Studying Numerical Literacy Across Generations in U.S. Households

Aws Alsaedi, Omar Talib
AI Researchers, University of Toronto
Toronto, ON, Canada
Emails: Aws.saedi@gmail.com, Omar.a.bintalib@gmail.com

***Abstract*—Numerical literacy, the ability to understand and work with numbers, is a foundational skill with long-term implications for economic well-being. Yet little is known about how it persists over time, how it is transmitted across generations, and how it relates to wealth. Using nationally representative, multigenerational data from the Panel Study of Income Dynamics (PSID) and its Child Development Supplement (CDS), we link individuals' childhood numerical literacy to their adult outcomes and their children's skills, forming matched longitudinal and parent–child pairs. Our analysis reveals modest but statistically significant correlations in numerical literacy across time and generations, suggesting that family environments shape quantitative skills and influence wealth accumulation. We apply both statistical and machine learning methods to identify patterns and predictors, providing new insights into how families impact long-term economic outcomes and highlighting the importance of early skill development.**

Project Mentor

Professor Michael Boutros
Email: michael.boutros@utoronto.ca

Group Members and Contributions

- **Aws Alsaedi** — Led project framing and developed the analytical framework from inception. Coordinated and organized all team activities, including full poster design, presentation preparation, and complete project documentation in LaTeX. Managed and structured the GitHub repository for version control and collaboration. Reviewed and refined statistical design strategies, conducted codebook verification, supported variable selection, and guided regression planning. Led major discussions in team meetings, particularly on methodology, interpretation, and strategic direction. Conducted statistical validation and reproducibility checks (seed control, train/test consistency). Drafted the abstract, poster, and final report to align with KAUST/UofT requirements. Oversaw data visualization standards (style, accessibility, color consistency).
- **Omar Talib** — Led the end-to-end data engineering for merging PSID and CDS datasets. Implemented preprocessing, missing-data handling, and exploratory analysis. Designed and optimized the pipeline for large joins across waves (memory-aware chunking). Performed feature engineering to construct socioeconomic indicators and harmonize test-score scales across years. Implemented and compared predictive models (regularized regressions, Gradient Boosting, Random Forest, XGBoost), tracked performance metrics, and produced feature-importance analyses to explain drivers of child scores.

Objectives

- Examine whether numerical literacy persists within individuals over time.
- Analyze whether skills are transmitted from parents to children.
- Link childhood numerical literacy to adult wealth outcomes.
- Identify predictors of intergenerational numerical literacy using statistical and ML methods.

Current Progress and Milestones Achieved

- Constructed matched parent–child datasets from multiple PSID and CDS waves.

- Standardized test scores and merged variables across survey years.
- Completed exploratory data analysis, data cleaning, and variable selection.
- Produced poster, presentation, and this final report integrating statistical and ML methods.

Data and Methods

- **PSID:** 6-question numerical literacy test (0–6 score) for parents.
- **CDS:** standardized child assessment (0–120 score) for children.
- Linked each child's test scores with their parents' test scores and their own future outcomes.
- Built matched parent–child datasets and longitudinal follow-ups.
- Applied regression and machine learning models (including XGBoost) to identify predictors.

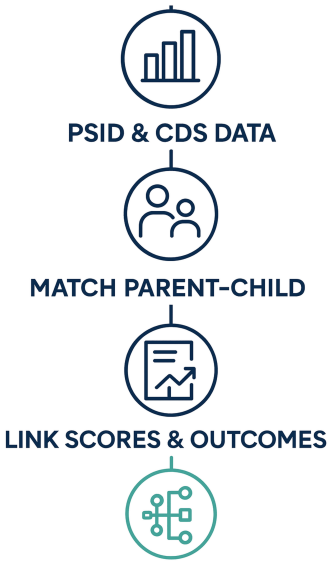

Fig. 1: Research pipeline: (1) collect PSID and CDS data, (2) match parent–child pairs, (3) link scores to adult outcomes, and (4) apply statistical and ML models.

