reduces the risk of inflated standard errors for other coefficients and facilitates interpretation in categorical comparisons (Allison, 1999). Substantively, agricultural households typically represent lower SES in the Chinese context and face more limited access to educational opportunities, making them a meaningful benchmark for assessing the effects of higher-status occupations (Sirin, 2005; Hannum & Park, 2007).

Cognitive Ability

Cognitive ability is measured using students' scores from CEPS standardised cognitive ability test. This assessment, designed to evaluate logical reasoning and problem-solving skills rather than curriculum-specific knowledge, yields result that are both internationally comparable and consistent across regions (Kuncel, 2010). The test consists of 22 questions covering verbal, figural and numerical/logical reasoning domains, and is administered with a 15-minute time limit.

The raw score, represented by the variable stcog, is the total number of correct answers, ranging from 0 to 22. Higher scores reflect a stronger cognitive ability. Although the score is technically discrete, its wide

range allows it to be treated as an approximately continuous variable in regression analysis (Zumbo & Zimmerman, 1993).

Family-Level Variables: Parental Expectations and Parental Involvement

Parental Involvement

Parental involvement was measured using seven items selected from the CEPS Grade 9 student questionnaire. These items aimed to capture two key domains: (1) parental academic supervision and behavioural monitoring, and (2) school-related communication with parents. Specifically, the items included (see Appendix B):

B22 (2 items): how frequently in the past week parents (a) checked the student's homework (b2201) and (b) provided instructions on homework (b2202)

B23 (3 items): perceived parental care and strictness concerning on (a) academic performance (b2301), (b) behaviour at school (b2302) and (c) school attendance (b2303)

B24 (2 items): frequency of school-related discussion with (a) the student's mother (b24a1) and (b) father (b24b1)

All items were ordinal in nature but varied in scale. For example, B22 used a 4-point frequency scale ranging from "Never" to "Almost every day", B23 used a 3-point scale ranging from "They don't care" to "Very strict", while B24 used a 3-point scale from "Never" to "Often". To ensure comparability across items with different response formats, the scales were harmonised by treating all variables as ordered categorical indicators, which were then used in the factor analysis.

To uncover the latent structure underlying these items, exploratory factor analysis (EFA) was employed using principal axis factoring (PAF) as the extraction method and oblimin as the rotation method. The suitability of the data for factor analysis was assessed using the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity. The overall KMO value was 0.71, indicating a middling level of sampling adequacy according to Kaiser's (1974) classification, and all individual items exceeded the recommended threshold of 0.50. Bartlett's test of sphericity was statistically significant, χ^2 (21) = 11064.1, $p < .001$, rejecting the null hypothesis that the correlation matrix is an identity matrix (Bartlett, 1954). These results suggest that the items share sufficient common variance and that the dataset is appropriate for exploratory factor analysis (Tabachnick & Fidell, 2013).

Exploratory factor analysis (EFA) was used to identify the latent construct of parental involvement. EFA is appropriate when the factor structure is not pre-specified, as it allows factor loadings and cross-loadings to emerge freely from the data, making it preferable to confirmatory factor analysis (CFA) in this context where no prior measurement model has been established for these items (Floyd & Widaman, 1995).

Within EFA, principal axis factoring (PAF) was selected as the extraction method because it isolates shared (common) variance, whereas principal component analysis (PCA) focuses on total variance and is primarily a data reduction technique. Since the aim of this study was to model the conceptual dimensions of parental involvement rather than simply reduce the number of observed variables, PAF was deemed the more appropriate choice (Fabrigar et al., 1999).

An oblique rotation (direct oblimin, $\gamma = 0$) was applied to the factor solution. Oblique rotation was selected over orthogonal alternatives because the underlying dimensions of parental involvement, such as supervision, discipline and communication are theoretically and empirically expected to be correlated rather than independent, with observed inter-factor correlations ranging from $r = .36$ to .45 (see Appendix A.4, Table A.4.3). Among oblique rotation methods, direct oblimin was chosen instead of promax because it provides a balanced and interpretable factor structure without imposing the more extreme simple-structure constraints characteristic of promax. Given the number of variables (seven items) and the large sample size ($N = 7{,}971$), computational efficiency was not a concern, making direct oblimin preferable for this exploratory context (Costello & Osborne, 2005).

Therefore, a parallel analysis was conducted to determine the optimal number of factors to retain, and the results supported a three-factor solution (see Appendix A.4, Figure A.4.1). All factor loadings exceeded the commonly recommended threshold of 0.50, ranging from 0.50 to 0.80, indicating strong associations between items and their respective factors (Hair et al., 2019). Furthermore, the three factors were moderately correlated ($r = 0.40$) (see Appendix A.4, Figure A.4.2), supporting the use of oblique rotation in the principal axis factoring analysis (Tabachnick & Fidell, 2013).

The rotated factor loadings revealed three interpretable dimensions: Factor 1 comprised b2201 ($\lambda = .8$) and b2202 ($\lambda = .7$), reflecting direct parental oversight of homework; Factor 2 included b2301 ($\lambda = .5$), b2302 ($\lambda = .8$), and b2303 ($\lambda = .5$), capturing perceived parental care regarding academic and behavioural issues; Factor 3 was represented by b24a1 ($\lambda = .6$) and b24b1 ($\lambda = .7$), indicating the frequency of parent–child discussions about school.

These three factors explained 47% of the total variance (PA1 = 18%, PA2 = 17%, PA3 = 12%), which falls within the acceptable range for social science research (Henson & Roberts, 2006). Regression-based factor score correlations with their respective factors were strong (ϱ = .79–.88), suggesting good measurement precision (DiStefano et al., 2009). These results support the construct validity of a multidimensional model of parental involvement.

To provide construct-level evidence of relevance (see Appendix A.4, Table A.4.4), Spearman's rank-order correlations were computed between each parental involvement item and students' educational aspirations. The correlations ranged from ϱ = 0.054 to 0.215, indicating very weak but consistently positive associations. The highest correlation was observed for b24a1 ("Mother: Discussed school events with you"; ϱ = 0.215), followed by b2301 ("Parents care about your homework/exam"; ϱ = 0.156) and b24b1 ("Father: Discussed school events with you"; ϱ = 0.130). Mid-range associations were found for b2302 ("Parents care about your behaviour at school"; ϱ = 0.090) and b2202 ("Parent gave instruction on your homework"; ϱ = 0.074). In contrast, the weakest associations were observed for b2303 ("Parents care about your school attendance"; ϱ = 0.062) and b2201 ("Parent checked your homework"; ϱ = 0.054). Overall, the pattern suggests that communicative forms of parental involvement, particularly discussions about school, may be more positively related to students' aspirations than supervisory behaviours (Erdem & Kaya, 2020; Hill & Tyson, 2009).

Parental Educational Expectations

To measure parental expectations, respondents were asked to report the highest level of education they believed their parents expected them to attain (see Appendix B). The original 10-category variable ranged from 1 ("drop out now") to 10 ("get a doctoral degree"). In line with the treatment of the educational aspirations variable (C22), the two lowest-response categories "drop out now" (code 1) and "they don't care" (code 10) were combined into a single "low expectation" category. This recoding is consistent with established practice in education research, where combining semantically and substantively equivalent low-frequency categories can improve interpretability, enhance model stability and avoid inflated standard errors (Agresti, 2010).

The dependant variable captures a spectrum of parental educational expectations, from low (1) to high (9 = "get a doctoral degree"), and is treated as a continuous regressor in linear regression analysis. This treatment is supported by the ordinal scale's sufficient granularity (9 levels), which meets the typical threshold for modelling ordinal data as continuous (Field, 2024; Zumbo & Zimmerman, 1993)

3.4 Research Design and Analytical Approach

Given the mixed numeric and categorical nature of the data and the large initial sample size (N = 9,208), listwise deletion of cases with missing values resulted in an analytic sample of N = 7,971. Sequential linear regression modelling was employed to examine the relationship between teacher expectations and students' educational aspirations, operationalised as the highest level of education students hope to attain (C22). Explanatory variables were introduced in theoretically informed blocks, enabling an assessment of how the association between teacher expectations and aspirations changes as additional covariates are included (Cohen et al., 2013). This stepwise strategy enhances interpretability and allows for evaluation of the extent to which the focal relationship may be confounded or mediated by other factors (Field, 2024; Hahs-Vaughn & Lomax, 2020).

The ordering of variables across the four models was theoretically guided. Model 1 included only the focal independent variable, students perceived teacher expectations (C29), to establish the baseline association with aspirations without the influence of other covariates. Teacher expectations are widely recognised as a strong explanatory factor of students' academic trajectories and aspirations, even in the absence of other controls (Jussim & Harber, 2005).

Model 2 added demographic background characteristics, specifically gender (A01), ethnicity (A03), and socioeconomic status as measured by parental occupation (B8). These background variables are relatively exogenous, shaped by structural and social contexts, and they represent foundational inequalities that may influence both teacher expectations and student aspirations (Sirin, 2005; Gamoran & Long, 2006). Introducing them at this stage allows the analysis to account for structural differences before adding more proximal or potentially mediating factors.

Model 3 incorporated students' cognitive ability test scores (stcog). Cognitive ability is a relatively stable individual trait that is strongly associated with academic achievement and longer-term educational trajectories, and it may also inform teachers' expectations (Carroll, 1993; Deary et al., 2007). It was entered after demographic background variables to ensure that its effect is estimated net of structural factors that often shape the development of cognitive skills (Hoge & Coladarci, 1989).

Model 4 introduced two family-level variables: parental educational expectations (B31) and parental involvement (latent factor scores derived from exploratory factor analysis). These variables capture proximal family processes that can shape students' educational aspirations through the goals parents set and the support they provide. They were entered after background characteristics and cognitive ability because such family processes are often influenced by structural factors and by students' demonstrated academic ability (Eccles, 2005; Fan & Chen, 2001; Jeynes, 2007). Positioning them last ensures that their estimated effects reflect contributions beyond those explained by background and ability and allows the analysis to assess whether teacher expectations remain significantly associated with aspirations once family influences are considered, as well as whether family-level effects operate independently or partly through students' ability.

To assess robustness, a logistic regression model was estimated as a sensitivity check. In this specification, the dependent variable (C22) was dichotomised to indicate whether students aspired to attain a bachelor's degree or higher (coded as 1) versus a lower level of education (coded as 0). The logistic model retained the same set of explanatory variables to determine whether the associations observed in the linear models held under an alternative model (Long & Freese, 2006; Menard, 2010).

Chapter 4: Statistical Analysis and Regression Results

4.1 Summary Statistics

Table 4.1.1 descriptive statistics and percentage of students aspiring to a bachelor's degree or higher by key variables

Variable	Number	Mean	SD	% Bachelor+
Perceived Teachers' Expectation				
Going to a key senior high school	4315	7.2	1.53	82.2
Going to an ordinary senior high school	2583	6.2	1.65	53.7
Going to a vocational high school	587	5.3	1.88	30.5
Going to a technical secondary school/technical school	497	4.6	2.20	23.5
Not going any further after graduating from junior high school	75	4.3	2.59	24.0
None	966	5.5	2.45	44.5
Gender				
Male	4523	6.2	2.12	57.1
Female	4545	6.7	1.73	68.7
Ethnicity				
Han	8338	6.4	1.96	62.8
Minority	790	6.6	1.90	63.2
Parental Occupation				
Government official, staff of publicin, stitutions, civil servant	629	6.8	1.94	73.6
Middle/Senior management personnel of enterprises/corporations	667	6.8	1.78	75.9
Teacher, engineer, doctor, lawyer	561	6.7	1.93	71.1
Technical worker (including driver)	1417	6.4	1.88	62.4
Ordinary staff or worker in production or manufacturing industry	893	6.3	1.91	59.4
Ordinary staff or worker in business or service industry	655	6.4	1.96	61.7
Self-employed worker	1591	6.4	1.95	62.4
Peasant	2159	6.2	2.04	56.1
Unemployed or laid-off worker	186	6.1	2.24	59.7
Other	214	6.4	1.87	58.9
Cognitive Ability				
Q1（lowest ability)	2492	5.8	2.07	46.0
Q2	2630	6.3	1.93	59.2
Q3	2125	6.8	1.76	71.5
Q4 (highest ability)	1725	7.1	1.78	81.4
Parental Involvement				
Q1（lowest involvement）	2112	6.3	2.00	60.7
Q2	2053	6.3	1.94	59.4
Q3	2080	6.4	1.92	62.2
Q4 （highest involvement）	2075	6.4	1.98	62.0
Parental Expectation				
Low expectation (Drop out / Don't care)	358	4.2	3.09	35.2
Graduate from junior high school	94	3.3	2.05	9.6
Go to technical secondary school	400	3.8	1.71	10.0
Go to vocational high school	415	4.9	1.70	18.6
Go to senior high school	1216	5.3	1.79	26.2
Graduate from junior college	1322	6.1	1.40	28.1
Get a bachelor degree	3656	7.0	1.19	88.4
Get a Master degree	1048	7.9	1.16	96.1
Get a Doctor degree	585	8.4	1.65	92.3
Grand mean		6.1	1.91	55.6

Note. N = number of respondents in each category. *Mean* = average score on the educational aspiration variable (C22). *SD* = standard deviation of C22 scores within each category. *% Bachelor+* = percentage of students in each category who reported aspiring to attain a bachelor's degree or higher.

Perceived Teachers' Expectation (C29)

Among students who believed their teacher expected them to go to a key senior high school (N = 4315), the mean aspiration score was 7.2 (SD = 1.53), and 82.2% aspired to a bachelor's degree or higher. This group exhibited the highest aspirations, reinforcing the idea that high perceived teacher expectations are strongly associated with elevated educational goals（Rosenthal & Jacobson, 1968).

Students who perceived their teachers expected them to attend an ordinary senior high school (N = 2583) had a lower average aspiration score of 6.2 (SD = 1.65), with 53.7% aiming for higher education. While aspirations remained moderate, this group already showed a noticeable decline compared to the key high school track.

Students who perceived that teachers expected them to attend vocational high schools (N = 587) or technical secondary schools (N = 497) reported substantially lower aspirations (vocational: M = 5.3, SD = 1.88; technical: M = 4.6, SD = 2.20). Consistent with this, only 30.5% and 23.5% of these students, respectively, aspired to a bachelor's degree or higher. This pattern is compatible with the idea that students internalise lower academic expectations, potentially steering them toward non-academic pathways at an early stage (Hanushek & Wößmann, 2006). However, the association should not be interpreted as causal. It may be driven by common factors, for example, the academic achievement, that might shape both teacher expectations and student aspirations and thus produce the observed correlation (Gentrup et al., 2020). To address this possibility, Subsequent models (2–4) are therefore employed to control for such potential confounders. Students who reported that their teachers expected them to end their education after junior high school (N = 75) had the lowest mean aspiration score of 4.3 (SD = 2.59), with just 24.0% expressing aspirations for higher education.

Interestingly, students who reported receiving no guidance from their teachers ("None"; N = 966) had a moderate mean aspiration score of 5.5 (SD = 2.45), and 44.5% still aspired to bachelor-level education or above. Notably, this percentage exceeds those of students who perceived their teachers encouraged vocational or technical tracks or not going to further education.

This finding prompts several possible interpretations. First, students in the "None" category might derive aspirational cues from other influential sources such as family, peers, or broader cultural norms (Moon, 1996). Second, the "None" response may not reflect an absolute absence of guidance but rather subtle, indirect messaging or student misinterpretation (Rubie-Davies, 2006).

A one-way ANOVA comparing aspirations across the perceived teacher-expectation groups revealed a large overall effect, $\eta^2 = 0.19$ and $\omega^2 = 0.19$ (95% one-sided CI lower bound = 0.18), indicating that approximately 19% of the variance in aspirations is attributable to differences in perceived teacher expectations at the bivariate level. According to conventional benchmarks, this represents a large effect size (Cohen, 1988; Lakens, 2013) and underscores the substantive differences in aspirations observed across the academic tracks.

Demographic Controls

Gender

Gender-based differences in educational aspiration were evident. Female students (N = 4545) reported a higher mean aspiration score (M = 6.7, SD = 1.73) than male students (N = 4523, M = 6.2, SD = 2.12). Furthermore, 68.7% of females aspired to a bachelor's degree or higher, compared to 57.1% of males. The mean difference corresponded to a small but statistically significant effect size, d = −0.24, 95% CI [−0.28, −0.20], indicating higher aspirations among females. These findings align with long-standing meta-analytic evidence showing that female students consistently outperform males in school grades across multiple educational levels (Voyer & Voyer, 2014). The narrower standard deviation among females in our data also suggests a more concentrated clustering around higher aspiration levels, whereas the broader spread among males reflects greater variability and potential polarization in aspirations (Field, 2024).

Ethnicity

A small but statistically significant difference in educational aspiration was observed between ethnic groups. Han students (N = 8,338) reported a mean aspiration score of 6.4 (SD = 1.96), with 62.8% aspiring to a bachelor's degree or higher. Minority students (N = 790) demonstrated slightly higher aspirations on average (M = 6.6, SD = 1.90), with 63.2% aiming for university-level education. The mean difference corresponded to a small effect size, d = 0.10, 95% CI [0.02, 0.17], indicating marginally higher aspirations among minority students.

Although the magnitude of the effect is small, these findings challenge stereotypes that ethnic minority students hold lower educational goals and align with research suggesting

that, when structural opportunities are accounted for, Chinese minority youth often display aspirations comparable to or exceeding those of their Han peers. For example, historical data show that post-reform-era enrolment among minorities aligned closely with the Han majority (Hannum, 2002). A longitudinal study of Hui and Kazakh adolescents in northwestern China reported no significant differences in future aspirations compared to Han youth (Zhang & Ng, 2025). Similarly, research on Yi students found that strong educational aspirations are a key driver of academic success (Wang, 2024). Chia (2023) also highlights that minority youth in certain regions outperform their Han counterparts in aspiration measures when socioeconomic and regional contexts are considered.

Socioeconomic Status (SES) - parental occupation

Students' educational aspirations varied by their parents' occupational background. The highest aspirations were found among students whose parents were government officials or civil servants (N = 629), with a mean score of 6.8 (SD = 1.94) and 73.6% aspiring to a bachelor's degree or higher. Similar high aspirations were seen in children of middle or senior corporate managers (N = 667; M = 6.8, SD = 1.78; 75.9% bachelor+), and professionals such as teachers, doctors, and lawyers (N = 561; M = 6.7, SD = 1.93; 71.1%).

In contrast, students from more manual or lower-status occupational backgrounds showed comparatively lower aspirations. Technical workers, including drivers (N = 1417), reported a mean score of 6.4 (SD = 1.88) with 62.4% aspiring to university. Those whose parents were ordinary production workers (N = 893) or business/service staff (N = 655) had mean scores of 6.3 and 6.4 respectively, with bachelor+ rates of 59.4% and 61.7%. Self-employed families (N = 1591) also fell in a mid-range category (M = 6.4, SD = 1.95; 62.4%), while children of peasants (N = 2159) and unemployed/laid-off workers (N = 186) reported the lowest aspirations (M = 6.2, SD = 2.04 and M = 6.1, SD = 2.24), with only 56.1% and 59.7% aspiring to higher education. The "Other" category (N = 214) had a mean score of 6.4 (SD = 1.87) and 58.9% bachelor+ rate.

A one-way ANOVA indicated that parental occupation had a statistically significant but small effect on students' aspirations, $\eta^2 = .03$, $\omega^2 = .03$, 95% CI [.03, 1.00], suggesting that approximately 3% of the variance in aspirations was attributable to differences in parental occupational categories. While the magnitude of this effect is modest, it is consistent with prior research showing that parental occupational prestige, often reflecting broader socioeconomic resources, shapes students' educational trajectories through both material advantages and normative expectations (Sirin, 2005). In the Chinese context, occupational status has been

shown to influence not only access to educational opportunities but also the transmission of academic values, thereby reinforcing aspirations toward higher education (Tan & Fang, 2023).

