Exploring the Link between Cognitive Abilities and Data Science Skills using Alternative Raven's Progressive Matrices*

Farshid Farzan[1,*,†], Hasan Mashrique[1,†] and Andrew M. Olney[1,†]

[1] *University of Memphis.*

Abstract

This study explored the relationship between performance on an alternative Raven's Progressive Matrices (aRPM) test and data science problem solving abilities, hypothesizing a strong link to relational thinking. In the experiment, 31 undergraduates engaged in a 2.5-hour session, including a worked example and four problem solving tasks, followed by data science problems. Our regression analysis confirmed that aRPM scores significantly predict data science problem solving performance, effectively capturing a moderate to strong variance in posttest out-comes. Additionally, aRPM was more predictive of performance than experience in related subjects. An investigation of model fairness indicated that the model may underestimate problem solving performance for male and non-white sub-groups. The findings of this study highlight the potential of using aRPM in traditional or intelligent tutoring systems for data science education to enhance personalization. aRPM can predict initial learning outcomes and identify students who may need additional support. However, further research is necessary to validate aRPM's effectiveness across different demographic groups.

Keywords

Cognitive Assessment, Raven's Progressive Matrices, Data Science Education, Problem Solving

1. Introduction

The association between cognitive ability and educational attainment is well known [1]. Indeed, multiple studies have found that the effect is bidirectional, with cognitive ability affecting educational attainment and long-term education improving cognitive ability [2, 3]. The relationship between cognitive ability and educational outcomes extends to learning programming, with application to the failure and dropout rates among programming students [4]. Raven's Progressive Matrices (RPM) [5], often referred to as a measure of fluid intelligence, has recently been proposed as the cognitive test most predictive of programming ability [4, 6]. Previous research highlights cognitive skills as crucial for programming success, but their impact on the broader field of data science remains underexplored. Given the interdisciplinary nature of data science, which encompasses a wide range of skills including programming, statis-tical analysis, and machine learning [7, 8], understanding the role of cognitive abilities in data science education presents an intriguing area for further exploration. Donoho [8] notes that data science is an evolving discipline that extends beyond traditional statistics by incorporating data analysis, modeling, and scientific inquiry. Exploring its cognitive foundations can enhance our understanding of what drives expertise in this dynamic field.

Originally intended as a broader study on learning gains in data science problem solving, high attrition led us to focus on the predictive role of alternative Raven's Progressive Matrices (aRPM) on data science problem solving (DSPS). This paper presents multiple regression analyses to explore three questions: whether aRPM scores predict DSPS, their predictive value after adjusting for experience in related fields, and their consistency across demographic groups to assess fairness.

* *9th Educational Data Mining in Computer Science Education (CSEDM) Workshop, July 20, 2025, Palermo, Sicily, Italy*
* Corresponding author.
† These authors contributed equally.
ffarzan@memphis.edu (F. Farzan); hmshrque@memphis.edu (H. Mashrique); aolney@memphis.edu (A. M. Olney)
0009-0004-7382-6863 (F. Farzan); 0009-0007-7842-3600 (H. Mashrique); 0000-0003-4204-6667 (A. M. Olney)

1.1. Predictive Power of Cognitive Diagnostics in Educational Success

Cognitive ability assessments effectively predict academic performance and chart learning progressions through data-driven analysis of attribute relationships [10, 11]. Other studies from psychometric [12] and neurocognitive assessment perspectives [6, 13] have also shown that cognitive abilities are key indicators of success in STEM [14]. However, the challenges associated with learning programming have captured researchers' attention, particularly due to historically high failure and dropout rates. To address this issue, researchers have explored the impact of cognitive abilities on programming outcomes [15, 16] and highlighted the necessity of cognitive abilities or functions for both learning and problem solving and showed programming also places demands on these cognitive faculties [17]. A broad variety of cognitive tests with different complexity have been used to evaluate the cognitive abilities [12, 13], but the RPM test, renowned for its non-verbal nature and emphasis on evaluating problem solving abilities devoid of prior knowledge and practice effects has emerged as a leading test for measuring programming ability [18]. Recent studies have validated the use of cognitive tests to enhance educational program designs in programming and mathematics [19]. RPM, designed to assess general intelligence, is critical in psychometric evaluations due to their ability to measure perceptual and analytic cognitive processes [20]. The accuracy and consistency of these tests, crucial for their application in educational and psychological contexts [21, 22].