Pipeline narrative: We first extracted and cleaned PSID/CDS data to ensure comparability across waves. We then created a matched parent–child panel, enabling us to trace how skills evolve and transfer. Next, we linked childhood scores to later-life outcomes to study persistence and economic relevance. Finally, we estimated both statistical models and ML models to uncover predictors and validate robustness.

Key Results

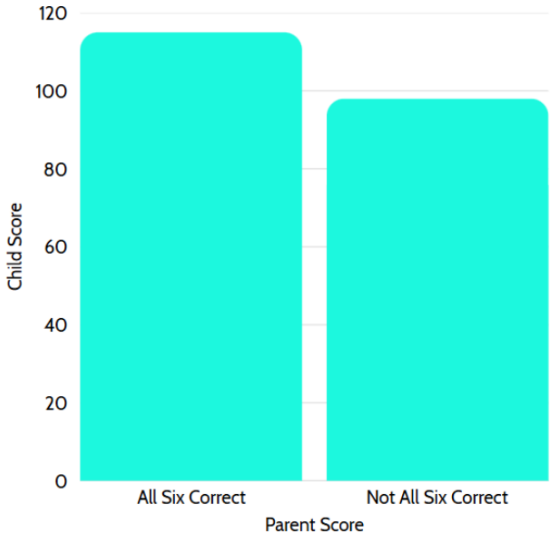

Fig. 2: Children of parents who answered all six numeracy questions correctly score higher on average.

Interpretation (why it matters): This result provides a direct signal of intergenerational skill transmission. When parents demonstrate perfect numeracy, their children's average scores are substantially higher, even before controlling for other factors. For our research question, it establishes that parental numeracy is not merely an individual attribute; it is associated with measurable differences in the next generation's performance, implying that early investments at the household level can propagate forward.

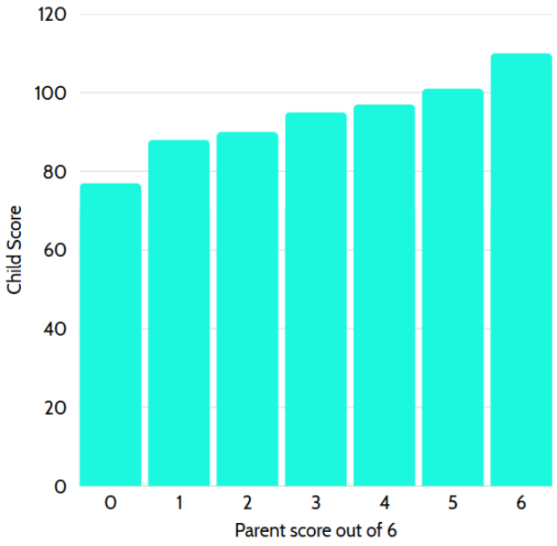

Fig. 3: Child score rises steadily with the parent's score (0–6 correct), indicating intergenerational transmission.

Interpretation (what it shows): The graded relationship reveals that transmission is continuous rather than binary. Each incremental improvement in the parent's score is associated with a corresponding improvement in the child's score. This pattern supports our hypothesis that families shape quantitative skill formation in a cumulative way and motivates models that capture smooth, monotonic effects rather than simple "high/low" categories.

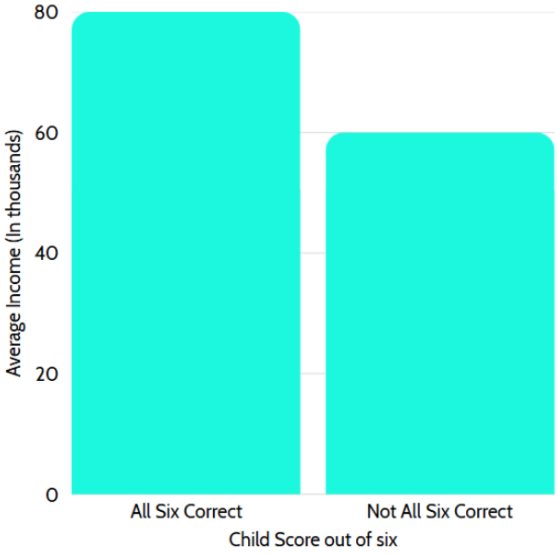

Fig. 4: Children who answered all six questions correctly have higher average adult income.