Cognitive Ability

Students' educational aspirations showed clear variation across levels of cognitive ability. Based on test scores, students were grouped into quartiles.

Among those in the lowest quartile (Q1, N = 2492), the average aspiration score was 5.8 (SD = 2.07), and only 46.0% aspired to a bachelor's degree or higher. In the second quartile (Q2, N = 2630), the mean aspiration score increased to 6.3 (SD = 1.93), with 59.2% aspiring to university-level education. Students in the third quartile (Q3, N = 2125) reported a further increase in aspirations, with a mean of 6.8 (SD = 1.76) and 71.5% aiming for a bachelor's degree or beyond. Finally, those in the highest quartile (Q4, N = 1725) demonstrated the strongest educational goals, with a mean aspiration score of 7.1 (SD = 1.78), and 81.4% aspiring to university.

An ANOVA confirmed that these differences were statistically significant, with an effect size of $\eta^2 = .05$, 95% CI [.05, 1.00], and $\omega^2 = .05$, indicating a small-to-moderate association (Field, 2024). This suggests that cognitive ability accounts for approximately 5% of the variance in students' aspirations. These findings align with prior research showing that higher cognitive performance is positively associated with elevated educational aspirations among Chinese adolescents (Zhou et al., 2019).

Family-Level Variables

Parental Involvement

When parental involvement was grouped into quartiles, no strictly monotonic pattern emerged, although averages were marginally higher in the upper quartiles. In the lowest involvement group (Q1, N = 2112), students reported a mean aspiration score of 6.3 (SD = 2.00), and 60.7% aspired to a bachelor's degree or higher. In Q2 (N = 2053), the mean remained at 6.3 (SD = 1.94), with 59.4% aspiring to higher education. Slight increases appeared in the higher quartiles. In Q3 (N = 2080), the average aspiration score rose to 6.4 (SD = 1.92), with 62.2% aiming for a university degree. A similar level was observed in Q4 (N = 2075), where the mean score was again 6.4 (SD = 1.98), and 62.0% aspired to a bachelor's or higher degree.

A one-way ANOVA indicated that differences in aspirations across parental involvement quartiles were minimal, with a very small effect size ($\eta^2 = 0.00024$, $\omega^2 \approx 0$). This suggests that while a slight upward trend was observable in the descriptive statistics, parental involvement as measured here accounts for virtually none of the variance in students' aspirations.

Several factors may explain this negligible association. First, the parental involvement index in this study primarily reflects behavioural aspects such as checking homework, attending school meetings, or discussing school events. Prior research indicates that such forms of involvement, while beneficial for achievement, are less predictive of aspirations compared to motivational or future-oriented involvement, such as fostering autonomy or discussing career goals (Hill & Tyson, 2009). Second, in the Chinese context, parental engagement in schoolwork is widespread and relatively homogeneous across socioeconomic groups, which limits variability and reduces its explanatory power (Fan & Chen, 2001). Finally, parental involvement may exert its influence on aspirations indirectly through academic performance, motivation or self-efficacy, rather than via a direct pathway (El Nokali et al., 2010).

Parental Expectations

Students' educational aspirations were sharply stratified by their parents' expectations. A strong positive gradient was evident across the educational expectation spectrum. At the lowest end, students whose parents were perceived to hold low expectations such as encouraging dropout or expressing indifference (N = 358) had a mean aspiration score of 4.2 (SD = 3.09), with only 35.2% aspiring to a bachelor's degree or higher. Similarly, aspirations were extremely low among students whose parents expected them to only graduate from junior high (N = 94; M = 3.3, SD = 2.05; 9.6% bachelor+), attend technical school (N = 400; M = 3.8, SD = 1.71; 10.0%), or vocational school (N = 415; M = 4.9, SD = 1.70; 18.6%).

A more moderate level of aspiration appeared among students whose parents expected them to complete senior high school (N = 1216). These students had a mean score of 5.3 (SD = 1.79), and 26.2% aspired to a bachelor's degree or above. Students whose parents expected them to complete junior college (N = 1322) showed slightly higher aspirations (M = 6.1, SD = 1.40), with 28.1% aspiring to university education. The highest levels of aspiration were found among students whose parents expected them to earn a bachelor's degree (N = 3656; M = 7.0, SD = 1.19; 88.4%), a master's degree (N = 1048; M = 7.9, SD = 1.16; 96.1%), or a doctoral degree (N = 585; M = 8.4, SD = 1.65; 92.3%).

A one-way ANOVA confirmed that these differences were statistically significant and corresponded to a very large effect size, $\eta^2 = .40$, 95% CI [.39, 1.00], $\omega^2 = .40$, 95% CI [.39, 1.00]. According to Cohen's (2013)

guidelines, this value far exceeds the conventional benchmark for a large effect, indicating that parental expectations account for approximately 40% of the variance in students' educational aspirations. This suggests that parental expectations are one of the most powerful correlates examined in this study.

Higher expectations appear to serve as motivational anchors, with students internalising those goals and aiming accordingly. However, it is notable that among students whose parents held the most ambitious expectations at the doctoral level, the proportion aspiring to higher education (92.3%) was slightly lower than in the master's expectation group (96.1%). This pattern may reflect a "Goldilocks effect," where expectations that are excessively high can become counterproductive because they exceed what students perceive as realistic, which may lead to disengagement or resistance (Marsh et al., 2023).

Conversely, a significant portion of students (35.2%) still aspired to university education despite minimal or absent parental expectations. This suggests that motivation can emerge from sources beyond the family, such as peer influence, personal ambition, and encouragement from teachers. Research in rural Chinese contexts has found that peer aspirational norms and close friends' goals significantly predict adolescents' educational aspirations, often mediated through students' self-regulation and self-concept, even when parental expectations are low (Chen et al., 2023). Similarly, studies using nationally representative United States data indicate that teacher expectations can compensate for weaker parental expectations in influencing academic outcomes (Benner et al., 2021). In addition, peer groups tend to reinforce educational norms among adolescents, forming homogeneous clusters of aspiration even in contexts where parental pressure is limited (Mundt, 2020).

In summary, the analysis demonstrates that parental expectations are a powerful and consistent explanatory variable of students' educational aspirations, explaining a substantial proportion of variance in aspiration levels. The positive gradient across expectation levels underscores the motivational role of family vision in shaping academic goals.

4.2 Regression Results

Table 4.2.1 sequential linear regression models for predicting educational aspirations

Group	Variable	model 1			model 2			model 3			model 4		
		Coeff.	SE	Sig.	Coeff.	SE	Sig.	Coeff.	SE	Sig.	Coeff.	SE	Sig.
Perceived Teachers' Expectation	Intercept	7.13	0.028	***	6.67	0.051	***	6.17	0.071	***	2.91	0.089	***
base = Going to a key senior high school	Going to an ordinary senior high school	-0.10	0.047	***	-0.94	0.046	***	-0.85	0.047	***	-0.41	0.042	***
	Going to a vocational high school	-1.75	0.081	***	-1.64	0.080	***	-1.49	0.081	***	-0.70	0.072	***
	Going to a technical secondary school/technical school	-2.52	0.087	***	-2.41	0.087	***	-2.24	0.088	***	-1.14	0.079	***
	Not going any further after graduating from junior high school	-2.85	0.210	***	-2.68	0.208	***	-2.49	0.208	***	-1.45	0.181	***
	None	-1.74	0.067	***	-1.65	0.066	***	-1.54	0.067	***	-0.81	0.060	***
Gender (base = male)	female				0.31	0.039	***	0.33	0.039	***	0.25	0.034	***
Ethnicity (base = Han)	ethnic minority				0.31	0.070	***	0.37	0.070	***	0.27	0.061	***
Parental Occupation	Government official, staff of public institutions, civil servant				0.65	0.090	***	0.49	0.091	***	0.28	0.079	***
base = peasant	Middle/Senior management personnel of enterprises/corporations				0.62	0.083	***	0.45	0.084	***	0.18	0.073	*
	Teacher, engineer, doctor, lawyer				0.56	0.089	***	0.39	0.090	***	0.23	0.078	**
	Technical worker (including driver)				0.23	0.063	***	0.13	0.063	*	0.06	0.055	
	Ordinary staff or worker in production or manufacturing industry				0.14	0.072	.	0.06	0.072		0.05	0.062	
	Ordinary staff or worker in business or service industry				0.32	0.082	***	0.23	0.082	**	0.11	0.072	
	Self-employed worker				0.32	0.061	***	0.22	0.062	***	0.14	0.054	*
	Unemployed or laid-off worker				0.01	0.140		-0.05	0.139		0.00	0.121	
	Other				-0.23	0.131	.	-0.25	0.130	.	-0.15	0.113	
Cognitive Ability	Cognitive Ability							0.05	0.005	***	0.03	0.005	***
Family-level Factor	Parental Involvement										0.04	0.020	*
	Parental Expectation										0.53	0.010	***
R-squared		0.1806		***	0.2000		***	0.2106		***	0.4053		***
degree of freedom		7965			7954			7953			7951		
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1													

Before running the regressions, the relevant assumptions were tested and met. Details are provided in Appendix C.

Model 1: Perceived Teacher Expectations and Student Aspirations

Model 1 examined the bivariate association between perceived teacher expectations and students' educational aspirations. The regression model was statistically significant, $F(5, 7965) = 350.6$, $p < .001$, and explained 18.1% of the variance in aspirations ($R^2 = .1806$), indicating that teacher expectations alone accounted for a substantial proportion of variability in students' aspirations regarding educational attainment.

Relative to the reference group, all other perceived teacher expectations were associated with significantly lower levels of aspiration. Students who believed their teachers expected them to attend an ordinary senior high school reported aspirations that were, on average, 0.10 points lower. Those who perceived expectations to attend a vocational high school scored 1.75 points lower, while perceived expectations for technical secondary school were associated with a 2.52-point decrease. Students who believed their teachers expected them not to pursue any further education after junior high reported the lowest aspirations, with scores 2.85 points below the reference group. Notably, even students who indicated that their teachers had expressed no clear educational expectations scored significantly lower than the reference group, by 1.74 points. Interestingly, students in the "no expectation" group still reported higher aspiration scores than those whose teachers were perceived to expect no further education at all, suggesting that ambiguity may be less detrimental than explicitly low expectations.

This graded pattern is consistent with the "Pygmalion effect" proposed by Rosenthal and Jacobson (1968), which posits that high teacher expectations can elevate student achievement and aspirations through self-fulfilling mechanisms. Conversely, the "Golem effect" describes how low expectations can diminish student motivation and perceived potential (Babad et al., 1982). Meta-analytic evidence indicates that interventions targeting teacher expectations, particularly those aimed at reshaping beliefs about student capability, can improve student outcomes (De Boer et al., 2018). Furthermore, longitudinal research indicates that teacher expectations are robustly associated with students' self-concept and long-term goals (Rubie-Davies, 2014) and remain linked to subsequent educational trajectories even after adjusting for family background (Benner et al., 2021).

Model 2: Adding Demographic Controls

Model 2 extended the baseline specification by incorporating three important demographic covariates: gender, ethnicity, and parental occupation. The model was statistically significant, $F(16, 7954) = 124.1$, $p < .001$, and explained 20.0% of the variance in educational aspirations ($R^2 = .2000$), marking a modest increase from Model 1 ($R^2 = .1806$). This suggests that gender, ethnicity, and parental socioeconomic background contribute additional explanatory power, though only incrementally beyond the effect of perceived teacher expectations alone.

Students who perceived their teachers as expecting them to pursue lower-tier educational pathways continued to report significantly lower aspirations than those in the reference group (students expected to attend an academic key high school). Compared with the reference group, students who perceived expectations for ordinary senior high school scored 0.94 points lower ($B = –0.94$, $p < .001$), while those associated with vocational or technical school expectations scored 1.64 and 2.41 points lower, respectively ($p < .001$). Students who believed their teachers expected them to stop education after junior high reported the lowest aspirations, 2.68 points lower, whereas those with no clear expectations still scored 1.65 points below the reference group. This consistent gradient suggests that both the content and clarity of perceived teacher expectations act as important motivational cues for students, echoing evidence that teacher beliefs shape aspirations even after controlling for sociodemographic characteristics (Benner et al., 2021; Rubie-Davies, 2014).

Gender emerged as a significant covariate, with female students reporting higher aspirations than their male peers ($B = 0.31$, $p < .001$). This is consistent with broader evidence showing that girls in China and internationally tend to have higher educational aspirations, often attributed to shifting labour market structures and changing gender norms (Buchmann et al., 2008; Voyer & Voyer, 2014; Wei & Ma, 2017).

Similarly, minority status showed a significant positive relationship with aspirations, with minority students reporting higher levels than Han peers ($B = 0.31$, $p < .001$). This accords with prior work attributing such patterns to mobility motivations and targeted policy initiatives in some regions (Hannum, 2002; Zhang & Ng, 2025).

Parental occupation, used as a proxy for socioeconomic status, was significantly associated with student aspirations. In this model, the reference category was students whose parents were classified as peasants. Compared to this group, students whose parents held public-sector jobs (civil servants or institutional staff)

reported higher aspirations (B = 0.65, p < .001), as did those from middle or senior management (B = 0.62, p < .001) and professional backgrounds such as teachers, doctors, or engineers (B = 0.56, p < .001). Children of technical workers (B = 0.23, p < .001), service sector workers (B = 0.32, p < .001), and self-employed workers (B = 0.32, p < .001) also reported higher aspirations. Students whose parents fell into the "Other" occupational category reported slightly lower aspirations (B = –0.23), while those whose parents worked as ordinary staff in production or manufacturing reported slightly higher aspirations (B = 0.14); both associations only approached marginal significance (p < .1). By contrast, students whose parents were unemployed or laid off reported no significant difference in aspirations compared to the peasant group (B = 0.01, p > .1).

These patterns are consistent with prior evidence that higher parental occupational status is positively associated with adolescents' educational and career ambitions, partly through greater access to resources and role modelling (Plenty, 2021). This aligns with the Wisconsin model of status attainment, which emphasises that parental occupational status and expectations transmit social advantage and influence the formation of adolescent aspirations (Haller & Woelfel, 1972; Sewell & Hauser, 1980; Sirin, 2005).

Most crucially, the effect of perceived teacher expectations remained strong, statistically robust and structurally consistent even after adjusting for demographic controls. Coefficients for each expectation category slightly attenuated relative to Model 1 but all remained highly significant, preserving the original rank order of aspiration levels. This suggests that teacher expectations function as an independent motivational force, shaping students' aspirations regardless of their gender, ethnicity or family background (Good et al., 2018).

Model 3: Adding Cognitive Ability

Building on Model 2, Model 3 introduced cognitive ability to assess whether differences in aspirations are partly attributable to students' academic competence. The model was statistically significant, F (17, 7953) = 124.6, p < .001, accounting for 21.1% of the variance in aspirations (R^2 = .2106), a modest increase from Model 2 (R^2 = .2000). Cognitive ability emerged as a positive, albeit modest, explanatory variable (B = 0.05, SE = 0.005, p < .001), aligning with prior evidence that stronger cognitive skills are linked to higher educational goals through enhanced academic performance and perceived competence (Liang et al., 2018; Ones et al., 2012).

Perceived teacher expectations still retained a strong, graded association with aspirations. Compared to students whose teachers were perceived to expect them to attend a key senior high school, aspirations were 0.85 points lower for ordinary senior high school expectations ($B = –0.85$, $p < .001$), 1.49 points lower for vocational high school ($p < .001$), and 2.24 points lower for technical secondary school ($p < .001$). Students who believed their teachers expected them to end education after junior high scored 2.49 points lower, and those with no clear expectations scored 1.54 points lower (both $p < .001$).

Compared with Model 2, coefficients for all expectation categories were slightly reduced, suggesting that part of the effect of perceived teacher expectations on aspirations overlaps with cognitive ability. This attenuation likely reflects that teacher expectations are partially shaped by student cognitive performance (Barry et al., 2022; Brophy & Good, 1974). Nonetheless, the persistence of large, statistically significant effects supports the Pygmalion effect framework, whereby teacher expectations influence students' self-beliefs, effort and ultimate attainment beyond the contribution of cognitive skills (Rosenthal & Jacobson, 1968; Rubie-Davies, 2006, 2010, 2014).

In relation to Model 2, the gender gap in aspirations widened slightly ($B = 0.31 \rightarrow 0.33$, $p < .001$), with female students continuing to report higher aspirations than males. This finding aligns with research showing rising female educational ambition in China and globally, often linked to shifting gender norms and expanded opportunities for women (Del Pero & Bytchkova, 2013; Liu, 2020). The increase in the coefficient after controlling for cognitive ability reflects a suppression effect, in which adding a control variable, in this case, cognitive ability, removes variance that is negatively associated with the independent variable (female gender) but positively associated with the dependent variable (educational aspiration) (MacKinnon et al., 2000). Given that female students, on average, score slightly lower on standardised cognitive tests in the CEPS dataset yet display stronger non-cognitive motivation and academic persistence (Duckworth & Seligman, 2005), cognitive ability in Model 2 had masked part of their relative advantage in aspirations, which became more apparent once this factor was held constant.

Ethnic minority students also exhibited a larger positive coefficient relative to Model 2 ($B = 0.31 \rightarrow 0.37$, $p < .001$). This pattern likewise reflects a suppression effect: minority students tend to score slightly lower on standardised cognitive assessments in the CEPS dataset, and cognitive ability is positively related to aspirations. Controlling for cognitive ability therefore removes a source of negative covariance between minority status and aspirations, amplifying the observed positive association (MacKinnon et al., 2000). This finding is consistent with evidence that minority students' high aspirations may be fuelled by strong mobility motives and reinforced by targeted educational support policies (Hannum, 2002; Zhang & Ng, 2025).

Turning to parental occupation, almost all significant coefficients were positive, indicating that students from non-agricultural households held higher aspirations than those from peasant households. The strongest effects were for government/public-sector employees ($B = 0.49$, $p < .001$), middle or senior enterprise management ($B = 0.45$, $p < .001$), and professional occupations such as doctors or teachers ($B = 0.39$, $p < .001$). Positive associations were also observed for service-sector clerical jobs ($B = 0.23$, $p < .01$), self-employment ($B = 0.22$, $p < .001$), and technical work ($B = 0.13$, $p < .05$). By contrast, manufacturing workers ($B = 0.06$, $p > .1$) and the unemployed/laid-off group ($B = –0.05$, $p > .1$) did not differ significantly from peasants, while the "other" category reported marginally lower aspirations ($B = –0.25$, $p < .1$). Compared with Model 2, most coefficients decreased, suggesting that part of the association between parental occupation and aspirations is mediated through cognitive ability.

Overall, Model 3 indicates that while cognitive ability accounts for only a modest and the perceived teacher expectations remain a strong association.

Model 4: Adding Family-Level Values

Model 4 incorporated family-level educational values operationalised through parental educational expectations and parental involvement. The model was statistically significant, F (19, 7951) = 284.8, $p < .001$, and explained 40.5% of the variance in educational aspirations ($R^2 = .4053$), marking a substantial increase from Model 3 ($R^2 = .2106$). This notable jump in explained variance suggests that family-level orientations towards education account for a large proportion of students' academic ambition (Englund et al., 2004; Wei and Ma, 2017).