1.2. The Role of Individual Differences

Individual differences in cognitive abilities significantly influence learning outcomes, highlighting the importance of tailored practice and engagement in domain-specific tasks [23]. As cognitive abilities like fluid intelligence decline with age, crystallized intelligence, which is based on accumulated knowledge, tends to remain stable or even increase, supporting competent functioning in various contexts [24, 25]. Additionally, working memory plays a critical role in cognitive development and education, with its effectiveness influenced by age-related strategies that adapt over time [26]. Previous research has identified gender-based differences in some cognitive processes and fundamental skills like problem solving [27, 28]. While this research is not settled, particularly given the multi-dimensional nature of the gender effect [29], it does suggest that findings relating cognitive ability to skill should consider individual differences, and if that ignores this consideration, it could potentially disadvantage some groups. These findings highlight the importance of developing educational pro-grams that adapt to the diverse learning and cognitive needs throughout an individual's life.

Tailoring instruction based on cognitive profiles, such as aRPM scores, can be implemented by human instructors or AI-driven educational systems. Adaptive learning technologies, including intelligent tutoring systems, have shown promise in personalizing instruction to match learner needs and abilities [30, 31]. Leveraging such systems enables scalable, data-driven scaffolding that adjusts to individual learners in real time, enhancing engagement and learning outcomes.

2. Method

The study utilized a 2x2x2 factorial design with a pretest/posttest setup, investigating the effects of programming (blocks vs. code), problem-solving explanation, and sub-goal-labeled materials on data science learning. Participants (N=31) were undergraduate psychology students recruited from an urban university in the southern United States, including 11 males and 20 females, with a racial composition spanning white (n=15) and non-white (n=16) categories. Participants' mean age was 22.93 years (SD= 8.60). Participants were randomly assigned to one of eight conditions in a 2x2x2 design varying in programming style (blocks/code), explanation prompt, and subgoal labeling. This structure was originally intended to explore instructional effects. However, due to the small sample per cell , we did not analyze condition effects separately. The distribution of participants was approximately even across conditions. Participants received course credit but were not otherwise compensated. The study was conducted online using Chrome on participant computers. It employed several measures: attitudinal surveys about learning data science, mathematical concepts, and statistical variable types, along with demographic questions and data science problem solving tests. These tests assessed procedural coding knowledge, data manipulation skills, and code tracing abilities, focusing on conceptual under-standing rather than complete problem resolution. The posttest comprised computational thinking questions designed without the use of coding [32, 33]. Participants used JupyterLab [34] for tasks that progressed from direct application to complex problem solving with minimal guidance. All activities and instructions were conducted through Qualtrics, with video tutorials for coding and problem-solving [34, 35]. Participants, after being randomized into eight groups, filled initial surveys assessing their foundational knowledge, followed by engaging with progressively challenging tasks through interactive notebooks. Posttest involved problem solving, the System Usability Scale [36], a cognitive load survey [37], an adapted version of Raven's Progressive Matrices, aRPM [5, 38], and demographic queries about programming, statistics, and data science experience (Figure 1). The aRPM used in this study is an 18-item, open-access version of Raven's Progressive Matrices designed to mirror the structure and difficulty of the original test while aligning with the appropriate timing of the study. Though not identical to the original version, it retains the core non-verbal reasoning features and our internal consistency analysis supports its reliability. Using the 18-question aRPM, a free version of the proprietary RPM, improves accessibility and practicality, facilitating wider use in educational settings without cost barriers. The 2.5-hour study concluded with a thorough debrief on its aims and structure. Because the planned study had high attrition such that it would require several years of data collection to complete, we focus our analysis on the relationship between aRPM and DSPS, a preregistered hypothesis . Therefore, our analysis collapses across all conditions to examine the relationship between tests and aRPM.