Interpretation (economic relevance): This figure links early numeracy to later-life economic outcomes. Higher childhood numeracy predicts higher adult income, suggesting that the benefits of these skills extend beyond school performance and into financial security. For policymakers and educators, this underscores that strengthening early numeracy can be a lever for improving long-run economic mobility.

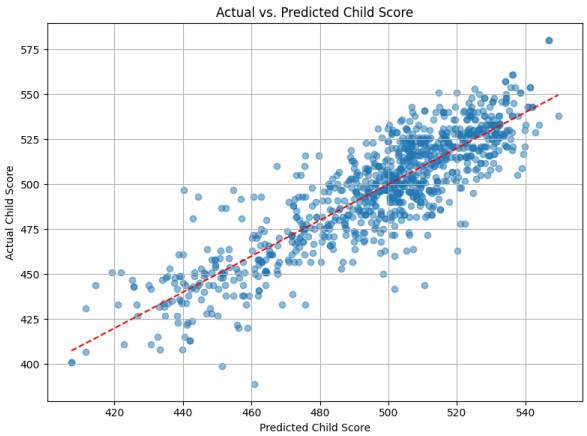

Fig. 5: Machine learning predictions vs. actual child scores. Close alignment along the dashed line demonstrates strong predictive performance.

Interpretation (model validity): The tight clustering around the 45-degree line indicates that our feature set—covering family background and socioeconomic indicators—captures meaningful variation in child scores. This provides external validation for our substantive claims: the same variables that theory suggests (parental income, expenditure patterns, educational expectations) are also empirically powerful in prediction, not just explanation.

Discussion

Numerical literacy shows persistence over time. Evidence of intergenerational transmission highlights the strong family influence on numerical literacy development. Early numerical skills correlate with adult wealth outcomes, emphasizing the importance of early education. Machine learning analysis identified key predictors (parental income, expenditure patterns, education expectations), reinforcing the role of socioeconomic factors in numerical literacy.

Challenges and Obstacles Encountered

- Large and complex PSID/CDS datasets required careful merging and principled treatment of missing data.
- Codebook ambiguities made variable harmonization and verification essential.
- Standardizing test-score scales across survey years required methodological adjustments and sensitivity checks.
- Balancing statistical methods and ML approaches involved trade-offs between interpretability and predictive power.

Conclusion

Our findings confirm that numerical literacy endures across time and is passed on between generations. Stronger early-life numeracy is consistently linked to better adult wealth outcomes, including higher income and improved financial security. The family environment plays a decisive role: parental income, household spending habits, and parents'

educational expectations all emerge as key predictors of a child's long-term outcomes. These insights indicate that improving early numeracy is not only an educational priority but also an economic strategy tied to intergenerational wealth dynamics. By combining traditional regression methods with modern machine learning, we substantially improve predictive accuracy, making it possible to identify at-risk groups and target interventions more effectively. Overall, nurturing quantitative skills in childhood is a pragmatic pathway to building financial resilience and narrowing inequality across generations.

Future Directions

- Extend analysis to additional PSID waves for longer follow-up and more robust cohort comparisons.
- Enrich socioeconomic covariates (occupation, neighborhood context, parental wealth).
- Explore advanced ML models and causal inference strategies to strengthen interpretability.
- Evaluate policy strategies for boosting early numeracy (school/home interventions, parental guidance).
- Compare results across countries to assess generalizability of the transmission patterns.

Tools and Resources

We used **Python**, **PyTorch**, and **GitHub**; core data are from the University of Michigan's **PSID/CDS**. The workflow included data preparation, exploratory analysis, regression modeling, and machine learning predictions.

Acknowledgements

We gratefully acknowledge: Prof. Michael Boutros for guidance and mentorship; KAUST Academy for funding and research support; University of Toronto – Data Sciences Institute for resources and collaboration; **NTDP** and **ALAGIL** for their support and partnership; and our colleagues and peers for valuable feedback.

Resources & Contacts