Among family level variables, parental educational expectation showed the largest partial association with aspirations in the fully adjusted model ($B = 0.53$, $p < .001$). Students who perceived that their parents aspired for higher educational attainment reported substantially greater aspirations than peers whose parents held lower expectations. In contrast, parental involvement in the form of routine support behaviours was statistically significant but much smaller in magnitude ($B = 0.04$, $p < .05$). This pattern suggests that not all forms of involvement relate to aspirations to the same extent, with aspirational and academically communicative practices appearing more closely linked than purely directive or routine support (Jung & Zhang, 2016; Karbach et al., 2013).

Even after adding family-level values, perceived teacher expectations retained a strong and graded relationship with aspirations. Compared with students whose teachers were perceived to expect them to attend a key senior high school, aspirations were 0.41 points lower for ordinary senior high expectations (B = –0.41, $p < .001$), 0.70 points lower for vocational high school expectations (B = –0.70, $p < .001$), and 1.14 points lower for technical secondary school expectations (B = –1.14, $p < .001$). Students who believed their teachers expected them to end education after junior high reported aspirations 1.45 points lower, and those perceiving no clear teacher expectation scored 0.81 points lower (both $p < .001$). These results mirror the persistent influence observed in prior models and echo evidence that teacher expectations convey unique motivational signals beyond family influence (Rosenthal & Jacobson, 1968; Rubie-Davies, 2006, 2010, 2014).

Gender and ethnicity effects also persisted. Boys reported aspirations 0.25 points lower than girls ($p < .001$), and ethnic minority students reported aspirations 0.27 points higher than Han students ($p < .001$). The slight reductions in these coefficients compared to Model 3 suggest that part of the gender and ethnicity differences in aspirations is explained by family-level values, echoing prior findings that supportive home environments can narrow demographic aspiration gaps (Buchmann et al., 2008; Hannum, 2002).

A comparison with Model 3 revealed that the inclusion of parental educational expectations and involvement reduced the size and significance of most parental occupation coefficients, indicating mediation effect (Field, 2024). The strong effects previously attributed to high-status occupations appear to be channelled through the educational norms and aspirations held within families (Sewell & Hauser, 1980; Sirin, 2005). Specifically, the coefficient for government/public-sector employees declined from 0.49 to 0.28 ($p < .001$), for middle or senior management from 0.45 to 0.18 ($p < .01$), and for professional occupations such as teachers, engineers, doctors and lawyers from 0.39 to 0.23 ($p < .05$). Notably, except for government/public-sector employees, all other occupational groups no longer reached the $p < .001$ significance threshold, underscoring the dominant role of family-level values over occupational status in predicting aspirations.

Turning to cognitive ability, it remained a positive and significant effect (B = 0.03, $p < .001$), although the coefficient was slightly attenuated relative to Model 3. This attenuation suggests that part of the association between cognitive ability and aspirations is shared with family-level educational values. Students with higher cognitive scores are often more likely to have parents who hold ambitious educational goals for them and to experience greater parental involvement, both of which can, in turn, elevate their aspirations (Neuenschwander et al., 2007; Mistry et al., 2009).

Overall, Model 4 demonstrates that parental educational expectations are the most powerful explanatory variable of student aspirations, explaining a substantial portion of the variance in educational goals (Kirk et al., 2011). Yet, teacher expectations remain robust s even after accounting for ability, background and family values, underscoring their complementary role in shaping students' academic futures, particularly for students who may not receive strong encouragement at home (Fan & Chen, 2001; Jeynes, 2007).

Figure 4.2.2 coefficient estimates for the association between perceived teacher expectations and students' educational aspirations across sequential regression models

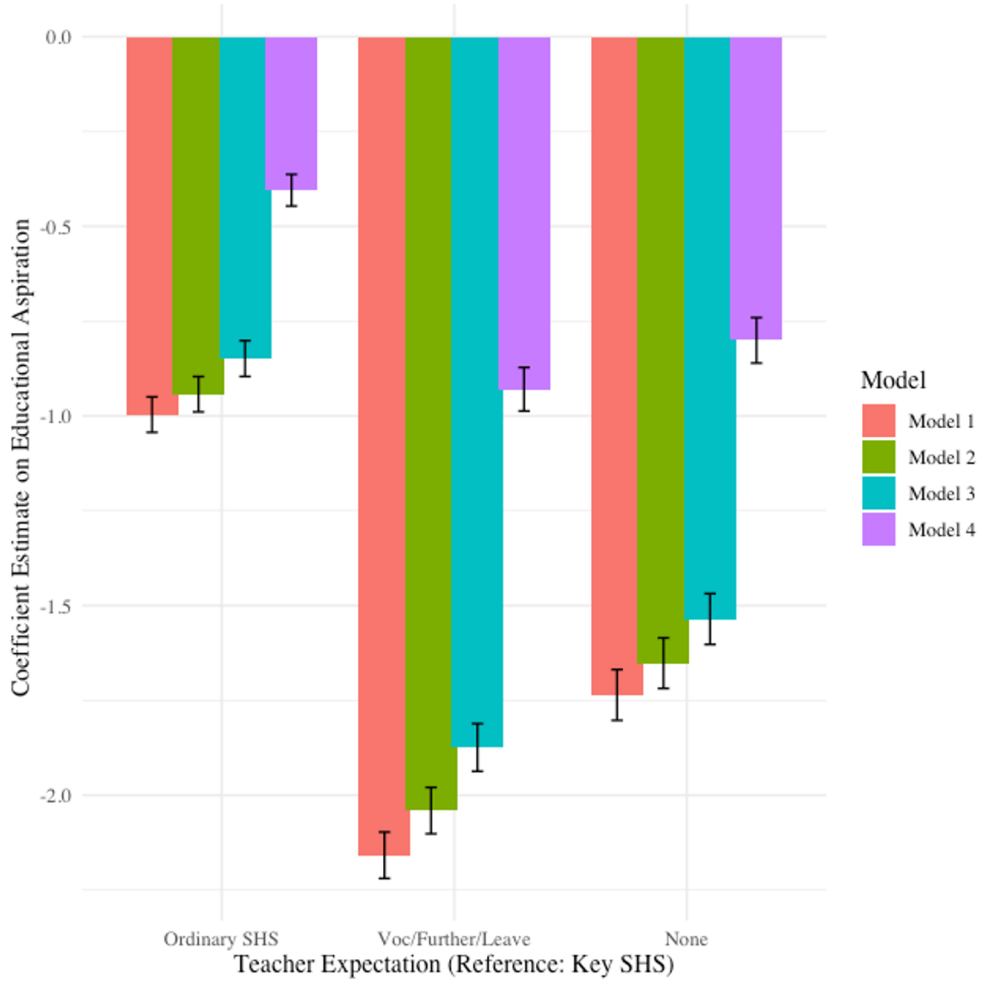


Figure 4.2.2 visually complements the regression tables by illustrating both the magnitude and pattern of coefficients across sequential models. It presents the estimated coefficients for perceived teacher expectations on students' educational aspirations across four sequential linear models, using students who perceived expectations to attend a key senior high school (Key SHS) as the reference group. Across all models, students in the Ordinary SHS, Vocational/Technical/No Further Education (Voc/Further/Leave), and None categories consistently reported significantly lower aspirations than those in the Key SHS group.

The figure offers two insights beyond the tabulated results. First, it reveals a clear graded pattern: as perceived teacher expectations decline, students' aspirations fall proportionally, with the steepest gap observed for the Vocational/Technical/No Further Education category, whose Model 1 coefficients approach –2 points relative to the reference group. Second, it shows a progressive attenuation of coefficients as theoretically ordered covariates are added. Demographic controls in Model 2 reduce effects only slightly, suggesting that background characteristics explain little of the association. The inclusion of cognitive ability in Model 3 yields further, though modest reductions, indicating that cognitive ability accounts for only a limited portion of the teacher expectation–aspiration link. The most pronounced attenuation occurs in Model 4 with the addition of family-level values, particularly parental expectations, which, as noted earlier, explain a substantial share of the variation in aspirations.

Despite this attenuation, all coefficients remain statistically significant in the fully adjusted model, underscoring the independent predictive value of teacher expectations. The persistence of these effects aligns with the Pygmalion effect framework, in which teacher expectations shape students' self-beliefs, motivation and eventual attainment beyond structural background, measured ability, and family involvement (Jussim & Harber, 2005; Rosenthal & Jacobson, 1968; Rubie-Davies, 2010, 2014). The sharp drop between Models 3 and 4 underscores the central role of family-level expectations in aspiration formation, suggesting that home-based values can amplify or buffer school-based influences, yet still leaving a sizeable residual effect uniquely attributable to teacher expectations.

4.3 Robustness Checking

Table 4.3.1 logistic regression models predicting the likelihood of aspiring to a bachelor's degree or higher

Group	Variable	model 1				model 2				model 3				model 4			
		Coeff.	SE	Sig.	Odd Ratio	Coeff.	SE	Sig.	Odd Ratio	Coeff.	SE	Sig.	Odd Ratio	Coeff.	SE	Sig.	Odd Ratio
Perceived Teachers' Expectation	Intercept	1.46	0.041	***	4.29	0.76	0.067	***	2.13	-0.08	0.09		0.09	-4.24	0.171	***	0.01
base = Going to a key senior high school	Going to an ordinary senior high school	-1.34	0.059	***	0.26	-1.32	0.061	***	0.27	-1.20	0.06	***	0.30	-0.83	0.068	***	0.44
	Going to a vocational high school	-2.27	0.102	***	0.10	-2.17	0.104	***	0.12	-1.96	0.11	***	0.14	-1.31	0.118	***	0.27
	Going to a technical secondary school/technical school	-2.63	0.118	***	0.07	-2.54	0.120	***	0.08	-2.30	0.12	***	0.10	-1.44	0.139	***	0.24
	Not going any further after graduating from junior high school	-2.61	0.281	***	0.07	-2.42	0.287	***	0.09	-2.16	0.29	***	0.12	-1.49	0.337	***	0.23
	None	-1.74	0.081	***	0.18	-1.69	0.083	***	0.19	-1.54	0.08	***	0.22	-0.99	0.097	***	0.37
Gender (base = male)	female					0.46	0.052	***	1.58	0.50	0.05	***	1.65	0.50	0.058	***	1.64
Ethnicity (base = Han)	ethnic minority					0.18	0.093	.	1.20	0.28	0.09	**	1.32	0.21	0.103	*	1.23
Parental Occupation	Government official, staff of public institutions, civil servant					1.34	0.140	***	3.83	1.13	0.14	***	3.08	1.13	0.167	***	3.11
	Middle/Senior management personnel of enterprises/corporations					1.18	0.120	***	3.24	0.92	0.12	***	2.50	0.73	0.138	***	2.07
	Teacher, engineer, doctor, lawyer					0.90	0.126	***	2.47	0.64	0.13	***	1.89	0.57	0.147	***	1.76
	Technical worker (including driver)					0.43	0.081	***	1.53	0.26	0.08	**	1.30	0.20	0.091	*	1.22
	Ordinary staff or worker in production or manufacturing industry					0.25	0.092	**	1.28	0.12	0.09		1.13	0.14	0.103		1.15
	Ordinary staff or worker in business or service industry					0.53	0.107	***	1.69	0.38	0.109	***	1.47	0.32	0.121	**	1.38
	Self-employed worker					0.58	0.080	***	1.79	0.44	0.08	***	1.55	0.39	0.091	***	1.47
	Unemployed or laid-off worker					0.36	0.178	*	1.44	0.27	0.18		1.31	0.35	0.202	.	1.42
	Other					-0.31	0.169	.	0.73	-0.35	0.170	*	0.70	-0.32	0.186	.	0.73
Cognitive Ability	Cognitive Ability									0.10	0.01	***	1.10	0.08	0.008	***	1.08
Family-level Factor	Parental Involvement													0.05	0.034		1.05
	Parental Expectation													0.67	0.022	***	1.95
R-squared		0.2181		***		0.2595		***		0.2821		***		0.4528		***	
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1																	

As a robustness check, educational aspiration was recoded as a binary outcome: aspiring to a bachelor's degree or higher (coded as 1) versus below that level (coded as 0). Logistic regression is well-suited for this purpose, as it can validate patterns observed in linear models when the dependent variable is categorical (Long & Freese, 2006). The logistic regression results closely mirrored those from the linear models, reinforcing the robustness of the earlier findings. Across all models, key correlators retained both their statistical significance and relative importance, and model fit improved substantially, with pseudo-R^2 increasing from 0.2181 in Model 1 to 0.4528 in Model 4.

Relative to students who perceived their teachers expected them to attend a key senior high school, the odds of aspiring to university were substantially lower for those perceiving expectations of an ordinary senior high school (OR = 0.44, $p < .001$), a vocational high school (OR = 0.27, $p < .001$), a technical secondary school (OR = 0.24, $p < .001$), or no further education after junior high (OR = 0.23, $p < .001$). Students reporting no clear teacher expectation also had lower odds (OR = 0.37, $p < .001$). In general, lower perceived expectations corresponded to reduced aspirations. Interestingly, the "no expectation" group reported higher odds than those with explicitly low expectations, which may suggest that ambiguity, and thus a lack of clear guidance is comparatively less restrictive than explicit discouragement. This pattern echoes broader findings in expectation research, where ambiguity is often less harmful than unequivocally low expectations (Brophy, 1970). This is consistent with the "Pygmalion effect" (Rosenthal & Jacobson, 1968) and meta-analytic evidence showing that high teacher expectations have a strong, graded relationship with student aspirations (Benner et al., 2021; Rubie-Davies, 2014).

Gender effects were consistent across methods: male students had significantly lower odds of aspiring to university than females (OR = 1.64, $p < .001$), echoing the gender gap in the linear models (B = 0.25 in Model 4). This pattern reflects the trends that girls often hold higher educational aspirations, partly due to changing gender norms and labour market shifts (Buchmann et al., 2008). Similarly, ethnic minority students maintained higher odds compared to Han students (OR = 1.23, $p < .05$), closely paralleling the linear results (B = 0.27 in Model 4), and aligning with findings that minority students in China may show equal or higher aspirations due to mobility motives and supportive educational policies (Hannum, 2002).

Occupational effects in Model 4 remained directionally consistent with the linear regressions. The three occupational categories with the highest odds were parents working as government officials or in the public sector (OR = 3.11, $p < .001$), middle or senior managers in enterprises (OR = 2.07, $p < .001$), and professionals such as teachers, engineers, doctors, or lawyers (OR = 1.76, $p < .001$). The public-sector category retained the largest advantage, reinforcing its persistent positive association with aspirations. The attenuation or loss of significance for several other occupational categories after the introduction of parental

expectations and parental involvement suggests that occupational influences are partly mediated by family educational values (Sewell & Hauser, 1980; Sirin, 2005).

Cognitive ability remained a significant correlate of aspiration in the logistic model (OR = 1.08, $p < .001$), indicating that a one-unit increase on the cognitive-ability scale was associated with an 8% increase in the odds of expressing higher aspirations. The magnitude of this association was smaller than the corresponding effects for teacher and parental expectations, a pattern consistent with the idea that ability may shape aspiration partly indirectly through performance feedback and the expectations that key adults form about the student (Ones & Viswesvaran, 2012).

Among family-level variables, parental educational expectation showed the strongest relationship with aspirations (OR = 1.95, $p < .001$), indicating that students whose parents aspired for them to attain higher education had nearly twice the odds of aiming for a bachelor's degree or above. This pattern is consistent with evidence from China and other contexts that parental expectations are among the most salient influences on students' educational goals (Kirk et al., 2011). Parental involvement (OR = 1.05, $p > .10$) was not statistically significant, suggesting that any positive link with aspiration is small and imprecisely estimated in this sample (Fan & Chen, 2001; Jeynes, 2007).

Taken together, results from both the sequential linear regression and the logistic-regression robustness check point in the same direction. Perceived teacher expectations and parental educational expectations are the variables most strongly associated with students' educational aspirations, and these associations remain after adjusting for demographic background, cognitive ability, and family involvement. While these background characteristics help account for variation in aspirations, they do not remove the graded association observed for teacher expectations. Consistency across modelling approaches increases confidence in the robustness of the observed associations and highlights the salience of adult expectations in adolescents reported educational goals.

Chapter 5: Discussion

5.1 General Finding discussion

Using a nationally representative Grade 9 sample from CEPS, this study finds a clear ordered association: students who perceive higher teacher expectations report higher educational aspirations. The association remains after adjusting for demographics, cognitive ability, and family factors, and it also appears in logistic models. The finding is consistent with international evidence on expectation effects and with Chinese scholarship that treats teachers as key socialising agents shaping students' self-beliefs and goals (Wang et al., 2018; Wang et al., 2021). Viewed through this theoretical lens, the findings underscore the central role of salient others in communicating high expectations, thereby elevating students' success expectancies and legitimising ambitious goals (Eccles & Wigfield, 2020; Merton, 1948; Rosenthal & Jacobson, 1968).

Consistent with this, students who report no clear teacher expectation have lower aspirations than those in the high-expectation group, yet higher than peers who perceive explicitly low expectations, indicating that unequivocally low signals are more inhibiting than ambiguity. This aligns with evidence that a warm classroom climate combined with challenging instruction fosters motivation and goal clarity (Rubie-Davies, 2014; De Boer et al., 2018) and accords with classroom findings from China showing that teacher feedback and affect jointly shape students' self-efficacy and future orientation (Ling et al., 2014; Pi, 2013).

Turning to the family context, parental educational expectations show the strongest relationship with aspirations among the observed covariates in the fully adjusted model. Students who perceive that their parents expect university level education report substantially higher aspirations than peers whose parents hold lower expectations, consistent with evidence on the role of family values in shaping students' future orientations (Kirk et al., 2011; Koshy et al., 2019; Wei & Ma, 2017). By contrast, parental involvement operationalised as routine school related support shows little additional relationship once expectations are accounted for. This distinction highlights that the frequency of involvement differs from the motivational quality of parental belief (Duan et al., 2018; Karbach et al., 2013; Xie & Zhao, 2022). In this context, higher expectations may provide purpose and direction that homework checking or attendance monitoring alone does not.

At the individual level, cognitive ability is statistically significantly related to aspirations, but the estimated magnitude is smaller than for perceived teacher and parental expectations. Although higher cognitive scores coincide with higher aspirations, the relationship is modest, suggesting that academic ability contributes to ambition yet does not outweigh the social and relational signals captured by adult expectations. In this

sample, aligned teacher and parental expectations are more closely linked to students' reported goals than cognitive skill alone (Liang et al., 2018; Liu et al., 2024; Zhou et al., 2016).

Demographically, girls report higher aspirations than boys. International evidence links this to more positive learning attitudes, greater engagement, fewer behavioural problems, and consistently higher grades (Del Pero & Bytchkova, 2013; Voyer & Voyer, 2014). Chinese studies likewise show that girls often voice higher ambitions and receive stronger encouragement from parents and teachers (Wei & Ma, 2017; Liu, 2020).

For ethnicity, minority students show slightly higher aspirations than Han peers after covariate adjustment. This challenges deficit views and aligns with work indicating that, once structural opportunities and supportive policies are considered, minority adolescents' aspirations are comparable to or higher than those of Han students (Hannum, 2002; Zhang & Ng, 2025; Wang, 2024).

Finally, using parental occupation as a proxy for socioeconomic status, aspirations differ by occupational status before accounting for family values. Once parental expectations are included, most occupational associations attenuate, with a persistent advantage for public-sector households (government officials, staff of public institutions, and civil servants). This pattern supports status-attainment accounts in which occupational position shapes children's goals chiefly through educational values and information channels transmitted within families (Sirin, 2005; Tan & Fang, 2023).

In sum, across school, family, and individual domains, two signals stand out: perceived teacher expectations and parental educational expectations are consistently linked to higher student aspirations.

5.2 Limitations

This study has several limitations that should be acknowledged.