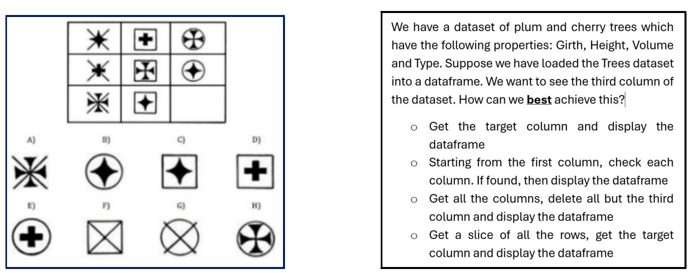

Figure 1: Sample questions of aRPM (left, answer E) and posttest (DSPS) (right)

3. Results

Since aRPM lacks a published psychometric evaluation, key metrics, including mean scores, internal consistency, and item-to-scale correlation, were examined. The mean correctness of .42, high internal consistency was confirmed by a Cronbach's alpha of .81, and an item correlation of .19 indicated low redundancy among items. These metrics are within published ranges of standard RPM [5]. Table 1 presents the descriptive statistics for scores and years of experience in programming, data science, and statistics among the participants. A Variance Inflation Factor (VIF) analysis indicated no significant multicollinearity among aRPM and experience predictors for DSPS.

Table 1
Mean values and standard deviations for test scores and years of experience

DSPS		aRPM		Stat. Exp.		Prog. Exp.		Data Sci. Exp.	
M	SD	M	SD	M	SD	M	SD	M	SD
.48	.29	.42	.23	.56	.64	.29	.68	.29	.90

3.1. Preregistered Model: Predicting DSPS with aRPM

A linear regression analysis was conducted to examine the extent to which aRPM scores, and other probable factors predict DSPS performance. The models was preregistered as part of the study's hypotheses. The results indicated that aRPM scores significantly predicted posttest performance, $B = .78$, $SE = .19$, $t(29) = 4.19$, $p < .001$, 95% CI [0.397, 1.156], such that each correctly answered question on aRPM predicts a 4.3% increase in DSPS score. The model accounted for 37.7% of the variance in posttest scores, supporting the hypothesis that aRPM, as a measure of cognitive ability, significantly predicts DSPS scores.

3.2. Exploratory Model: aRPM Prediction, Controlling for Experience

A second exploratory regression analysis was conducted to examine whether aRPM predicts DSPS beyond prior experience in statistics, programming, and data science. The extended model with these experience predictors remained significant, explaining 50% of the variance in posttest scores ($p < .001$). aRPM scores continued to be a strong and significant predictor of posttest performance, $B = .84$, $SE = .18$, $t(26) = 4.62$, $p < .001$, 95% CI [0.469, 1.218], such that each correctly answered question on aRPM predicts a 4.7% increase in DSPS score. Programming experience was also a significant predictor of DSPS, $B = .18$, $SE = .07$, $t(26) = 2.45$, $p = .021$, 95% CI [0.028, 0.324], suggesting that each additional year of programming experience increased posttest performance by 18 %. However, statistics experience ($p = .247$) and data science experience ($p = .179$) were not significant predictors. In terms of effect, four correct questions on aRPM are equivalent to one year of programming experience, and programming experience explains only an additional 12.3% of the variance compared to 37.7% explained by aRPM alone.

3.3. Model Fairness: Predicting Posttest Performance Across Subgroups

We conducted an exploratory analysis to see if our base model, which uses aRPM scores to predict DSPS, performs consistently across demographic groups (gender and race). This aimed to verify the model's fairness in reflecting diverse individual scores.