First, the key independent variable teacher expectations are captured indirectly through students' perceptions. While this proxy offers insight into how students interpret their teachers' expectations, it introduces potential measurement bias. Perceptions may be shaped by students' self-concept, classroom interactions, or differential treatment, and may not accurately reflect teachers' actual beliefs or intentions (Rubie-Davies, 2006). Thus, the variable reflects students' subjective interpretations rather than an objective measure of teacher expectations.

Second, the study relies on a single wave of data, the 2013–2014 CEPS baseline, and focuses exclusively on Grade 9 students. This restriction, necessitated by the data structure (as teacher expectation perceptions are only available for this group), limits both the temporal scope and the generalisability of the findings across other age groups. Therefore, the cross-sectional nature of the data precludes strong causal inference (Morgan & Winship, 2015). Although the study includes theoretically informed controls and robustness checks (e.g., logistic regression), the analysis identifies associations rather than confirming causal directionality.

Third, aspirations are measured using a single survey item capturing the highest level of education students hope to attain. While this is common in aspiration research, it simplifies a multidimensional construct and does not distinguish between short-term and long-term goals (Khattab, 2015).

5.3 Future Research

Future research should address these limitations in several ways.

First, future research should adopt multi-informant designs incorporating teacher-reported expectations, peers' perspectives and classroom observations to strengthen the measurement of expectations and aspirations (Ready & Wright, 2011; Rubie-Davies, 2006, 2010).

Second, longitudinal data would enable researchers to track how teacher expectations and student aspirations co-evolve over time, thereby facilitating stronger causal inferences. In addition, a more nuanced measurement approach should be adopted, one that utilises multiple items or scales to distinguish between different types and temporal dimensions of educational aspirations. This would provide a richer understanding of how teacher expectations shape both short-term and long-term goals, such as aspirations for high school versus university (Wang et al., 2018).

Third, understanding not only the effects but also the antecedents of teacher expectations are essential for promoting equitable educational environments. Investigating these antecedents involves examining the psychological, cultural and experiential factors that influence how teachers come to form expectations about their students. For instance, teachers' personality traits, prior teaching experiences, cultural beliefs about intelligence and effort, and implicit biases may all shape their interpretive lens (Rubie-Davies, 2006). Gaining a deeper understanding of these formative processes is crucial for designing teacher education and

professional development programs that reduce bias, challenge stratified perceptions, and promote more inclusive and supportive classroom practices.

Lastly, future research should further explore how teacher expectations influence students' educational aspirations, with particular attention to the underlying mechanisms. For example, high teacher expectations may strengthen student motivation, academic self-concept and perceived competence, thereby enhancing their perceived capacity to pursue ambitious educational goals (Hornstra et al., 2018). Beyond self-beliefs, relational and emotional factors may also mediate the link between teacher expectations and student aspirations (Korpershoek et al., 2020). When students feel valued and respected by their teachers, they are more likely to imagine broader future possibilities and invest in long-term educational plans (Roorda et al., 2011). Unpacking these mediating processes is essential not only for advancing theoretical understanding but also for informing teacher training programs aimed at fostering equitable and aspirational learning environments.

Chapter 6: Conclusion

From a policy perspective, these findings have several implications. First, teacher training programs should include components that help educators understand the powerful influence of their expectations. Raising awareness of how verbal and non-verbal cues communicate academic potential may help mitigate unintentional discouragement. Schools might also implement systems to monitor student perceptions of teacher support and identify those who feel overlooked or underestimated.

Second, schools should engage parents more directly in expectation-setting conversations. Workshops, information sessions, or home-school communication tools that help parents formulate and express high educational hopes may prove more effective than traditional parent involvement strategies. This could be especially useful for students from low-SES or migrant backgrounds, who may otherwise lack exposure to aspirational role models.

Finally, future research should prioritise longitudinal designs that can better assess causality and explore how early signals from teachers and parents shape longer-term academic trajectories. Mixed-methods studies incorporating interviews, classroom observations, and teacher self-reports would help triangulate the mechanisms through which expectations are formed and internalised. Additionally, studies examining variations across urban/rural contexts or provincial education systems could uncover important structural moderators of these dynamics.

In conclusion, this study indicates that perceived teacher expectations and parental educational expectations are robust correlates of adolescents' educational aspirations in China. Although cognitive ability, gender, and socioeconomic background also matter, adults' expectation signals are more consistently linked to students' educational goals. Strengthening the clarity and positivity of expectation messages at school and at home therefore represents a practical lever for supporting ambitious yet realistic educational plans. However, the cross-sectional design limits causal interpretation, so these results should be interpreted as associations rather than causal effects.

References

Agresti, A. (2010). *Analysis of ordinal categorical data*. John Wiley & Sons.

Ajzen, I. (1991). The theory of planned behaviour. *Organizational behaviour and human decision processes*, *50*(2), 179-211.

Allison, P. D. (1999). *Multiple regression: A primer*. Pine Forge Press.

An, X. (2005). 教育期望、社会资本与贫困地区教育发展 [Educational aspiration, social capital, and educational development in disadvantaged regions]. *教育与经济 (Education and Economy)*, (04), 31–35. https://doi.org/10.3969/j.issn.1003-4870.2005.04.006

Archambault, I., Janosz, M., & Chouinard, R. (2012). Teacher Beliefs as Predictors of Adolescents' Cognitive Engagement and Achievement in Mathematics. *The Journal of Educational Research*, *105*(5), 319–328. https://doi.org/10.1080/00220671.2011.629694

Babad, E. Y., Inbar, J., & Rosenthal, R. (1982). Pygmalion, Galatea, and the Golem: Investigations of biased and unbiased teachers. *Journal of educational psychology*, *74*(4), 459.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of school psychology*, *48*(1), 5-37.

Barry, C. J. S., Davies, N. M., & Morris, T. T. (2022). Investigating how the accuracy of teacher expectations of pupil performance relates to socioeconomic and genetic factors. *Scientific Reports*, *12*(1), 7120.

Bartlett, M. S. (1954). A note on the multiplying factors for various χ 2 approximations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 296-298.

Benner, A. D., Fernandez, C. C., Hou, Y., & Gonzalez, C. S. (2021). Parent and teacher educational expectations and adolescents' academic performance: Mechanisms of influence. *Journal of community psychology*, *49*(7), 2679-2703.

Berger, R. (2015). Now I see it, now I don't: Researcher's position and reflexivity in qualitative research. *Qualitative Research*, *15*(2), 219–234. https://doi.org/10.1177/1468794112468475

Berrington, A., Roberts, S., & Tammes, P. (2016). Educational aspirations among UK Young Teenagers: Exploring the role of gender, class and ethnicity. *British Educational Research Journal*, *42*(5), 729–755. https://doi.org/10.1002/berj.3235

Bognar, C. J. (1983). Teacher expectations and student characteristics. *Canadian Journal of Education/Revue canadienne de l'education*, 47-56.

Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of educational psychology*, *75*(5), 631.

Brophy, J. E., & Good, T. L. (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of educational psychology*, *61*(5), 365.

Brophy, J. E., & Good, T. L. (1974). *Teacher-student relationships: Causes and consequences*. Holt, Rinehart and Winston.

Bryman, A. (2016). *Social research methods*. Oxford university press.

Buchmann, C., DiPrete, T. A., & McDaniel, A. (2008). Gender inequalities in education. *Annu. Rev. Sociol*, *34*(1), 319-337.

Campbell, R. T., & Parker, R. N. (1983). Substantive and statistical considerations in the interpretation of multiple measures of SES. *Social Forces*, *62*(2), 450-466.

Caplin, M. D. (1968). Self-concept, level of aspiration, and academic achievement. *The Journal of Negro Education*, *37*(4), 435-439.

Carpenter, J. R., & Smuk, M. (2021). Missing data: A statistical framework for practice. *Biometrical Journal*, *63*(5), 915-947.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies* (No. 1). Cambridge university press.

Castro, M., Expósito-Casas, E., López-Martín, E., Lizasoain, L., Navarro-Asencio, E., & Gaviria, J. L. (2015). Parental involvement on student academic achievement: A meta-analysis. *Educational research review*, *14*, 33-46.

Chen, L., Li, T., King, R. B., Du, H., Wu, K., & Chi, P. (2022). Gender Inequality lowers educational aspiration for adolescent boys and girls: a multi-level and longitudinal study in China. *Sex Roles*, *86*(5), 320-333.

Chen, X., Allen, J. L., & Hesketh, T. (2023). The influence of individual, peer, and family factors on the educational aspirations of adolescents in rural China. *Social Psychology of Education*, *26*(3), 735-759.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. routledge.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Cohen, L., Manion, L., & Morrison, K. (2002). *Research methods in education*. routledge.

Chia, T., & Hruschka, D. (2023). Educational & income disparities among ethnic minorities of China. International Journal of Educational Development, 102, 102846.

Cooper, H. M. (1979). Pygmalion grows up: A model for teacher expectation communication and performance influence. *Review of Educational research*, *49*(3), 389-410.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, *10*(1).

Crotty, M. J. (1998). The foundations of social research: Meaning and perspective in the research process.

Danişman, Ş. (2017). The effect of expectation on student achievement. In *The factors effecting student achievement: Meta-analysis of empirical studies* (pp. 227-245). Cham: Springer International Publishing.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*(1), 13-21.

De Boer, H., Bosker, R. J., & Van Der Werf, M. P. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology*, *102*(1), 168.

De Boer, H., Timmermans, A. C., & Van Der Werf, M. P. (2018). The effects of teacher expectation interventions on teachers' expectations and student achievement: Narrative review and meta-analysis. *Educational Research and Evaluation*, *24*(3-5), 180-200.

Del Pero, A. S., & Bytchkova, A. (2013). A bird's eye view of gender differences in education in OECD countries.

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical assessment, research, and evaluation*, *14*(1).

Duan, W., Guan, Y., & Bu, H. (2018). The effect of parental involvement and socioeconomic status on junior school students' academic achievement and school behavior in China. *Frontiers in psychology*, *9*, 952.

Duckworth, A. L., & Seligman, M. E. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological science*, *16*(12), 939-944.

Dupriez, V., Monseur, C., Van Campenhoudt, M., & Lafontaine, D. (2012). Social Inequalities of Post-Secondary Educational Aspirations: Influence of Social Background, School Composition and Institutional Context. *European Educational Research Journal*, *11*(4), 504–519. https://doi.org/10.2304/eerj.2012.11.4.504

Eccles, J. S. (2005). Influences of parents' education on their children's educational attainments: The role of parent and child perceptions. *London review of education*, *3*(3).

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

El Nokali, N. E., Bachman, H. J., & Votruba-Drzal, E. (2010). Parent involvement and children's academic and social development in elementary school. *Child development*, *81*(3), 988-1005.

England, K. V. (1994). Getting personal: Reflexivity, positionality, and feminist research. *The professional geographer*, *46*(1), 80-89.

Englund, M. M., Luckner, A. E., Whaley, G. J. L., & Egeland, B. (2004). Children's Achievement in Early Elementary School: Longitudinal Effects of Parental Involvement, Expectations, and Quality of Assistance. *Journal of Educational Psychology*, *96*(4), 723–730. https://doi.org/10.1037/0022-0663.96.4.723

Erdem, C., & Kaya, M. (2020). A Meta-Analysis of the Effect of Parental Involvement on Students' Academic Achievement. *Journal of Learning for Development*, *7*(3), 367-383.

Erikson, R. (1984). Social class of men, women and families. *Sociology*, *18*(4), 500-514.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods*, *4*(3), 272.

Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational psychology review*, *13*(1), 1-22.

Fang, G., & Hou, Y. (2019). 家庭社会经济地位如何影响初中生认知能力的发展 [How does family socioeconomic status influence the development of junior high school students' cognitive ability?]. *全球教育展望 (Global Education Outlook)*, (9), 68–76. https://doi.org/CNKI:SUN:WGJN.0.2019-09-006

Field, A. (2024). *Discovering statistics using IBM SPSS statistics* (6th edition). Sage.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological assessment*, *7*(3), 286.

Friedrich, A., Flunger, B., Nagengast, B., Jonkmann, K., & Trautwein, U. (2015). Pygmalion effects in the classroom: Teacher expectancy effects on students' math achievement. *Contemporary Educational Psychology*, *41*, 1-12.

Gamoran, A., & Long, D. A. (2007). Equality of educational opportunity a 40-year retrospective. In *International studies in educational inequality, theory and policy* (pp. 23-47). Dordrecht: Springer Netherlands.

Gan, Y., & Bilige, S. (2019). Parental involvement in home-based education and children's academic achievement in China. *Social Behavior and Personality: an international journal*, *47*(12), 1-15.

Ganzach, Y. (2000). Parents' education, cognitive ability, educational expectations and educational attainment: Interactive effects. *British Journal of Educational Psychology*, *70*(3), 419–441. https://doi.org/10.1348/000709900158218

Ganzeboom, H. B., & Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Social science research*, *25*(3), 201-239.

Gentrup, S., Lorenz, G., Kristen, C., & Kogan, I. (2020). Self-fulfilling prophecies in the classroom: Teacher expectations, teacher feedback and student achievement. *Learning and Instruction*, *66*, 101296.

Gil-Hernández, C. J., & Gracia, P. (2018). Adolescents' educational aspirations and ethnic background: The case of students of African and Latin American migrant origins in Spain. *Demographic Research*, *38*, 577-618.

Good, T. L., Sterzinger, N., & Lavigne, A. (2018). Expectation effects: Pygmalion and the initial 20 years of research. *Educational Research and Evaluation*, *24*(3-5), 99-123.

Hadjar, A., Scharf, J., & Hascher, T. (2021). Who aspires to higher education? Axes of inequality, values of education and higher education aspirations in secondary schools in Luxembourg and the Swiss Canton of Bern. *European Journal of Education*, *56*(1), 9–26. https://doi.org/10.1111/ejed.12435

Hahs-Vaughn, D. L., & Lomax, R. (2020). *An introduction to statistical concepts*. Routledge.

Hair, J. F. (2009). Multivariate data analysis.

Haller, A. O., & Woelfel, J. (1972). *Significant others and their expectations: Concepts and instruments to measure interpersonal influence on status aspirations*. Institute for Research on Poverty, University of Wisconsin.

Hannum, E. (2002). Educational stratification by ethnicity in China: Enrollment and attainment in the early reform years. *Demography*, *39*(1), 95-117.

Hannum, E., & Park, A. (Eds.). (2007). *Education and reform in China*. London: Routledge.

Hanson, S. L. (1994). Lost talent: Unrealized educational aspirations and expectations among US youths. *Sociology of education*, 159-183.

Hanushek, E. A., & W ößmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal*, *116*(510), C63-C76.

Harris, M. J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological bulletin*, *97*(3), 363.

Hauser, R. M., & Warren, J. R. (1997). 4. Socioeconomic Indexes for Occupations: A Review, Update, and Critique. *Sociological Methodology*, *27*(1), 177–298. https://doi.org/10.1111/1467-9531.271028

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological measurement*, *66*(3), 393-416.

Hill, N. E., & Tyson, D. F. (2009). Parental involvement in middle school: a meta-analytic assessment of the strategies that promote achievement. *Developmental psychology*, *45*(3), 740.

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of educational research*, *59*(3), 297-313.

Hopper, E. I., Coxon, A. P. M., Blau, P. M., & Duncan, O. D. (1968). The American Occupational Structure. *The British Journal of Sociology*, *19*(4), 453. https://doi.org/10.2307/588184

Hornstra, L., Stroet, K., van Eijden, E., Goudsblom, J., & Roskamp, C. (2018). Teacher expectation effects on need-supportive teaching, student motivation, and engagement: A self-determination perspective. *Educational Research and Evaluation*, *24*(3-5), 324-345.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.

Hu, Y. P. (2004). 新疆回族教育观念的调查与分析 [An investigation and analysis of Hui people's educational attitudes in Xinjiang]. *回族研究 (The Hui Studies)*, (3), 65–69. https://doi.org/10.16023/j.cnki.cn64-1016/c.2004.03.010

Huang, C., & Wu, Y. X. (2016). 中学生教育期望的性别差异：表现与成因 [Gender differences in educational expectations among secondary school students: Manifestations and causes]. *江苏社会科学 (Jiangsu Social Sciences)*, (04), 121–132. https://doi.org/10.13858/j.cnki.cn32-1312/c.2016.04.017

Huang, L. (2016). 家长参与学校教育对初中学生认知能力表现影响的实证研究—基于中国教育追踪调查基线数据的分析 [An empirical study on the impact of parental involvement in school education on junior high school students' cognitive performance: Based on CEPS baseline data]. *教育科学研究* (*Educational Research*), (12), 53–59. https://doi.org/CNKI:SUN:JYKY.0.2016-12-012

Hynds, A., Averill, R., Hindle, R., & Meyer, L. (2017). School expectations and student aspirations: The influence of schools and teachers on Indigenous secondary students. *Ethnicities*, *17*(4), 546-573.

Jeynes, W. H. (2007). The relationship between parental involvement and urban secondary school student academic achievement: A meta-analysis. *Urban Education, 42*(1), 82–110. https://doi.org/10.1177/0042085906293818

Jung, E., & Zhang, Y. (2016). Parental involvement, children's aspirations, and achievement in new immigrant families. *The Journal of Educational Research*, *109*(4), 333–350. https://doi.org/10.1080/00220671.2014.959112

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and social psychology review*, *9*(2), 131-155.

Kao, G., & Tienda, M. (1998). Educational aspirations of minority youth. *American journal of education*, *106*(3), 349-384.

Kaiser, H. F. (1974). An index of factorial simplicity. *psychometrika*, *39*(1), 31-36.

Karbach, J., Gottschling, J., Spengler, M., Hegewald, K., & Spinath, F. M. (2013). Parental involvement and general cognitive ability as predictors of domain-specific academic achievement in early adolescence. *Learning and Instruction*, *23*, 43–51. https://doi.org/10.1016/j.learninstruc.2012.09.004

Khattab, N. (2015). Students' aspirations, expectations and school achievement: What really matters?. *British educational research journal*, *41*(5), 731-748.

Khattab, N. (2018). Ethnicity and higher education: The role of aspirations, expectations and beliefs in overcoming disadvantage. *Ethnicities*, *18*(4), 457-470.

Kirk, C. M., Lewis-Moss, R. K., Nilsen, C., & Colvin, D. Q. (2011). The role of parent expectations on adolescent educational aspirations. *Educational Studies*, *37*(1), 89–99. https://doi.org/10.1080/03055691003728965

Korpershoek, H., Canrinus, E. T., Fokkens-Bruinsma, M., & De Boer, H. (2020). The relationships between school belonging and students' motivational, social-emotional, behavioural, and academic outcomes in secondary education: A meta-analytic review. *Research papers in education*, *35*(6), 641-680.

Koshy, P., Dockery, A. M., & Seymour, R. (2019). Parental expectations for young people's participation in higher education in Australia. *Studies in Higher Education*, *44*(2), 302–317. https://doi.org/10.1080/03075079.2017.1363730

Krosnick, J. a., & Presser, S.(2010). Question and questionnaire design. *Handbook of survey research*, *94305*(886), 1432-1033.

Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science*, *19*(6), 339-345.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*, *4*, 863.

Lazarides, R., Viljaranta, J., Aunola, K., Pesu, L., & Nurmi, J.-E. (2016). The role of parental expectations and students' motivational profiles for educational aspirations. *Learning and Individual Differences*, *51*, 29–36. https://doi.org/10.1016/j.lindif.2016.08.024

Li, H., Loyalka, P., Rozelle, S., Wu, B., & Xie, J. (2015). Unequal access to college in China: How far have poor, rural students been left behind?. *The China Quarterly*, *221*, 185-207.

Li, H., Loyalka, P., Rozelle, S., & Wu, B. (2017). Human Capital and China's Future Growth. *Journal of Economic Perspectives*, *31*(1), 25–48. https://doi.org/10.1257/jep.31.1.25

Liang, W., Ye, X., & Li, T. (2018). 父母参与如何影响流动儿童认知能力—基于 CEPS 基线数据的实证研究 [How parental involvement affects the cognitive ability of migrant children: An empirical study based on CEPS baseline data]. *教育学报 (Journal of Educational Studies)*, *14*(1), 80–94. https://doi.org/10.14082/j.cnki.1673-1298.2018.01.011

Ling, H., Lu, A., Feng, Y., & Li, H. (2014). 教师期望、自我效能感与成绩：基于小学语文学习的研究 [Teacher expectations, student self-efficacy and achievement: A study based on primary Chinese education]. *湖北函授大学学报 (Journal of Hubei Correspondence University), 27*, 115–116.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American statistical Association*, *83*(404), 1198-1202.

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons.

Liu, J. (2012). 少数民族贫困地区基础教育发展的思考—以云南金平县者米乡为例[Rethinking basic education development in ethnic minority poverty areas: A case study of Zhemi Township, Jinping County, Yunnan]. *红河学院学报 (Journal of Honghe University)*, (02), 109–111. https://doi.org/10.13963/j.cnki.hhuxb.2012.02.052

Liu, J., Cheng, M., & Wu, C. (2024). 父母受教育水平与子女认知能力——来自中国教育追踪调查的经验证据 [Parental educational attainment and children's cognitive ability: Empirical evidence from the China Education Panel Survey]. *教育与经济* (*Education & Economy*), *(4)*, 56–68. https://doi.org/CNKI:SUN:JYJI.0.2024-04-006

Liu, W. (2020). 中学生高等教育期望的性别差异及其成因 [Gender differences in secondary school students' higher education expectations and their causes]. *当代青年研究 (Contemporary Youth Research)*, (03), 78–84. https://doi.org/CNKI:SUN:QING.0.2020-03-012

Liu, Y. (2016). *Higher education, meritocracy and inequality in China*. Springer.

Long, J. S., & Freese, J. (2006). *Regression models for categorical dependent variables using Stata* (Vol. 7). Stata press.

Lundberg, S. (2020). Educational gender gaps. *Southern economic journal*, *87*(2), 416-439.

Ma, X. & Wei, Y. (2017). 家长教育期望中的"罗森塔尔效应"循环模型探析—基于 CEPS 的模型检验 [A cyclical model of the Rosenthal effect in parental educational expectations: An empirical test based on CEPS data]. *新疆社会科学* (*Xinjiang Social Sciences*), (01), 135–140. https://doi.org/CNKI:SUN:XJSH.0.2017-01-023

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention science*, *1*(4), 173-181.

Madarasova Geckova, A., Tavel, P., Van Dijk, J. P., Abel, T., & Reijneveld, S. A. (2010). Factors associated with educational aspirations among adolescents: Cues to counteract socioeconomic differences? *BMC Public Health*, *10*(1), 154.

Marginson, S. (2016). High participation systems of higher education. *The Journal of Higher Education*, *87*(2), 243-271.

Marks, G. N. (2014). *Education, social background and cognitive ability: The decline of the social*. Routledge.

Marsh, H. W., Pekrun, R., Guo, J., Hattie, J., & Karin, E. (2023). Too much of a good thing might be bad: The double-edged sword of parental aspirations and the adverse effects of aspiration-expectation gaps. *Educational Psychology Review*, *35*(2), 49.

Menard, S. W. (2010). *Logistic regression: From introductory to advanced concepts and applications*. Sage.

Meng, X., Qi, S., & Kan, B. (1998). 青海藏族的教育需求—对两个藏族社区的调查 [Educational demands among Tibetan communities in Qinghai: A study of two villages]. *中国社会科学 (Social Sciences in China)*, (03). https://doi.org/10.16063/j.cnki.cn11-1083/c.1998.03.010

Merton, R. K. (1948). The Self-Fulfilling Prophecy. *The Antioch Review*, *8*(2), 193. https://doi.org/10.2307/4609267

Millsap, R. E., & Maydeu-Olivares, A. (2009). *The SAGE handbook of quantitative methods in psychology*. SAGE.

Mistry, R. S., White, E. S., Benner, A. D., & Huynh, V. W. (2009). A longitudinal study of the simultaneous influence of mothers' and teachers' educational expectations on low-income youth's academic achievement. *Journal of youth and adolescence*, *38*(6), 826-838.

Moon, S. M. (1996). Beyond the classroom: Why school reform has failed and what parents need to do. *Gifted and Talented International*, *11*(2), 89-90.

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: Methods and principles for social research* (Second edition). Cambridge University press.

Mundt, S. D., & Mundt, M. P. (2020). The role of peer groups in adolescents' educational expectations: a stochastic actor-based model. *International Journal of Adolescence and Youth*, *25*(1), 1009-1021.

Neuenschwander, M. P., Vida, M., Garrett, J. L., & Eccles, J. S. (2007). Parents' expectations and students' achievement in two western nations. *International Journal of Behavioral Development*, *31*(6), 594-602.

Nitardy, C. M., Duke, N. N., Pettingell, S. L., & Borowsky, I. W. (2015). Racial and Ethnic Disparities in Educational Achievement and Aspirations: Findings from a Statewide Survey from 1998 to 2010. *Maternal and Child Health Journal.*, *19*(1), 58–66. https://doi.org/10.1007/s10995-014-1495-y

Ojeda, L., & Flores, L. Y. (2008). The influence of gender, generation level, parents' education level, and perceived barriers on the educational aspirations of Mexican American high school students. *The Career Development Quarterly*, *57*(1), 84-95.

Ones, D. S., Dilchert, S., & Viswesvaran, C. (2012). Cognitive abilities. *The Oxford handbook of personnel assessment and selection*, 179-224.

Osborne, J. W., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical assessment, research, and evaluation*, *8*(1).

Otani, M. (2019). Relationships between parental involvement and adolescents' academic achievement and aspiration. *International Journal of Educational Research*, *94*, 168–182. https://doi.org/10.1016/j.ijer.2019.01.005

Pan, L. (2012). 白族家庭教育价值观考察 [An investigation of Bai ethnic families' educational values]. *四川民族学院学报 (Journal of Sichuan Nationalities College)*, (01), 85–91. https://doi.org/10.13934/j.cnki.cn51-1729/g4.2012.01.023

Pang, W., Xu, X., Lin, L., & Ren, Y. (2013). 家庭社会经济地位与中学生学业成绩的关系研究 [The relationship between family socioeconomic status and academic performance among junior middle school students]. *全球教育展望 (Global Education Outlook)*, (02), 12–21. https://doi.org/10.3969/j.issn.1672-6218.2013.02.005

Pawson, R. (2006). Evidence-based policy: a realist perspective.

Pi, W. (2013). 英语教师期望对学生英语学习成绩的影响机制研究—基于英语学习自我效能感的中介效应分析 [The mechanism of the effects of teacher's expectations on university students' English learning performance: A research based on English learning self-efficacy]. *浙江外国语学院学报 (Journal of Zhejiang International Studies University)*, 37–45.

Plenty, S. M., & Jonsson, J. O. (2021). Students' occupational aspirations: Can family relationships account for differences between immigrant and socioeconomic groups?. *Child development*, *92*(1), 157-173.

Qian, Z., & Blair, S. L. (1999). Racial/ethnic differences in educational aspirations of high school seniors. *Sociological perspectives*, *42*(4), 605-625.

Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal*, *48*(2), 335-360.

Ree, M. J., Carretta, T. R., & Steindl, J. R. (2001). Cognitive ability. *Handbook of industrial, work and organizational psychology*, *1*, 219-232.

Richardson, J. T., Mittelmeier, J., & Rienties, B. (2020). The role of gender, social class and ethnicity in participation and academic attainment in UK higher education: an update. *Oxford Review of Education*, *46*(3), 346-362.

Roorda, D. L., Koomen, H. M., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement: A meta-analytic approach. Review of educational research, 81(4), 493-529.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, *3*(1), 16–20. https://doi.org/10.1007/bf02322211

Rosenthal, R. (1973). The mediation of Pygmalion effects: A four factor "theory". *Papua New Guinea Journal of Education*, *9*(1), 1-12.

Rubie-Davies, C. M. (2006). Teacher expectations and student self-perceptions: Exploring relationships. *Psychology in the Schools*, *43*(5), 537-552.

Rubie-Davies, C. M. (2010). Teacher expectations and perceptions of student attributes: Is there a relationship? *British Journal of Educational Psychology*, *80*(1), 121–135. https://doi.org/10.1348/000709909x466334

Rubie-Davies, C. M. (2014). *Becoming a high expectation teacher: Raising the bar*. Routledge.

Schoon, I., & Parsons, S. (2002). Teenage aspirations for future careers and occupational outcomes. *Journal of Vocational Behavior*, *60*(2), 262-288.

Schoon, I., & Polek, E. (2011). Teenage career aspirations and adult career attainment: The role of gender, social background and general cognitive ability. *International Journal of Behavioral Development*, *35*(3), 210–217. https://doi.org/10.1177/0165025411398183

Sewell, W. H., Haller, A. O., & Portes, A. (1969). The Educational and Early Occupational Attainment Process. *American Sociological Review*, *34*(1), 82. https://doi.org/10.2307/2092789

Sewell, W. H., & Hauser, R. M. (1993). A review of the Wisconsin Longitudinal Study of social and psychological factors in aspirations and achievements 1963-1992. *Center for Demography and Ecology, Working Papers*, *92*(01).

Sheng, X. (2014). Parental expectations relating to children's higher education in urban China: Cultural capital and social class. *Journal of Sociology*, *50*(4), 560-576.

Shi, W. (2013). Issues and problems in the current development of vocational education in China. *Chinese Education & Society*, *46*(4), 12-21.

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, *75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational researcher*, *31*(7), 15-21.

Sternberg, R. J. (2011). Individual differences in cognitive development.

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics, 6th edn Boston. *Ma: Pearson.*

Tan, G. L. C., & Fang, Z. (2023). Family social and cultural capital: an analysis of effects on adolescents' educational outcomes in China. *The Journal of Chinese Sociology*, *10*(1), 21.

Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of educational psychology*, *99*(2), 253.

Tian, J. (2022). 论民族地区留守儿童教育理想期望与实质期望的差异—以湖北省恩施州为例 [Differences between ideal and actual educational expectations among left-behind children in ethnic regions: A case study of Enshi Prefecture, Hubei Province]. *西部学刊 (Western Journal)*, (17), 135–138. https://doi.org/10.16721/j.cnki.cn61-1487/c.2022.17.029

Turcios-Cotto, V. Y., & Milan, S. (2013). Racial/Ethnic Differences in the Educational Expectations of Adolescents: Does Pursuing Higher Education Mean Something Different to Latino Students Compared to White and Black Students? *Journal of Youth and Adolescence*, *42*(9), 1399–1412. https://doi.org/10.1007/s10964-012-9845-9

Turner, H., Rubie-Davies, C. M., & Webber, M. (2015). Teacher Expectations, Ethnicity and the Achievement Gap. *New Zealand Journal of Educational Studies*, *50*(1), 55–69. https://doi.org/10.1007/s40841-015-0004-1

Tyler, K. M., & Boelter, C. M. (2008). Linking black middle school students' perceptions of teachers' expectations to academic engagement and efficacy. *Negro Educational Review*, *59*(1/2), 27.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, *140*(4), 1174–1204. https://doi.org/10.1037/a0036620

Walkey, F. H., McClure, J., Meyer, L. H., & Weir, K. F. (2013). Low expectations equal no expectations: Aspirations, motivation, and achievement in secondary school. *Contemporary educational psychology*, *38*(4), 306-315.

Wang, L., & Lehtomaki, E. (2024). Explaining persistent disparities in educational achievement: a comparative study of Yi minority and Han majority in China. *Educational Research and Evaluation*, *29*(1-2), 27-49.

Wang, M. T., & Eccles, J. S. (2013). School context, achievement motivation, and academic engagement: A longitudinal study of school engagement using a multidimensional perspective. *Learning and instruction*, *28*, 12-23.

Wang, S., Meissel, K., & Rubie-Davies, C. M. (2021). Teacher expectation effects in Chinese junior high schools: Exploring links between teacher expectations and student achievement using a hierarchical linear modelling approach. *Social Psychology of Education*, *24*(5), 1305–1333. https://doi.org/10.1007/s11218-021-09654-7

Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation*, *24*(3-5), 124-179.

Wang, Y., Qin, Y., Gao, K., Qi, W., & Sang, G. (2025). Perceived teacher expectations and the academic engagement of junior high school students in rural China: The role of academic self-efficacy. *School Psychology International*, *46*(1), 51-67.

Wei, Y., & Ma, X. (2017). 中学生自我教育期望的影响因素研究：基于 CEPS 的实证分析 [A study on factors influencing secondary school students' self-educational aspirations: An empirical analysis based on CEPS]. *教育学术月刊 (Educational Academic Monthly)*, (10), 69–78.

Weinstein, R. S. (2002). *Reaching higher: The power of expectations in schooling*. Harvard University Press.

Wen, Q. (2022). Trends in college–high school wage differentials in China: The role of cohort-specific labor supply shift. *Sustainability*, *14*(24), 16917.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wu, X., & Ye, Y. (2018). *Technical and vocational education in China* (pp. 175-204). Springer Singapore.

Xie, R. J. & Zhao, Y. (2022). 家长参与对子女学业成绩的影响研究—基于 CEPS2013—2014 数据 [Research on the effect of parental involvement on children's academic achievement: Based on CEPS 2013–2014 data]. *林区教学* (*Teaching of Forestry Region*), (01), 109–112. https://doi.org/CNKI:SUN:LQXJ.0.2022-01-027

Xu, X., & Zhang, Q. (2015). 小学生的教师期望、数学作业情绪与数学成绩的关系研究 [A study on the relations between primary students' teacher expectations, mathematics assignment emotion, and mathematics achievement]. *现代中小学教育 (Modern Primary and Secondary Education)*, *31*, 52–54.

Yang, F., Qian, Y., & Xia, Z. (2024). Cognitive ability and locus of control: The effect of parental involvement on the academic performance of elementary and secondary school students. *Current Psychology*, *43*(3), 2816–2831. https://doi.org/10.1007/s12144-023-04572-7

Yang, X., Yao, Y., & Zhang, S. (2016). 家庭社会地位对青少年教育期望影响研究—基于 CEPS 2014 调查数据的实证分析 [The influence of family socioeconomic status on adolescents' educational aspirations: An empirical study based on CEPS 2014 data]. *中国青年研究 (China Youth Study)*, (07), 67-73. https://doi.org/10.19633/j.cnki.11-2579/d.2016.07.011

Yiu, L., & Adams, J. (2013). Reforming Rural Education in China: Understanding Teacher Expectations for Rural Youth. *The China Quarterly*, *216*, 993–1017. https://doi.org/10.1017/s0305741013001136

Yu, X. (2020). 父母社会背景、教育价值观及其教育期望 [Parental social background, educational values, and educational expectations]. *南京师大学报（社会科学版） [Journal of Nanjing Normal University (Social Science Edition)]*, (04), 62-74. https://doi.org/10.3969/j.issn.1001-4603.2020.04.005

Zhang, D., Li, X., & Xue, J. (2015). Education inequality between rural and urban areas of the People's Republic of China, migrants' children education, and some implications. *Asian development review*, *32*(1), 196-224.

Zhang, M. R., & Ng, F. F. Y. (2025). Ethnic and Gender Variations in Adolescents' Future Aspirations. *Journal of Adolescence*.

Zhao, M., & Glewwe, P. (2010). What determines basic school attainment in developing countries? Evidence from rural China. *Economics of Education Review*, *29*(3), 451–460. https://doi.org/10.1016/j.econedurev.2009.10.008

Zhou, F., & Cheng, T. (2016). 中学生教育期望的性别差异：父母教育卷入的影响效果分析 [Gender differences in educational aspirations among middle school students: An analysis of the effect of parental involvement]. *教育研究与实验 (Educational Research and Experiment)*, (6), 7–16.

Zhou, M., Sun, X., Kang, Z., & Huang, L. (2016). 父母教育期望对农村儿童认知能力的影响——基于 CFPS 面板数据的实证研究 [The impact of parental educational expectations on rural children's cognitive ability: An empirical study based on CFPS panel data]. *湖南农业大学学报(社会科学版) (Journal of Hunan Agricultural University: Social Sciences),* (4), 57–62, 90. https://doi.org/10.13331/j.cnki.jhau(ss).2019.04.008

Zumbo, B. D., & Zimmerman, D. W. (1993). Is the selection of statistical methods governed by level of measurement? *Canadian Psychology / Psychologie Canadienne*, *34*(4), 390–400. https://doi.org/10.1037/h0078865

Appendices

Appendix A

A.1 Comparison of Included and Excluded Samples on Key Study Variables

Variable	Included Sample	Excluded Sample	Test Statistic	p-value
% student aspiration bachelor's degree or above	65.0%	48.2%	χ^2 = 127.07	< .001
% Female	51.2%	42.5%	χ^2 = 30.49	< .001
% Ethnic Minority	8.0%	12.9%	χ^2 = 31.46	< .001
% Teacher Expectation: key high school & ordinary high school	74.4%	81.7%	χ^2 = 29.87	< .001
SES (mean)	5.49	6.01	t = -6.76	< .001
Cognitive Ability (mean)	9.28	7.72	t = 13.88	< .001
Parental Expectation (mean)	6.32	5.62	t = 11.07	< .001
Parental Involvement (mean)	-0.03	0.16	t = -6.71	< .001

Notes. χ^2 tests were used for categorical variables and t tests for continuous variables. See Section 3.3.1 for discussion of representativeness and limitations.

A.2 Distribution of Students by Ethnic Group

	Code	Label	Frequency	Percentage
1	1	The Han nationality	8398	91.20
2	2	The Mongol nationality	12	0.13
3	3	The Manchu nationality	97	1.05
4	4	The Hui nationality	60	0.65
5	5	The Tibetan nationality	7	0.08
6	6	The Zhuang nationality	12	0.13
7	8	Other	607	6.59
8	<NA>	Missing	15	0.16

A.3 Distribution of Parental Occupations

	Label	Frequency	Percentage
1	Government / civil servant	645	7.14
2	Manager	685	7.58
3	Professional (e.g., teacher/doctor)	572	6.33
4	Technical worker	1421	15.72
5	Production/Factory worker	895	9.90
6	Service industry worker	658	7.28
7	Self-employed	1602	17.72
8	Peasant	2160	23.90
9	Unemployed	187	2.07
10	Other	214	2.37
11	Missing	169	1.87

A.4 Parental involvement by using EFA (Exploratory Factor Analysis - principal axis factoring)

Including B22 (all 2 items); B23 (first 3 items); B24 (first item for both parents)

B22. How often did your parents do the following to check up on your study last week?

	Never	One or two days	Three or four days	Almost everyday
Checking up on your homework	1	2	3	4
Giving instruction on your homework	1	2	3	4

B23. Do your parents care and are they strict with you about the following?

	They don't care.	They do care about it, but are not strict.	They are very strict about it.
Your homework and examination	1	2	3
Your behavior at school	1	2	3
Attendances at school everyday	1	2	3
Time when you get home everyday	1	2	3
Whom you make friends with	1	2	3
Your dress style	1	2	3
Time you spend on the Internet	1	2	3
Time you spend on watching TV	1	2	3

B24. How often do your parents discuss the following with you?