Table2

Subgroups means and standard deviations for test scores and years of experience

Subgroup	N	DSPS		aRPM		Stat. Exp.		Prog. Exp.		Data Sci. Exp.	
		M	SD	M	SD	M	SD	M	SD	M	SD
Female-Nonwhite	12	.46	.3	.42	.23	.67	.58	.25	.40	.42	1.16
Male-Nonwhite	4	.68	.38	.54	.28	.25	.50	0	0	0	0
Female-White	9	.40	.27	.35	.19	.61	.70	.56	1.13	.44	1.01
Male-White	6	.52	.27	.44	.28	.50	.84	.17	.41	0	0

For male participants, a simple linear regression analysis revealed that RPM scores were a strong predictor of posttest performance, $B = .94$, $p = .003$, 95% CI [0.421 ,1.464], a stronger effect than found in the base model ($B = .78$). This model suggests that the relationship between RPM scores and posttest performance is underestimated by the base model for male participants. In contrast, the regression model for female participants showed that RPM scores, while still significant, had a weaker predictive power, $B = .61$, $p = .036$, 95% CI [0.043 ,1.167], compared to the base model. This model suggests that the relationship between RPM scores and posttest performance is overestimated by the base model for female participants. Regarding racial subgroups, the regression model for white participants indicated a marginally significant prediction of posttest performance by RPM scores, $B = .61$, $p = .0506$ weaker than the base model. This model suggests that the relationship between RPM scores and posttest performance is overestimated by the base model for white participants.

Conversely, for non-white participants, RPM scores showed a strong and significant effect on posttest performance, B = .90, p = .004, 95% CI [0.338 ,1.464], exceeding the base model's prediction. This model explained indicating that the base model underestimates the strength of the RPM score's predictive power on DSPS score for non-white participants.

4. Discussion

This research showed that aRPM scores was a significant predictor of data science posttest performance, demonstrating 37.7% of the variance in the regression model for posttest scores, underscoring a moderate-to-strong effect of aRPM. Carpenter et al. [39] suggest that Raven's test performance predicts ability on new cognitive problems. This study found that aRPM predicts early-stage data science problem solving in participants new to data science. As learners gain experience, aRPM's predictive value may lessen, though it remains an effective early indicator.

Our findings indicate that only programming experience, not statistics or data science knowledge, predicted DSPS. Given the study's use of block and traditional programming, this influence of programming on DSPS is expected. Notably, a year of programming experience had an impact equivalent to four correct aRPM responses. Our analyses investigating subgroup model fairness suggest the potential for the model to both overestimate and underestimate performance for different demographic groups. These results are concerning and should be considered in terms of scale. The base model predicts a 4.3% increase in DSPS for each correct aRPM question. In the subgroup analyses, the predicted increase ranged from 3.4-5.2%, i.e. approximately 1% different in the worst cases. The difference could accumulate to 18% if all aRPM questions were correctly answered. Future research should investigate these relationships more closely with a larger sample size to confirm these estimates.

Our findings enrich our understanding of the interplay between instructional strategies, individual differences, and cognitive capabilities in the context of data science education among undergraduate psychology students. By demonstrating the importance of cognitive abilities in predicting educational outcomes, the study supports refined educational interventions that act as bridges, connecting sides of the zone of proximal development [40]. Using the insights from Raven's matrices, educators can effectively scaffold learning experiences to not only meet students where they are but also extend their reach, seamlessly connecting the phases of learning that lie just within and just beyond their immediate grasp. This approach, whether implemented through intelligent or traditional adaptive systems, ensures that every student receives tailored support to enhance their data science skills and understanding, regardless of their starting level.

As our study's limitations, the use of small sample sizes, especially in subgroup analyses, may limit the generalizability and statistical power to detect significant effects accurately. Secondly, our methodological choice to collapse data across the factorial design could mask variations in posttest scores attributable to different conditions, potentially obscuring how specific interventions may influence outcomes.

Many participants did not complete our study, so our results only include those who completed aRPM towards the end of the study session. Therefore, it is possible that non completers may have a different relationship between posttest scores and aRPM than completers.

Additionally, our study experienced differential attrition, such that participants in the block programming condition were less likely to complete the study than participants in the coding condition. Therefore, it is possible that blocks condition participants may have a different relationship between posttest scores and aRPM than coding condition participants, but we do not have enough data to make this comparison.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant 1918751.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT (GPT-4) in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] M. Lövden, L. Fratiglioni, M. M. Glymour, U. Lindenberger, E. M. Tucker-Drob, Education and cognitive functioning across the life span, Psychol. Sci. Public Interest **21** (2020) 6–41.

[2] T. Falch, S. Sandgren Massih, The effect of education on cognitive ability, Econ. Inq. **49** (2011) 838–856.

[3] S. J. Ritchie, T. C. Bates, I. J. Deary, Is education associated with improvements in general cognitive ability, or in specific skills?, Dev. Psychol. **51** (2015) 573–582.

[4] A. A. Farghaly, P. M. El-Kafrawy, Exploring the use of cognitive tests to predict programming performance: A systematic literature review, in: Proceedings of the 31st International Conference on Computer Theory and Applications (ICCTA), IEEE, 2021, pp. 40–48.

[5] J. Raven, J. Raven, Uses and Abuses of Intelligence: Studies Advancing Spearman and Raven's Quest for Non-Arbitrary Metrics, Royal Fireworks Press, 2008.

[6] C. S. Prat, T. M. Madhyastha, M. J. Mottarella, C. H. Kuo, Relating natural language aptitude to individual differences in learning programming languages, Sci. Rep. **10** (2020).

[7] R. D. De Veaux, M. Agarwal, M. Averett, B. S. Baumer, A. Bray, T. C. Bressoud, L. Bryant, L. Z. Cheng, A. Francis, R. Gould, Curriculum guidelines for undergraduate programs in data science, Annu. Rev. Stat. Appl. **4** (2017) 15–30.

[8] D. Donoho, 50 years of data science, J. Comput. Graph. Stat. **26** (2017) 745–766.

[9] G. S. Halford, W. H. Wilson, S. Phillips, Relational knowledge: The foundation of higher cognition, Trends Cogn. Sci. **14** (2010) 497–505.

[10] C.-J. Yen, T. R. Konold, P. A. McDermott, Does learning behavior augment cognitive ability as an indicator of academic achievement?, J. Sch. Psychol. **42** (2004) 157–169.

[11] X. Wu, R. Wu, Y. Zhang, D. Arthur, H. H. Chang, Research on construction method of learning paths and learning progressions based on cognitive diagnosis assessment, Assess. Educ. **28** (2021) 657–675.

[12] D. M. Kurland, R. D. Pea, C. Clement, R. Mawby, A study of the development of programming ability and thinking skills in high school students, in: Studying the Novice Programmer, Psychology Press, London, 1989, pp. 83–112.

[13] B. Helmlinger, M. Sommer, M. Feldhammer-Kahr, G. Wood, M. E. Arendasy, S. E. Kober, Programming experience associated with neural efficiency during figural reasoning, Sci. Rep. **10** (2020).

[14] M. Berkowitz, E. Stern, Which cognitive abilities make the difference? Predicting academic achievements in advanced STEM studies, J. Intell. **6** (2018) 48.

[15] S. H. Fletcher, Cognitive Abilities and Computer Programming, unpublished manuscript, 1984.

[16] M. C. Linn, J. Dalbey, Cognitive consequences of programming instruction: Instruction, access, and ability, Educ. Psychol. **20** (1985) 191–206.