	Your mother			Your father		
	Never	Sometimes	Often	Never	Sometimes	Often
Things happened at school	1	2	3	1	2	3
The relationship between you and your friends	1	2	3	1	2	3
The relationship between you and your teachers	1	2	3	1	2	3
Your feelings	1	2	3	1	2	3
Your worries and troubles	1	2	3	1	2	3

Figure A.4.1 Scree plot for parental involvement

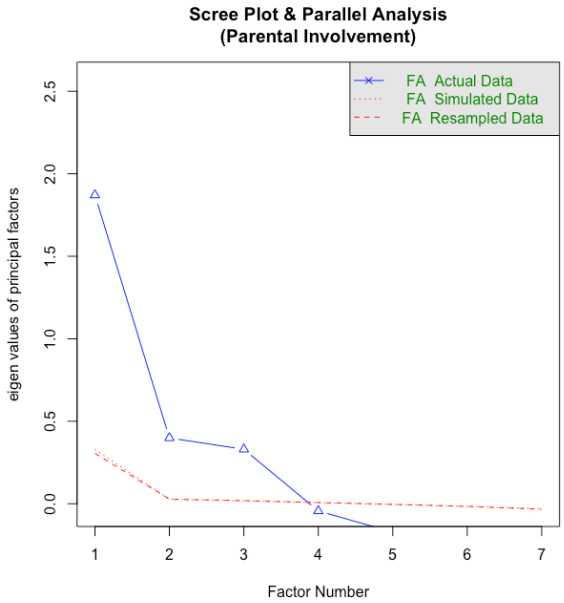


Figure A.4.2 Factor Structure of Parental Involvement from EFA

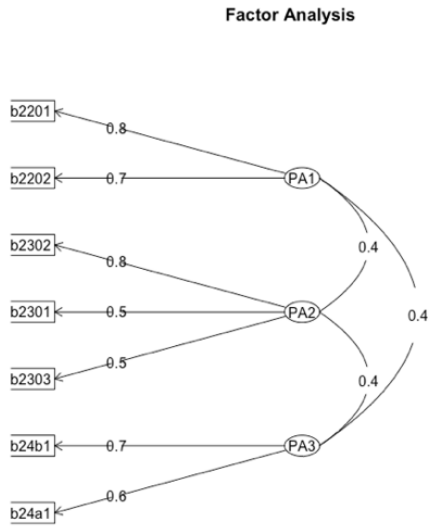


Table A.4.3 Factor Correlation Matrix for Parental Involvement Factors

```
 With factor correlations of
     PA1  PA2  PA3
PA1 1.00 0.45 0.38
PA2 0.45 1.00 0.36
PA3 0.38 0.36 1.00
```

Table A.4.4 Spearman's Rho Correlations among Parental Involvement Items

Variable	Label	Spearman_rho
b2201	Parent checked your homework	0.054
b2202	Parent gave instruction on your homework	0.074
b2301	Parents care about your homework/exam	0.156
b2302	Parents care about your behavior at school	0.090
b2303	Parents care about your school attendance	0.062
b24a1	Mother: Discussed school events with you	0.215
b24b1	Father: Discussed school events with you	0.130

Appendix B

relevant questions in the questionnaire:

dependant variable: student educational aspiration

C22. What is the highest level of education you expect yourself to receive?

1. Drop out now
2. Graduate from junior high school
3. Go to technical secondary school or technical school
4. Go to vocational high school
5. Go to senior high school
6. Graduate from junior college
7. Get a bachelor degree
8. Get a Master degree
9. Get a Doctor degree
10. I don't care

key independent variable: perception of teacher's expectation

C29. What kind of advice do you think your homeroom teacher will give to you about your education after junior high school?

1. Going to a key senior high school
2. Going to an ordinary senior high school
3. Going to a vocational high school
4. Going to a technical secondary school/technical school
5. Not going any further after graduating from junior high school
6. None

control variable:

gender

A1. Your sex is:

1. Male 2. Female

ethnicity

A3. Your ethnic nationality is:

1. The Han nationality
2. The Mongol nationality
3. The Manchu nationality
4. The Hui nationality
5. The Tibetan nationality
6. The Zhuang nationality
7. The Uygur nationality
8. Other (Please specify: _____)

social economic status – parental occupation

B8. What are the occupations of your parents?

	Mother	Father
Government official, staff of public institutions, civil servant	1	1
Middle/Senior management personnel of enterprises/corporations	2	2
Teacher, engineer, doctor, lawyer	3	3
Technical worker (including driver)	4	4
Ordinary staff or worker in production or manufacturing industry	5	5
Ordinary staff or worker in business or service industry	6	6
Self-employed worker	7	7
Peasant	8	8
Unemployed or laid-off worker	9	9
Other	(Please specify:_____)	(Please specify:_____)

parental involvement

B22. How often did your parents do the following to check up on your study last week?

	Never	One or two days	Three or four days	Almost everyday
Checking up on your homework	1	2	3	4
Giving instruction on your homework	1	2	3	4

B23. Do your parents care and are they strict with you about the following?

	They don't care.	They do care about it, but are not strict.	They are very strict about it.
Your homework and examination	1	2	3

Your behavior at school	1	2	3
Attendances at school everyday	1	2	3
Time when you get home everyday	1	2	3
Whom you make friends with	1	2	3
Your dress style	1	2	3
Time you spend on the Internet	1	2	3
Time you spend on watching TV	1	2	3

B24. How often do your parents discuss the following with you?

	Your mother			**Your father**		
	Never	**Sometimes**	**Often**	**Never**	**Sometimes**	**Often**
Things happened at school	1	2	3	1	2	3
The relationship between you and your friends	1	2	3	1	2	3
The relationship between you and your teachers	1	2	3	1	2	3
Your feelings	1	2	3	1	2	3
Your worries and troubles	1	2	3	1	2	3

parental expectation:

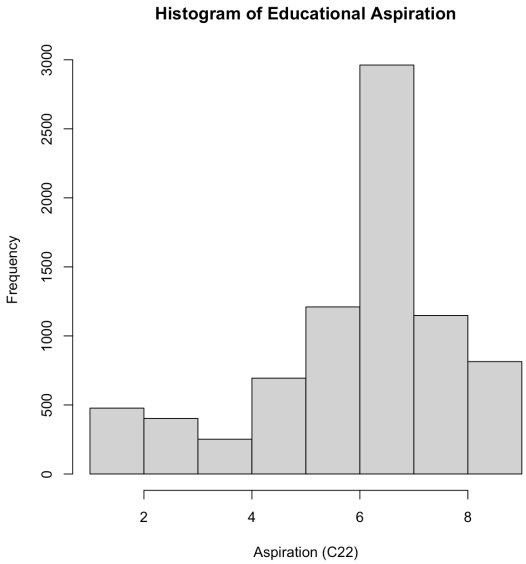
B31. What is the highest level of education your parents expect you to receive?

1. Drop out now
2. Graduate from junior high school
3. Go to technical secondary school or technical school
4. Go to vocational high school
5. Go to senior high school
6. Graduate from junior college
7. Get a bachelor degree
8. Get a Master degree
9. Get a Doctor degree
10. They don’t care

Appendix C: Assumptions Checking

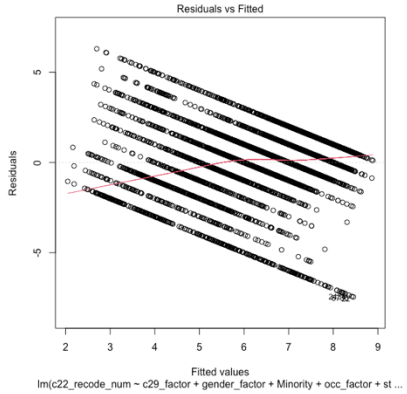
Testing Assumption for Sequential Linear Regression

In the variable construction section, it was ensured that the dependent variable was continuous. The independent variables were either continuous (e.g., cognitive ability, parental involvement, parental expectation) or binary (e.g., gender, ethnicity). Nominal variables such as student aspiration and socioeconomic status were dummy coded accordingly.

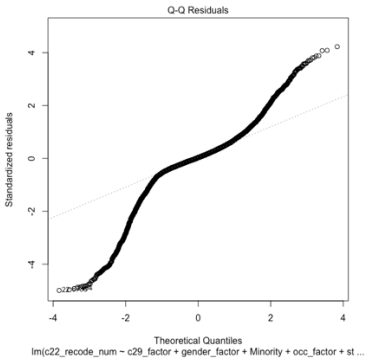


	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	7959	6.38	1.94	7	6.61	1.48	1	9	8	-1.14	1.03	0.02

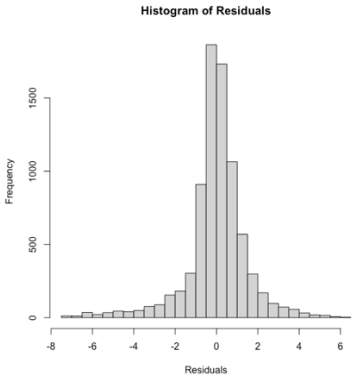
Although the histogram suggested a slight positive skew, both the skewness (1.03) and kurtosis (-1.14) were within acceptable thresholds (|skewness| < 2, |kurtosis| < 7), indicating no serious deviation from normality (Hair, 2009)



To assess the assumption of linearity in the regression model, a residual versus fitted values plot was inspected. In a well-fitting linear model, residuals should be randomly scattered around zero with no clear pattern (Field, 2024; Hair, 2019). In this case, the scatterplot did not show strong systematic deviations, although a slight curve was visible in the smoothing line, suggesting a minor departure from perfect linearity. However, the deviation was not substantial. In addition to linearity, this same residuals-versus-fitted plot was used to evaluate homoscedasticity, the assumption that the variance of residuals remains constant across levels of the fitted values (Tabachnick & Fidell, 2013).



To assess the normality of residuals, a Q–Q plot was examined. Most points closely followed the 45-degree reference line, particularly in the central range, suggesting that the residuals were approximately normally distributed (Field, 2024). Some deviations were observed at the tails, indicating minor departures from normality; however, these were not severe and are unlikely to substantially impact the validity of the model (Hair, 2009).



The normality assumption was also assessed through visual inspection of the residual distribution. A histogram of the standardised residuals indicated an approximately bell-shaped and symmetric distribution, with most residuals clustering around zero and tapering gradually at both tails. This pattern

suggests that the residuals were approximately normally distributed, with only minor deviations at the extremes that are unlikely to materially affect model validity (Field, 2004; Osborne & Waters, 2002).

```
                    GVIF Df GVIF^(1/(2*Df))
c29_factor      1.314761  5        1.027743
gender_factor   1.021006  1        1.010448
Minority        1.053542  1        1.026422
occ_factor      1.172183  9        1.008865
stcog           1.224336  1        1.106497
b31_recode      1.247629  1        1.116973
parental_factor 1.012579  1        1.006270
```

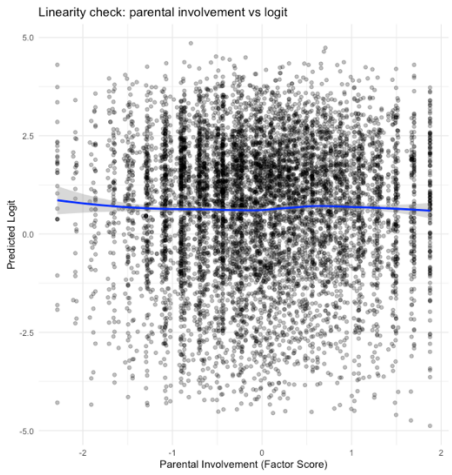
To evaluate potential multicollinearity among predictors, Variance Inflation Factors (VIFs) were calculated. All predictors exhibited VIF values well below the conventional threshold of 5 (range: 1.01–1.31), indicating no serious multicollinearity concerns (Field, 2004; Osborne & Waters, 2002)

Taken together, the diagnostic results suggest that the regression model satisfied the core assumptions of normality and absence of problematic multicollinearity. Therefore, the sequential linear regression analysis proceeded with confidence in the robustness of the model estimates.

Testing Assumption for Logistic Regression

Before estimating the logistic regression model, the dependent variable (educational aspiration, C22) was transformed into a binary variable to meet the requirements of binary logistic regression. Students who reported an aspiration level equal to or above "bachelor's degree" (C22 = 7) were coded as 1, and those below that threshold were coded as 0.

To check for multicollinearity, Variance Inflation Factors (VIFs) were examined for all predictors in the final model (logit_model4). All VIF values were well below the conventional threshold of 5, indicating no evidence of multicollinearity among the predictors (Hair, 2019; Field, 2024).



To assess the assumption of linearity of the logit, the predicted logit values from the model were plotted against each continuous predictor. For example, the plot of the parental involvement factor versus predicted logit showed a reasonably linear trend, supporting the linearity assumption. Additionally, interaction terms between each continuous predictor and its log-transformed value were tested in a supplemental model, and none of these terms reached statistical significance, further confirming that the linearity assumption was met (Hosmer et al., 2013; Menard, 2010).

Taken together, the logistic regression model satisfied the key assumptions of binary logistic regression, including binary outcome, no multicollinearity, and linearity of the logit for continuous predictors.

Appendix D: output for R instruction code

```r
library(haven)
library(dplyr)
library(forcats)
library(psych)
library(ggplot2)
library(tidyr)
library(GPArotation)
library(broom)
library(rcompanion)
library(car)
library(effectsize)

ceps <- read_dta("/Users/aoao/Desktop/Dissertation/CEPS 13-14/cepsw1studentEN（原始数据）.dta")

# we only look at G9,because G7 they don't have question for c29 (perceived teachers'expectation)
ceps_g9 <- ceps %>%
  filter(grade9 == 1)
table(ceps_g9$grade9)

## c22 student aspiration
# output c22 descriptive data (frequency and percentage) and with missing data
attr(ceps$c22, "labels")
ceps_g9 <- ceps_g9 %>%
  mutate(c22_label = fct_recode(as.factor(c22),
                                "Drop out now"                          = "1",
                                "Graduate from junior high school"      = "2",
                                "Go to technical secondary school"      = "3",
                                "Go to vocational high school"          = "4",
                                "Go to senior high school"              = "5",
                                "Graduate from junior college"          = "6",
                                "Get a bachelor degree"                 = "7",
                                "Get a Master degree"                   = "8",
                                "Get a Doctor degree"                   = "9",
                                "They don’t care"                       = "10"
    )
  )
sasp_tab <- ceps_g9 %>%
  count(c22_label, name = "Frequency") %>%
  mutate(
    Label = fct_explicit_na(c22_label, na_level = "Missing"),
    Percentage = round(100 * Frequency / sum(Frequency[Label != "Missing"]), 2)
  ) %>%
  select(Code = c22_label, Label, Frequency, Percentage)

print(sasp_tab)

# combine 10 and 1
ceps_g9 <- ceps_g9 %>%
  mutate(
    c22_combined = fct_recode(as.factor(c22),
                              "Drop out / Don't care"           = "1",
                              "Graduate from junior high school" = "2",
                              "Go to technical school"          = "3",
                              "Go to vocational high school"    = "4",
                              "Go to senior high school"        = "5",
                              "Graduate from junior college"    = "6",
                              "Get a bachelor degree"           = "7",
                              "Get a Master degree"             = "8",
                              "Get a Doctor degree"             = "9",
                              "Drop out / Don't care"           = "10"
    )
  )

c22_tab <- ceps_g9 %>%
  count(c22_combined, name = "Frequency") %>%
  mutate(
    Label = fct_explicit_na(c22_combined, na_level = "Missing"),
    Percentage = round(100 * Frequency / sum(Frequency[Label != "Missing"]), 2)
  ) %>%
  select(Code = c22_combined, Label, Frequency, Percentage)

print(c22_tab)

```

```r
# justification for combining code 1 and 10
ceps_g9 %>%
  filter(c22 %in% c(1, 10)) %>%
  group_by(c22) %>%
  summarise(
    mean_score = mean(stcog, na.rm = TRUE),
    sd_score = sd(stcog, na.rm = TRUE),
    n = n()
  )
t.test(stcog ~ as.factor(c22), data = ceps_g9 %>% filter(c22 %in% c(1, 10)))

ceps_g9 %>%
  filter(c22 %in% 1:10) %>%
  group_by(c22) %>%
  summarise(
    mean_score = mean(stcog, na.rm = TRUE),
    sd_score = sd(stcog, na.rm = TRUE),
    n = n()
  ) %>%
  arrange(as.numeric(c22))

# convert c22 to a continuous variable
ceps_g9_clean <- ceps_g9 %>%
  mutate(
    c22_recode = case_when(
      as.character(c22) %in% c("1", "10") ~ "1_10",
      TRUE ~ as.character(c22)
    ),
    c22_recode_num = case_when(
      c22_recode == "1_10" ~ 1,
      c22_recode == "2" ~ 2,
      c22_recode == "3" ~ 3,
      c22_recode == "4" ~ 4,
      c22_recode == "5" ~ 5,
      c22_recode == "6" ~ 6,
      c22_recode == "7" ~ 7,
      c22_recode == "8" ~ 8,
      c22_recode == "9" ~ 9,
      TRUE ~ NA_real_
    )
  ) %>%
  drop_na(c22_recode_num)

sasp_numeric <- ceps_g9_clean$c22_recode_num
summary(sasp_numeric)

# plot the histogram
labels_ordered <- c(
  "Low aspiration (Drop out / Don't care)",
  "Graduate from junior high school",
  "Go to technical secondary school or technical school",
  "Go to vocational high school",
  "Go to senior high school",
  "Graduate from junior college",
  "Get a bachelor degree",
  "Get a Master degree",
  "Get a Doctor degree"
)

sasp_tab <- ceps_g9_clean %>%
  count(c22_recode_num, name = "Frequency") %>%
  mutate(
    Percentage = round(100 * Frequency / sum(Frequency), 2),
    Label = factor(labels_ordered, levels = labels_ordered)
  )

ggplot(sasp_tab, aes(x = Label, y = Frequency)) +
  geom_col(fill = "#002147", color = "White") +
  geom_text(
    aes(label = paste0(Frequency, "\n(", Percentage, "%)")),
    vjust = -0.5, size = 3, color = "black"
  ) +
  labs(
    title = "Distribution of Students' Educational Aspirations (C22)",
    x = "Educational Aspiration Category",
    y = "Frequency"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 30, hjust = 1),
```

```
    plot.title = element_text(face = "bold", size = 14)
  )

# check normality
describe(sasp_numeric)

qqnorm(sasp_numeric, main = "Q-Q Plot of Educational Aspirations (C22)",
       ylab = "Observed Quantiles", xlab = "Theoretical Quantiles")
qqline(sasp_numeric, col = "red")

## c29 perceived teachers' expectation
attr(ceps$c29, "labels")
c29_labels <- c(
  "1" = "Going to a key senior high school",
  "2" = "Going to an ordinary senior high school",
  "3" = "Going to a vocational high school",
  "4" = "Going to a technical secondary/technical school",
  "5" = "Not going any further after junior high",
  "6" = "None"
)
c29_tab <- ceps_g9 %>%
  count(c29, name = "Frequency") %>%
  mutate(
    Label = c29_labels[as.character(c29)],
    Label = ifelse(is.na(Label), "Missing", Label),
    Percentage = round(100 * Frequency / sum(Frequency[Label != "Missing"]), 2)
  ) %>%
  select(Code = c29, Label, Frequency, Percentage)

print(c29_tab)

# Summary table
ceps_model_data <- ceps_g9 %>%
  filter(!is.na(c22), !is.na(c29)) %>%
  mutate(
    c22_recode = case_when(
      as.character(c22) %in% c("1", "10") ~ "1_10",
      TRUE ~ as.character(c22)
    ),
    c22_recode_num = dplyr::recode(
      c22_recode,
      "1_10" = 1,
      "2" = 2, "3" = 3, "4" = 4, "5" = 5,
      "6" = 6, "7" = 7, "8" = 8, "9" = 9,
      .default = NA_real_
    ),
    c29_factor = factor(
      c29,
      levels = 1:6,
      labels = c(
        "Going to a key senior high school",
        "Going to an ordinary senior high school",
        "Going to a vocational high school",
        "Going to a technical secondary/technical school",
        "Not going any further after junior high",
        "None"
      )
    ),
    c29_factor = fct_relevel(c29_factor, "Going to a key senior high school")
  ) %>%
  drop_na(c22_recode_num)

teacher_summary <- ceps_model_data %>%
  group_by(c29_factor) %>%
  summarise(
    N = n(),
    Mean_Aspiration = round(mean(c22_recode_num), 2),
    SD_Aspiration = round(sd(c22_recode_num), 2),
    Percent_BachelorPlus = round(100 * mean(c22_recode_num >= 7), 2)
  )

print(teacher_summary)

# effect size
anova_model <- aov(c22_recode_num ~ teachers_expectation_label, data = ceps_g9_clean)

eta_squared(anova_model)

omega_squared(anova_model)
```

```r
### individual factors
## gender
# Frequency & Percentage Table (with missing)
attr(ceps$a01, "labels")
gender_labels <- c("1" = "Male", "2" = "Female")
ceps_model_data <- ceps_g9 %>%
  mutate(
    c22_recode = case_when(
      as.character(c22) %in% c("1", "10") ~ "1_10",
      TRUE ~ as.character(c22)
    ),

    c22_recode_num = case_when(
      c22_recode == "1_10" ~ 1,
      c22_recode == "2" ~ 2,
      c22_recode == "3" ~ 3,
      c22_recode == "4" ~ 4,
      c22_recode == "5" ~ 5,
      c22_recode == "6" ~ 6,
      c22_recode == "7" ~ 7,
      c22_recode == "8" ~ 8,
      c22_recode == "9" ~ 9,
      TRUE ~ NA_real_
    ),

    gender_factor = factor(a01, levels = c(1, 2), labels = c("Male", "Female")),
    gender_factor = fct_relevel(gender_factor, "Male")
  ) %>%
  drop_na(c22_recode_num, a01) ❱

gender_summary <- ceps_model_data %>%
  group_by(gender_factor) %>%
  summarise(
    N = n(),
    Mean_Aspiration = round(mean(c22_recode_num), 2),
    SD_Aspiration = round(sd(c22_recode_num), 2),
    Percent_BachelorPlus = round(100 * mean(c22_recode_num >= 7), 2)
  )

print(gender_summary)
```

```r

#effect size
cohens_d_result <- cohens_d(c22_recode_num ~ gender_factor,
                            data = ceps_model_data,
                            pooled_sd = TRUE,
                            ci = 0.95)
print(cohens_d_result)

## ethnicity
# Frequency & Percentage Table (with missing)
attr(ceps_g9$a03, "labels")
ethnicity_labels <- c("1" = "Han", "0" = "Minority")
ethnicity_tab <- ceps_g9 %>%
  mutate(
    ethnicity_binary = case_when(
      a03 == 1 ~ 1,
      a03 %in% 0:8 ~ 0,
      TRUE ~ NA_real_
    )
  ) %>%
  count(ethnicity_binary, name = "Frequency") %>%
  mutate(
    Label = ethnicity_labels[as.character(ethnicity_binary)],
    Label = ifelse(is.na(Label), "Missing", Label),
    Percentage = round(100 * Frequency / sum(Frequency[Label != "Missing"]), 2)
  ) %>%
  select(Code = ethnicity_binary, Label, Frequency, Percentage)

print(ethnicity_tab)

ceps_model_data <- ceps_g9 %>%
  mutate(
    c22_recode = case_when(
      as.character(c22) %in% c("1", "10") ~ "1_10",
      TRUE ~ as.character(c22)
    ),
    c22_recode_num = dplyr::recode(
      c22_recode,
      "1_10" = 1,
      "2" = 2, "3" = 3, "4" = 4, "5" = 5,
      "6" = 6, "7" = 7, "8" = 8, "9" = 9,
```

```r
      .default = NA_real_
    ),
    ethnicity_binary = case_when(
      a03 == 1 ~ 1,
      a03 %in% 0:8 ~ 0,
      TRUE ~ NA_real_
    )
  ) %>%
  drop_na(c22_recode_num, ethnicity_binary)

# summary table
ethnicity_summary <- ceps_model_data %>%
  group_by(ethnicity_binary) %>%
  summarise(
    N = n(),
    Mean_Aspiration = round(mean(c22_recode_num), 2),
    SD_Aspiration = round(sd(c22_recode_num), 2),
    Percent_BachelorPlus = round(100 * mean(c22_recode_num >= 7), 2)
  ) %>%
  mutate(Ethnicity = ethnicity_labels[as.character(ethnicity_binary)]) %>%
  select(Ethnicity, everything(), -ethnicity_binary)

print(ethnicity_summary)

# effect size
ceps_model_data$ethnicity_factor <- factor(
  ceps_model_data$ethnicity_binary,
  levels = c(0, 1),
  labels = c("Minority", "Han")
)

cohens_d_ethnicity <- cohens_d(
  c22_recode_num ~ ethnicity_factor,
  data = ceps_model_data,
  pooled_sd = TRUE,
  ci = 0.95
)
print(cohens_d_ethnicity)

```

```r
## SES
## parental occupation
# take the lowest among the two, output b08a b08b descrptive data (frequency and percentage) and with missing data
ceps_g9 <- ceps_g9 %>%
  mutate(rowid = row_number())

ceps_g9_clean <- ceps_g9_clean %>%
  mutate(rowid = row_number())

ceps_g9 <- ceps_g9 %>%
  mutate(parent_occ_max = pmin(b08a, b08b, na.rm = TRUE))

ceps_model_data <- ceps_g9_clean %>%
  left_join(ceps_g9 %>% select(rowid, parent_occ_max), by = "rowid")

ceps_model_data <- ceps_model_data %>%
  drop_na(parent_occ_max)

occ_labels <- setNames(
  c(
    "Government official, staff of public institutions, civil servant",
    "Middle/Senior management personnel of enterprises/corporations",
    "Teacher, engineer, doctor, lawyer",
    "Technical worker (including driver)",
    "Ordinary staff or worker in production or manufacturing industry",
    "Ordinary staff or worker in business or service industry",
    "Self-employed worker",
    "Peasant",
    "Unemployed or laid-off worker",
    "Other"
  ),
  as.character(1:10)
)

ceps_model_data <- ceps_model_data %>%
  mutate(
    occ_factor = factor(parent_occ_max, levels = 1:10, labels = unname(occ_labels)),
    occ_factor = fct_relevel(occ_factor, "Peasant")
  )

```

```r
occupation_summary <- ceps_model_data %>%
  group_by(occ_factor) %>%
  summarise(
    N = n(),
    Mean_Aspiration = round(mean(c22_recode_num), 2),
    SD_Aspiration = round(sd(c22_recode_num), 2),
    Percent_BachelorPlus = round(100 * mean(c22_recode_num >= 7), 2)
  )

print(occupation_summary)

# check for the missing value
missing_c22 <- sum(is.na(ceps_g9$c22))
missing_occ <- sum(is.na(ceps_g9$parent_occ_max))

missing_both <- sum(is.na(ceps_g9$c22) & is.na(ceps_g9$parent_occ_max))

cat("Missing in c22: ", missing_c22, "\n")
cat("Missing in parent_occ_max: ", missing_occ, "\n")
cat("Missing in both: ", missing_both, "\n")

cat("Unique missing count (union): ",
    sum(is.na(ceps_g9$c22) | is.na(ceps_g9$parent_occ_max)), "\n")

cat("Expected: ", missing_c22 + missing_occ - missing_both, "\n")

# effect size
ceps_model_data <- ceps_model_data %>%
  mutate(
    occ_factor = factor(
      parent_occ_max,
      levels = 1:10,
      labels = c(
        "Government official/public institution staff",
        "Middle/Senior management",
        "Teacher/engineer/doctor/lawyer",
        "Technical worker",
        "Manufacturing worker",
        "Service worker",
        "Self-employed",
        "Peasant",
        "Unemployed/laid-off",
        "Other"
      )
    )
  ) %>%
  tidyr::drop_na(c22_recode_num, occ_factor)

anova_occ <- aov(c22_recode_num ~ occ_factor, data = ceps_model_data)
summary(anova_occ)
eta_squared(anova_occ, partial = FALSE, ci = 0.95)
omega_squared(anova_occ, partial = FALSE, ci = 0.95)

## cognitive score
quartile_breaks <- quantile(ceps_model_data$stcog, probs = c(0, 0.25, 0.5, 0.75, 1), na.rm = TRUE)

quartile_labels <- c("Q1 (Lowest Ability)", "Q2", "Q3", "Q4 (Highest Ability)")

ceps_model_data <- ceps_model_data %>%
  mutate(
    cog_quartile = cut(
      stcog,
      breaks = quartile_breaks,
      include.lowest = TRUE,
      labels = quartile_labels
    )
  )

cog_tab <- ceps_model_data %>%
  count(cog_quartile, name = "Frequency") %>%
  mutate(
    Percentage = round(100 * Frequency / sum(Frequency), 2)
  )

print(cog_tab)

cog_summary <- ceps_model_data %>%
  group_by(cog_quartile) %>%
  summarise(
    N = n(),
    Mean_Aspiration = round(mean(c22_recode_num, na.rm = TRUE), 2),
```

```r
    SD_Aspiration = round(sd(c22_recode_num, na.rm = TRUE), 2),
    Percent_BachelorPlus = round(100 * mean(c22_recode_num >= 7, na.rm = TRUE), 2)
  )

print(cog_summary)

# effect size
anova_cog <- aov(c22_recode_num ~ cog_quartile, data = ceps_model_data)

eta_squared(anova_cog, partial = FALSE, ci = 0.95)

omega_squared(anova_cog, partial = FALSE, ci = 0.95)

### family factors
## parental involvement
# Run EFA (Principal Axis Factoring)
involve_vars <- c("b2201", "b2202", "b2301", "b2302", "b2303", "b24a1", "b24b1")

involve_labels <- c(
  b2201 = "Parent checked your homework",
  b2202 = "Parent gave instruction on your homework",
  b2301 = "Parents care about your homework/exam",
  b2302 = "Parents care about your behavior at school",
  b2303 = "Parents care about your school attendance",
  b24a1 = "Mother: Discussed school events with you",
  b24b1 = "Father: Discussed school events with you"
)

get_freq_pct_labeled <- function(varname) {
  tab <- as.data.frame(table(ceps_g9[[varname]], useNA = "ifany"))
  names(tab) <- c("Value", "Frequency")
  tab$Value <- as.character(tab$Value)
  tab$Value[is.na(tab$Value)] <- "Missing"
  total <- sum(tab$Frequency[tab$Value != "Missing"])
  tab$Percentage <- round(100 * tab$Frequency / total, 2)
  tab$Variable <- varname
  tab$Label <- involve_labels[[varname]]
  return(tab[, c("Variable", "Label", "Value", "Frequency", "Percentage")])
}

```

```r
involve_summary <- do.call(rbind, lapply(involve_vars, get_freq_pct_labeled))
print(involve_summary)

efa_data <- ceps_g9[, involve_vars]
efa_data_complete <- na.omit(efa_data)

KMO_result <- KMO(efa_data_complete)
bartlett_result <- cortest.bartlett(efa_data_complete)
print(KMO_result)
print(bartlett_result)

# Scree Plot
involve_vars <- c("b2201", "b2202", "b2301", "b2302", "b2303", "b24a1", "b24b1")
efa_data <- ceps_g9[, involve_vars]

efa_data_complete <- na.omit(efa_data)

fa.parallel(efa_data_complete, fa = "fa", fm = "pa", main = "Scree Plot & Parallel Analysis\n(Parental Involvement)")

efa_result <- fa(efa_data_complete, nfactors = 3, fm = "pa", rotate = "oblimin")
print(efa_result)
fa.diagram(efa_result)
factor_scores <- as.data.frame(efa_result$scores)
colnames(factor_scores) <- c("involve_supervision", "involve_discipline", "involve_communication")

ceps_g9$row_id <- 1:nrow(ceps_g9)

factor_scores$row_id <- as.numeric(rownames(factor_scores))

ceps_g9 <- left_join(ceps_g9, factor_scores, by = "row_id")

# spearman correlation
spearman_rho_results <- sapply(involve_vars, function(var) {
  cor(ceps_g9[[var]], ceps_g9$c22_recode_num, method = "spearman", use = "complete.obs")
})

spearman_rho_df <- data.frame(
  Variable = involve_vars,
  Label = involve_labels,
  Spearman_rho = round(spearman_rho_results, 3)
```

```r
)

print(spearman_rho_df)

# plot the histogram
efa_scores_df <- data.frame(
  parental_factor = efa_result$scores[, 1],
  row_id = 1:nrow(efa_result$scores)
)

ceps_g9$row_id <- 1:nrow(ceps_g9)

print(nrow(ceps_g9))
print(nrow(efa_scores_df))

ceps_g9 <- left_join(ceps_g9, efa_scores_df, by = "row_id")

bin_width <- 0.5

hist_data <- ceps_g9 %>%
  filter(!is.na(parental_factor.x)) %>%
  mutate(bin = cut(parental_factor.x,
                   breaks = seq(floor(min(parental_factor.x)), ceiling(max(parental_factor.x)), by = bin_width),
                   include.lowest = TRUE)) %>%
  group_by(bin) %>%
  summarise(Frequency = n(), .groups = "drop") %>%
  mutate(Percentage = round(100 * Frequency / sum(Frequency), 2),
         Label = paste0(Frequency, "\n(", Percentage, "%)"))

ggplot(hist_data, aes(x = bin, y = Frequency)) +
  geom_col(fill = "#002147", color = "white", width = 1) +
  geom_text(aes(label = Label), vjust = -0.5, size = 3.5) +
  labs(
    title = "Distribution of Parental Involvement (Factor Score)",
    x = "Parental Involvement Factor Score Bin",
    y = "Frequency"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

```r
ceps_g9$parental_factor <- ceps_g9$parental_factor.x

ceps_g9$row_id <- 1:nrow(ceps_g9)
efa_scores_df$row_id <- 1:nrow(efa_scores_df)
ceps_g9 <- left_join(ceps_g9, efa_scores_df[, c("row_id", "parental_factor")], by = "row_id")
summary(ceps_g9$parental_factor)
describe(ceps_g9$parental_factor)

# Q-Q Plot to assess normality
qqnorm(ceps_g9$parental_factor,
       main = "Normal Q-Q Plot for Parental Involvement Factor Score")
qqline(ceps_g9$parental_factor, col = "red", lwd = 2)

# summary table
efa_complete <- na.omit(ceps_g9[, c("c22", involve_vars)])
valid_row_ids <- as.numeric(rownames(efa_complete))

ceps_g9_filtered <- ceps_g9[valid_row_ids, ]

efa_result <- fa(efa_complete[, involve_vars], nfactors = 3, fm = "pa", rotate = "oblimin")
parental_factor <- efa_result$scores[, 1]
ceps_g9_filtered$parental_factor <- parental_factor

ceps_g9_filtered <- ceps_g9_filtered %>%
  mutate(
    c22_recode = case_when(
      as.character(c22) %in% c("1", "10") ~ "1_10",
      TRUE ~ as.character(c22)
    ),
    c22_recode_num = dplyr::recode(
      c22_recode,
      "1_10" = 1,
      "2" = 2, "3" = 3, "4" = 4, "5" = 5,
      "6" = 6, "7" = 7, "8" = 8, "9" = 9,
      .default = NA_real_
    )
  )

quartile_breaks <- quantile(ceps_g9_filtered$parental_factor, probs = seq(0, 1, 0.25), na.rm = TRUE)
quartile_labels <- c("Q1 (Lowest)", "Q2", "Q3", "Q4 (Highest)")
```

```r

ceps_g9_filtered <- ceps_g9_filtered %>%
  mutate(
    parental_quartile = cut(
      parental_factor,
      breaks = quartile_breaks,
      include.lowest = TRUE,
      labels = quartile_labels
    )
  )

quartile_summary <- ceps_g9_filtered %>%
  group_by(parental_quartile) %>%
  summarise(
    N = n(),
    Mean_Aspiration = round(mean(c22_recode_num, na.rm = TRUE), 2),
    SD_Aspiration = round(sd(c22_recode_num, na.rm = TRUE), 2),
    Percent_BachelorPlus = round(100 * mean(c22_recode_num >= 7, na.rm = TRUE), 2)
  )

print(quartile_summary)

# effect size
efa_complete <- na.omit(ceps_g9[, involve_vars])
valid_ids    <- as.numeric(rownames(efa_complete))

fa1 <- fa(efa_complete, nfactors = 3, fm = "pa", rotate = "oblimin")
parental_factor <- as.numeric(fa1$scores[, 1])

dat <- ceps_g9[valid_ids, , drop = FALSE] %>%
  mutate(
    parental_factor = parental_factor,
    c22_recode = ifelse(as.character(c22) %in% c("1","10"), "1_10", as.character(c22)),
    c22_recode_num = dplyr::recode(
      c22_recode,
      "1_10" = 1,
      "2" = 2, "3" = 3, "4" = 4, "5" = 5,
      "6" = 6, "7" = 7, "8" = 8, "9" = 9,
      .default = NA_real_
    )
  ) %>%
  tidyr::drop_na(parental_factor, c22_recode_num)

qbreaks <- quantile(dat$parental_factor, probs = seq(0,1,0.25), na.rm = TRUE)
qlabels <- c("Q1 (Lowest)", "Q2", "Q3", "Q4 (Highest)")
dat <- dat %>%
  mutate(parental_quartile = cut(parental_factor, breaks = qbreaks,
                                 include.lowest = TRUE, labels = qlabels))

aov_pi <- aov(c22_recode_num ~ parental_quartile, data = dat)
eta_squared(aov_pi,  partial = FALSE, ci = 0.95)
omega_squared(aov_pi, partial = FALSE, ci = 0.95)

## parental expectation
# output b31 descrptive data (frequency and percentage) and with missing data
# plot the histogram
b31_clean <- ceps_g9$b31

b31_clean[!is.na(b31_clean) & b31_clean %in% c(1, 10)] <- 1

b31_clean[!is.na(b31_clean) & b31_clean > 1] <- b31_clean[!is.na(b31_clean) & b31_clean > 1] - 1

b31_clean <- na.omit(b31_clean)

describe(b31_clean)

b31_df <- data.frame(b31 = b31_clean) %>%
  mutate(bin = factor(b31)) %>%
  group_by(bin) %>%
  summarise(Frequency = n(), .groups = "drop") %>%
  mutate(Percentage = round(100 * Frequency / sum(Frequency), 2),
         Label = paste0(Frequency, " (", Percentage, "%)"))

ggplot(b31_df, aes(x = bin, y = Frequency)) +
  geom_col(fill = "#002147", color = "white", width = 1) +
  geom_text(aes(label = Label), vjust = -0.5, size = 4) +
  labs(title = "Distribution of Parental Expectations (B31)",
       x = "Parental Expectation Score (1 = Low, 9 = Doctoral Degree)",
       y = "Frequency") +
  theme_minimal()

```

```r
# Q–Q Plot to assess normality
qqnorm(b31_clean,
       main = "Normal Q–Q Plot for Parental Expectations (B31)",
       pch = 19, col = "black")
qqline(b31_clean, col = "red", lwd = 2)