[17] A. P. Ambrósio, F. M. Costa, L. Almeida, A. Franco, J. Macedo, Identifying cognitive abilities to improve CS1 outcome, in: 2011 Frontiers in Education Conference, IEEE, Rapid City, SD, 2011, pp. F3G-1–F3G-7.

[18] J. Raven, The Raven's progressive matrices: Change and stability over culture and time, Cogn. Psychol. **41** (2000) 1–48.

[19] A. Ghosh, L. Malva, A. Singla, Analyzing-evaluating-creating: Assessing computational thinking and problem solving in visual programming domains, in: Proceedings of the 55th

ACM Technical Symposium on Computer Science Education, vol. 1, ACM, Portland, OR, 2024, pp. 387–393.

[20] N. J. Mackintosh, E. Bennett, What do Raven's Matrices measure? An analysis in terms of sex differences, Intelligence **33** (2005) 663–674.

[21] H. R. Burke, Raven's Progressive Matrices: A review and critical evaluation, J. Genet. Psychol. **93** (1958) 199–228.

[22] Y. Yang, M. Kunda, Computational models of solving Raven's Progressive Matrices: A comprehensive introduction, arXiv preprint arXiv:2302.04238 (2023).

[23] K. A. Ericsson, N. Charness, Expert performance: Its structure and acquisition, Am. Psychol. **49** (1994) 725–747.

[24] T. Salthouse, Consequences of age-related cognitive declines, Annu. Rev. Psychol. **63** (2012) 201–226.

[25] M. E. Beier, P. L. Ackerman, Age, ability, and the role of prior knowledge on the acquisition of new domain knowledge, Psychol. Aging **20** (2005) 341–355.

[26] N. Cowan, Working memory underpins cognitive development, learning, and education, Educ. Psychol. Rev. **26** (2014) 197–223.

[27] L. Beckwith, M. Burnett, Gender: An important factor in end-user programming environments?, in: IEEE Symp. on Visual Languages – Human Centric Computing, IEEE, Rome, 2004, pp. 107–114.

[28] M. Burnett, S. D. Fleming, S. Iqbal, G. Venolia, V. Rajaram, U. Farooq, V. Grigoreanu, M. Czerwinski, Gender differences and programming environments: Across programming populations, in: Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, ACM, Bolzano, 2010, pp. 1–10.

[29] S. Beyer, K. Rynes, J. Perrault, K. Hay, S. Haller, Gender differences in computer science students, in: Proceedings of the 34th SIGCSE Technical Symposium on Computer Science Education, ACM, Reno, NV, 2003, pp. 49–53.

[30] A. Adair, M. S. Pedro, J. Gobert, E. Segan, Real-time AI-driven assessment and scaffolding that improves students' mathematical modeling during science investigations, in: International Conference on Artificial Intelligence in Education, Springer, Cham, 2023, pp. 202–216.

[31] S. Durrani, D. S. Durrani, Intelligent tutoring systems and cognitive abilities, in: Proceedings of the Graduate Colloquium on Computer Sciences (GCCS), 2010.

[32] M. Román-González, J.-C. Pérez-González, J. Moreno-León, G. Robles, Can computational talent be detected? Predictive validity of the Computational Thinking Test, Int. J. Child Comput. Interact. **18** (2018) 47–58.

[33] M. Román-González, J.-C. Pérez-González, C. Jiménez-Fernández, Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test, Comput. Hum. Behav. **72** (2017) 678–691.

[34] GitHub, JupyterLab computational environment. https://github.com/jupyterlab/jupyterlab (accessed 2024/03/20).

[35] A. M. Olney, S. D. Fleming, JupyterLab extensions for blocks programming, self-explanations, and HTML injection, in: Joint Proceedings of the Workshops at the 14th International Conference on Educational Data Mining, EDM, Virtual Event, 2021.

[36] J. Brooke, SUS: A quick and dirty usability scale, in: Usability Evaluation in Industry, 1996, pp. 189–194.