# summary graph
ceps_g9 <- ceps_g9 %>%
  mutate(
    c22_recode = case_when(
      as.character(c22) %in% c("1", "10") ~ "1_10",
      TRUE ~ as.character(c22)
    ),

    c22_recode_num = case_when(
      c22_recode == "1_10" ~ 1,
      c22_recode == "2" ~ 2,
      c22_recode == "3" ~ 3,
      c22_recode == "4" ~ 4,
      c22_recode == "5" ~ 5,
      c22_recode == "6" ~ 6,
      c22_recode == "7" ~ 7,
      c22_recode == "8" ~ 8,
      c22_recode == "9" ~ 9,
      TRUE ~ NA_real_
    ),

    b31_recode = case_when(
      b31 %in% c(1, 10) ~ 1,
      TRUE ~ as.numeric(b31)
    )
  ) %>%
  drop_na(c22_recode_num, b31_recode)
ceps_g9 <- ceps_g9 %>%
  mutate(
    c22_recode = case_when(
      as.character(c22) %in% c("1", "10") ~ "1_10",
      TRUE ~ as.character(c22)
    ),

    c22_recode_num = case_when(
      c22_recode == "1_10" ~ 1,
      c22_recode == "2" ~ 2,
      c22_recode == "3" ~ 3,
      c22_recode == "4" ~ 4,
      c22_recode == "5" ~ 5,
      c22_recode == "6" ~ 6,
      c22_recode == "7" ~ 7,
      c22_recode == "8" ~ 8,
      c22_recode == "9" ~ 9,
      TRUE ~ NA_real_
    ),

    b31_recode = case_when(
      b31 %in% c(1, 10) ~ 1,
      TRUE ~ as.numeric(b31)
    )
  ) %>%
  drop_na(c22_recode_num, b31_recode)


labels_b31 <- c(
  "Low expectation (Drop out / Don't care)",
  "Graduate from junior high school",
  "Go to technical secondary school",
  "Go to vocational high school",
  "Go to senior high school",
  "Graduate from junior college",
  "Get a bachelor degree",
  "Get a Master degree",
  "Get a Doctor degree"
)

b31_expectation_summary <- ceps_b31_summary %>%
  group_by(b31_recode) %>%
  summarise(
    N = n(),
    Mean_Aspiration = round(mean(c22_recode_num), 2),
    SD_Aspiration = round(sd(c22_recode_num), 2),
    Percent_BachelorPlus = round(100 * mean(c22_recode_num >= 7), 2)
```

```r
  ) %>%
  mutate(Parental_Expectation = labels_b31[b31_recode]) %>%
  select(Parental_Expectation, everything())
print(b31_expectation_summary )

# effect size
ceps_b31 <- ceps_g9 %>%
  mutate(
    c22_recode = if_else(as.character(c22) %in% c("1","10"), "1_10", as.character(c22)),
    c22_recode_num = dplyr::recode(
      c22_recode, "1_10"=1, "2"=2, "3"=3, "4"=4, "5"=5,
      "6"=6, "7"=7, "8"=8, "9"=9, .default = NA_real_
    ),
    b31_recode = dplyr::case_when(
      is.na(b31) ~ NA_real_,
      b31 %in% c(1,10) ~ 1,
      TRUE ~ as.numeric(b31)
    )
  ) %>%
  select(c22_recode_num, b31_recode) %>%
  drop_na()

labels_b31 <- c(
  "Low expectation (Drop out / Don't care)",
  "Graduate from junior high school",
  "Go to technical secondary school",
  "Go to vocational high school",
  "Go to senior high school",
  "Graduate from junior college",
  "Get a bachelor degree",
  "Get a Master degree",
  "Get a Doctor degree"
)

ceps_b31 <- ceps_b31 %>%
  mutate(b31_factor = factor(b31_recode, levels = 1:9, labels = labels_b31))

anova_b31 <- aov(c22_recode_num ~ b31_factor, data = ceps_b31)

eta_sq_b31   <- eta_squared(anova_b31, partial = FALSE, ci = 0.95)
```

```r
omega_sq_b31 <- omega_squared(anova_b31, partial = FALSE, ci = 0.95)

print(eta_sq_b31)
print(omega_sq_b31)

# plot figure for regression coefficient of teacher expectation across the model
ceps_plot <- ceps_final %>%
  mutate(
    teacher_expectation_group = case_when(
      c29_factor == "Key SHS" ~ "Key SHS",
      c29_factor == "Ordinary SHS" ~ "Ordinary SHS",
      c29_factor %in% c("Vocational", "Technical", "No further") ~ "Voc/Further/Leave",
      c29_factor == "None" ~ "None",
      TRUE ~ NA_character_
    )
  ) %>%
  drop_na(teacher_expectation_group)

ceps_plot$teacher_expectation_group <- factor(
  ceps_plot$teacher_expectation_group,
  levels = c("Key SHS", "Ordinary SHS", "Voc/Further/Leave", "None")
)

model1_simple <- lm(c22_recode_num ~ teacher_expectation_group, data = ceps_plot)
model2_simple <- lm(c22_recode_num ~ teacher_expectation_group + gender_factor + Minority + occ_factor, data = ceps_plot)
model3_simple <- lm(c22_recode_num ~ teacher_expectation_group + gender_factor + Minority + occ_factor + stcog, data = ceps_plot)
model4_simple <- lm(c22_recode_num ~ teacher_expectation_group + gender_factor + Minority + occ_factor + stcog + parental_factor + b31_recode, data = ceps_plot)

extract_coefs <- function(model, model_name) {
  broom::tidy(model) %>%
    dplyr::filter(str_detect(term, "teacher_expectation_group")) %>%
    mutate(
      Expectation = str_remove(term, "teacher_expectation_group"),
      Model = model_name
    ) %>%
    select(Expectation, estimate, std.error, Model)
}

```

```r
coef_df <- dplyr::bind_rows(
  extract_coefs(model1_simple, "Model 1"),
  extract_coefs(model2_simple, "Model 2"),
  extract_coefs(model3_simple, "Model 3"),
  extract_coefs(model4_simple, "Model 4")
)

coef_df$Expectation <- factor(coef_df$Expectation,
                              levels = c("Ordinary SHS", "Voc/Further/Leave", "None"))

ggplot(coef_df, aes(x = Expectation, y = estimate, fill = Model)) +
  geom_col(position = position_dodge(width = 0.8)) +
  geom_errorbar(aes(ymin = estimate - std.error, ymax = estimate + std.error),
                width = 0.2, position = position_dodge(width = 0.8)) +
  labs(
    title = "Regression Coefficients of Teacher Expectations Across Models",
    x = "Teacher Expectation (Reference: Key SHS)",
    y = "Coefficient Estimate on Educational Aspiration"
  ) +
  theme_minimal() +
  theme(
    text = element_text(family = "serif", size = 12),
    plot.title = element_text(face = "bold", hjust = 0.5)
  )

### sequential linear regression
## check assumptions
# Linear model
lm_model <- lm(c22_recode_num ~ c29_factor + gender_factor + Minority +
                occ_factor + stcog + b31_recode + parental_factor, data = ceps_final)

describe(ceps_final$c22_recode_num)

```

```r
# Histogram
hist(ceps_final$c22_recode_num,
     breaks = 7,
     main = "Histogram of Educational Aspiration",
     xlab = "Aspiration (C22)",
     col = "lightgray",
     border = "white")

plot(lm_model, which = 1)  # Residuals vs Fitted
plot(lm_model, which = 2)  # Q-Q Plot

hist(residuals(lm_model),
     breaks = 20,
     main = "Histogram of Residuals",
     xlab = "Residuals")

library(car)
vif(lm_model)  # VIF > 5 --> multicollinearity

# listwise delete all the missing values
ceps_final <- ceps_g9 %>%
  mutate(
    c22_recode = case_when(as.character(c22) %in% c("1", "10") ~ "1_10", TRUE ~ as.character(c22)),
    c22_recode_num = dplyr::recode(c22_recode,
                                   "1_10" = 1, "2" = 2, "3" = 3, "4" = 4, "5" = 5,
                                   "6" = 6, "7" = 7, "8" = 8, "9" = 9,
                                   .default = NA_real_
    ),

    c29_factor = factor(c29,
                        levels = 1:6,
                        labels = c("Key SHS", "Ordinary SHS", "Vocational", "Technical", "No further", "None")
    ) %>% fct_relevel("Key SHS"),

    gender_factor = factor(a01, levels = c(1, 2), labels = c("Male", "Female")) %>% fct_relevel("Male"),

    Minority = case_when(a03 == 1 ~ 0, a03 %in% 0:8 ~ 1, TRUE ~ NA_real_),

    parent_occ_max = pmin(b08a, b08b, na.rm = TRUE),
    occ_factor = factor(parent_occ_max, levels = 1:10) %>% fct_relevel("8"),
```

```r

    b31_recode = case_when(b31 %in% c(1, 10) ~ 1, TRUE ~ as.numeric(b31))
  ) %>%

  select(c22_recode_num, c29_factor, gender_factor, Minority,
         occ_factor, b31_recode, stcog, parental_factor) %>%
  drop_na()

# Model 1: Teacher expectations only
model1 <- lm(c22_recode_num ~ c29_factor, data = ceps_final)
summary(model1)

# Model 2: gender + ethnicity + SES
model2 <- lm(c22_recode_num ~ c29_factor + gender_factor + Minority + occ_factor, data = ceps_final)
summary(model2)

# Model 3: + Cognitive ability
model3 <- lm(c22_recode_num ~ c29_factor + gender_factor + Minority + occ_factor + stcog, data = ceps_final)
summary(model3)

# Model 4: + Family-level (B31 + EFA-based involvement)
model4 <- lm(c22_recode_num ~ c29_factor + gender_factor + Minority + occ_factor + stcog +
               b31_recode + parental_factor, data = ceps_final)
summary(model4)

## compare teacher's expecation vs parent's expecation
# Model A: teacher's expecation only (C29)
model_teacher_only <- lm(c22_recode_num ~ c29_factor, data = ceps_final)
summary(model_teacher_only)
# Model B: parent's expecation only (B31)
model_parent_only <- lm(c22_recode_num ~ b31_recode, data = ceps_final)
summary(model_parent_only)

names(coef(model4))
linearHypothesis(model4, c("c29_factorOrdinary SHS = 0",
                           "c29_factorVocational = 0",
                           "c29_factorTechnical = 0",
                           "c29_factorNo further = 0",
                           "c29_factorNone = 0"))
linearHypothesis(model4, "b31_recode = 0")
```

```r


### Robutness Checking by using logistic model
# check assumptions
# check for multicollinearity
ceps_logit_df <- model.frame(logit_model4)
ceps_logit_df$logit_pred <- predict(logit_model4, type = "link")

vif_table <- vif(logit_model4)
print(vif_table)

ggplot(ceps_logit_df, aes(x = stcog, y = logit_pred)) +
  geom_point(alpha = 0.3) +
  geom_smooth(method = "loess", color = "blue") +
  labs(
    title = "Linearity Check: Cognitive Ability vs Logit",
    x = "Cognitive Ability (stcog)",
    y = "Predicted Logit"
  ) +
  theme_minimal()

ggplot(ceps_logit_df, aes(x = b31_recode, y = logit_pred)) +
  geom_point(alpha = 0.3) +
  geom_smooth(method = "loess", color = "blue") +
  labs(
    title = "Linearity Check: Parental Expectation vs Logit",
    x = "Parental Expectation (b31_recode)",
    y = "Predicted Logit"
  ) +
  theme_minimal()

ggplot(ceps_logit_df, aes(x = parental_factor, y = logit_pred)) +
  geom_point(alpha = 0.3) +
  geom_smooth(method = "loess", color = "blue") +
  labs(
    title = "Linearity Check: Parental Involvement vs Logit",
    x = "Parental Involvement (Factor Score)",
    y = "Predicted Logit"
  ) +
  theme_minimal()
```

```r

ceps_logit_df$pearson_resid <- residuals(logit_model4, type = "pearson")

hist(ceps_logit_df$pearson_resid, breaks = 30,
     main = "Histogram of Pearson Residuals",
     xlab = "Pearson Residuals",
     col = "gray", border = "white")

qqnorm(ceps_logit_df$pearson_resid,
       main = "Normal Q-Q Plot of Pearson Residuals",
       pch = 20, col = "black")
qqline(ceps_logit_df$pearson_resid, col = "red", lwd = 2)

# check for linearity
ceps_final$logit_pred <- predict(logit_model4, type = "link")

# Plot for cognitive ability
ggplot(ceps_final, aes(x = stcog, y = logit_pred)) +
  geom_point(alpha = 0.3) +
  geom_smooth(method = "loess", color = "blue") +
  labs(title = "Linearity check: stcog vs logit",
       x = "Cognitive Ability (stcog)",
       y = "Predicted Logit") +
  theme_minimal()

# Plot for parental involvement
ggplot(ceps_final, aes(x = parental_factor, y = logit_pred)) +
  geom_point(alpha = 0.3) +
  geom_smooth(method = "loess", color = "blue") +
  labs(title = "Linearity check: parental involvement vs logit",
       x = "Parental Involvement (Factor Score)",
       y = "Predicted Logit") +
  theme_minimal()

```

```r
# Plot for parental expectation
ggplot(ceps_final, aes(x = b31_recode, y = logit_pred)) +
  geom_point(alpha = 0.3) +
  geom_smooth(method = "loess", color = "blue") +
  labs(title = "Linearity check: parental expectation vs logit",
       x = "Parental Educational Expectation (b31_recode)",
       y = "Predicted Logit") +
  theme_minimal()

ceps_final <- ceps_final %>%
  mutate(
    aspiration_binary = ifelse(c22_recode_num >= 7, 1, 0)
  )

# Logistic Model 1: only teacher expectation
logit_model1 <- glm(aspiration_binary ~ c29_factor, data = ceps_final, family = binomial)

# Logistic Model 2: add gender, ethnicity, SES
logit_model2 <- glm(aspiration_binary ~ c29_factor + gender_factor + Minority + occ_factor, data = ceps_final, family = binomial)

# Logistic Model 3: add cognitive ability
logit_model3 <- glm(aspiration_binary ~ c29_factor + gender_factor + Minority + occ_factor + stcog, data = ceps_final, family = binomial)

# Logistic Model 4: add family-level variables
logit_model4 <- glm(aspiration_binary ~ c29_factor + gender_factor + Minority + occ_factor + stcog + b31_recode + parental_factor, data = ceps_final, family = binomial)

summary(logit_model1)
summary(logit_model2)
summary(logit_model3)
summary(logit_model4)

# Odds Ratio summary function
get_odds_ratios <- function(model) {
  tidy(model, exponentiate = TRUE, conf.int = TRUE) %>%
    select(term, estimate, conf.low, conf.high, p.value) %>%
    rename(
      Odds_Ratio = estimate,
      CI_lower = conf.low,
      CI_upper = conf.high,
      p = p.value
```

```
    )
}
or1 <- get_odds_ratios(logit_model1)
or2 <- get_odds_ratios(logit_model2)
or3 <- get_odds_ratios(logit_model3)
or4 <- get_odds_ratios(logit_model4)

print(or1)
print(or2)
print(or3)
print(or4)

# Run Nagelkerke R² for each model
nagelkerke(logit_model1)
nagelkerke(logit_model2)
nagelkerke(logit_model3)
nagelkerke(logit_model4)

## chekcing for all the missingness (Appendix A.1)
vars <- c("c22", "c29", "a01", "a03", "b08a", "b08b", "b31", "stcog",
          "b2201", "b2202", "b2301", "b2302", "b2303", "b24a1", "b24b1")

unique_missing_each <- sapply(vars, function(var) {
  this_na <- is.na(ceps_g9[[var]])

  other_vars <- setdiff(vars, var)
  others_not_na <- apply(ceps_g9[, other_vars], 1, function(x) all(!is.na(x)))

  sum(this_na & others_not_na)
})

print(unique_missing_each)

vars <- c("c22", "c29", "a01", "a03", "b08a", "b08b", "b31", "stcog",
          "b2201", "b2202", "b2301", "b2302", "b2303", "b24a1", "b24b1")

included_data <- ceps_g9[complete.cases(ceps_g9[, vars]), ]
excluded_data <- ceps_g9[!complete.cases(ceps_g9[, vars]), ]

included_data$high_aspiration <- included_data$c22 %in% 7:9
```

```
excluded_data$high_aspiration <- excluded_data$c22 %in% 7:9

included_data$SES_numeric <- pmin(included_data$b08a, included_data$b08b, na.rm = TRUE)
excluded_data$SES_numeric <- pmin(excluded_data$b08a, excluded_data$b08b, na.rm = TRUE)

included_data$is_female <- included_data$a01 == 2
excluded_data$is_female <- excluded_data$a01 == 2

included_data$Minority <- ifelse(included_data$a03 == 1, 0, 1)
excluded_data$Minority <- ifelse(excluded_data$a03 == 1, 0, 1)

included_data$parental_expectation <- ifelse(included_data$b31 %in% c(1, 10), 1, included_data$b31)
excluded_data$parental_expectation <- ifelse(excluded_data$b31 %in% c(1, 10), 1, excluded_data$b31)

included_data$parental_factor <- ceps_g9_full$parental_factor[match(rownames(included_data), rownames(ceps_g9_full))]
excluded_data$parental_factor <- ceps_g9_full$parental_factor[match(rownames(excluded_data), rownames(ceps_g9_full))]

included_data$perceived_expectation <- ceps_g9_full$c29[match(rownames(included_data), rownames(ceps_g9_full))]
excluded_data$perceived_expectation <- ceps_g9_full$c29[match(rownames(excluded_data), rownames(ceps_g9_full))]

included_data$teacher_expectation_high <- included_data$perceived_expectation %in% c(1, 2)
excluded_data$teacher_expectation_high <- excluded_data$perceived_expectation %in% c(1, 2)
compare_group_props <- function(var, label) {
  p1 <- mean(included_data[[var]], na.rm = TRUE)
  p2 <- mean(excluded_data[[var]], na.rm = TRUE)
  tbl <- matrix(c(
    sum(included_data[[var]], na.rm = TRUE), sum(!included_data[[var]], na.rm = TRUE),
    sum(excluded_data[[var]], na.rm = TRUE), sum(!excluded_data[[var]], na.rm = TRUE)
  ), nrow = 2, byrow = TRUE)
  chisq <- chisq.test(tbl)
  cat(sprintf(" %s\n", label))
  cat(sprintf("  Included sample:    %.1f%%\n", 100 * p1))
  cat(sprintf("  Excluded sample:    %.1f%%\n", 100 * p2))
  cat(sprintf("  Chi-sq = %.2f, p = %.4f\n\n", chisq$statistic, chisq$p.value))
}
```

```r
compare_group_means <- function(var, label) {
  m1 <- mean(included_data[[var]], na.rm = TRUE)
  m2 <- mean(excluded_data[[var]], na.rm = TRUE)
  ttest <- t.test(included_data[[var]], excluded_data[[var]])
  cat(sprintf(" %s\n", label))
  cat(sprintf("  Included sample mean:    %.2f\n", m1))
  cat(sprintf("  Excluded sample mean:    %.2f\n", m2))
  cat(sprintf("  t = %.2f, p = %.4f\n\n", ttest$statistic, ttest$p.value))
}

compare_group_props("high_aspiration", "% Expecting Bachelor's Degree or Above")
compare_group_props("is_female", "% Female")
compare_group_props("Minority", "% Ethnic Minority")
compare_group_props("teacher_expectation_high", "% Expecting KSH/OHS from Teachers")

compare_group_means("SES_numeric", "Mean SES (Parental Occupation)")
compare_group_means("stcog", "Mean Cognitive Ability")
compare_group_means("parental_expectation", "Mean Parental Expectation")
compare_group_means("parental_factor", "Mean Parental Involvement")

```