A Large-Scale Study on the Effects of Motivation on Academic Test Performance

Michael D. Kickmeier-Rust and Katharina Richter
St.Gallen University of Teacher Education, St.Gallen, Switzerland

michael.kickmeier@phsg.ch
katharina.richter@phsg.ch

Abstract: Motivation plays a crucial role in academic performance. In educational research, particularly in technology-enhanced learning, a strong correlation between motivation and performance is often assumed. However, existing studies present inconsistent findings, with some showing only low to moderate correlations between self-reported motivation and actual academic achievement. This study investigates the relationship between intrinsic motivation, restedness, and test performance, including response times, in large-scale assessments in Mathematics. These tests were administered to an extensive sample of over 26,000 and 27,000 students, respectively. The findings indicate significant but relatively small correlations between motivational and performance variables. Notably, post-test motivation proves to be a stronger predictor of test outcomes than pre-test motivation. This suggests a complex dynamic between intrinsic and extrinsic motivational factors, implying that extrinsic incentives may have a more substantial impact on actual test performance. Stepwise linear regression further differentiates the influences, illustrating the roles of personal background, prior academic achievements, and motivation as latent factors contributing to test scores. Although the overall explained variance remains moderate, this study provides valuable insights into a rarely explored aspect of motivation—its assessment after a test and its predictive capacity. While much of the existing literature emphasizes motivation before a test, these findings suggest that motivation measured after an academic task may better capture students' effort and engagement. Additionally, the study raises methodological concerns regarding the reliability of self-reported motivation, highlighting potential issues with survey validity, such as response tendencies like straight-lining. The results have significant implications for educational assessment practices and research methodologies. They underscore the need for a holistic approach to understanding academic performance and the diverse motivational processes involved.

Keywords: Intrinsic Motivation, Restedness, Test Performance, Secondary School Level, Representative Sample

1. Introduction

Human behavior, including academic performance, is shaped by various factors, with motivation playing a central role. As the driving force behind goal pursuit, motivation lacks a universal definition (Bandhu et al., 2024). However, several theoretical models provide insight, such as Arousal Theory (Reisenzein, 2017), Attribution Theory (Weiner, 1972), the ARCS model (Keller, 1987), and Self-Determination Theory (Deci & Ryan, 2008). These frameworks emphasize self-efficacy, attribution styles, and the distinction between intrinsic and extrinsic motivation. In learning contexts, motivation influences engagement and knowledge acquisition (Svinicki & Vogler, 2012), making it a crucial factor in educational success. It also impacts instructional design, didactics, and the development of educational technologies (Urhahne & Wijnia, 2023). Literature on instructional design highlights the importance of intrinsic motivation, notably advanced by Malone and Lepper's (1987) work on making learning enjoyable. While motivation is often assumed to be a key predictor of academic performance, particularly in technology-based education like gamification and serious games, its relationship with achievement remains complex. Empirical evidence is inconsistent, hindered by small sample sizes and methodological limitations, raising concerns about generalizability. Furthermore, the individualized nature of motivational factors in education remains a subject of debate. Motivation is a key factor in academic performance, but its influence varies across different contexts. A critical distinction is made between intrinsic and extrinsic motivation, as well as low-stakes and high-stakes scenarios. High-stakes assessments, such as exams, typically elicit strong extrinsic motivation, driven by rewards, grades, or consequences. Conversely, in low-stakes contexts, such as formative assessments and adaptive learning technologies, intrinsic motivation plays a more central role. Understanding these dynamics is crucial for educators seeking to optimize student learning experiences (Liu, Bridgeman & Adler, 2012).

1.1 The Impact of Motivation on Academic Performance

Intrinsic motivation, characterized by genuine interest in learning, fosters deeper engagement and knowledge retention. In contrast, extrinsic motivation enhances performance, particularly in structured assessments (Cerasoli et al., 2014). While intrinsic motivation encourages persistence and meaningful learning, extrinsic incentives are often better predictors of the quantity of effort exerted. Yu et al. (2020) found that educational games enhance motivation and engagement, suggesting that enjoyable activities promote better learning outcomes. However, when extrinsic rewards are linked directly to performance, intrinsic motivation tends to decline, a trend observed by Akhtar and Firdiyanti (2023). Afzal et al. (2010) emphasizes the role of educators in fostering student interest, which enhances intrinsic motivation. The combination of gamification and instructional design strategies can increase motivation and engagement, especially in low-stakes assessments where the perceived importance of the test may be diminished (Shell et al., 2019). The relationship between motivation and cognitive learning strategies has been widely researched. A strong correlation exists between motivation and student performance, particularly in settings that encourage deeper engagement with content (Ramírez-Arellano et al., 2018). Motivation also influences students' ability to adapt to academic environments, affecting their resilience and overall learning experience (Lo et al., 2022). Wang (2022) highlights that intrinsic motivation plays a critical role in students' performance in online learning environments, reinforcing the idea that motivation enhances focus and persistence in academic pursuits. Research by Chan et al. (2015) supports these findings, demonstrating that intrinsically motivated students report higher satisfaction with online learning, which improves engagement and academic success. This underscores the need for educators to cultivate motivation through interactive and engaging educational methodologies.

1.2 Nuanced Perspectives on Motivation and Performance

While many studies find a moderate to strong correlation between intrinsic motivation and performance, recent analyses introduce a more complex perspective. Silm, Pedaste, and Tährt (2020) found that self-reported effort showed a moderate correlation with academic achievement ($r = .33$), whereas response time effort (RTE) had a much stronger correlation ($r = .72$). Akhtar and Firdiyanti (2023) further explored test motivation in low-stakes assessments, comparing Self-Reported Effort (SRE) and RTE. Their findings reveal that while these indicators capture different aspects of motivation, RTE is a stronger predictor of test outcomes ($\beta = 0.67$ vs. $\beta = 0.12$ for SRE), suggesting that behavioral measures of effort may provide a more reliable gauge of academic performance. Xu et al. (2020) confirmed these observations in a large sample of 8th graders, finding only moderate correlations between self-reported intrinsic motivation and mathematics test performance ($r = .37$ and $r = .29$). In adults, Ulum and Küçükaydın (2024) found only a weak correlation ($r = .16$) between motivation and math achievements. The research base on motivation in primary and secondary education is smaller than for higher education. Silm, Pedaste, and Tährt (2020) report that test-taking effort correlates more strongly with performance among university students than among school-aged learners. This suggests that academic experience and maturity influence how motivation translates into achievement. These findings highlight the complex interplay between intrinsic and extrinsic motivation and its impact on academic success. While intrinsic motivation fosters engagement and long-term retention, extrinsic incentives drive immediate performance, particularly in structured assessments. However, response time effort (RTE) may serve as a more reliable behavioral indicator of exertion than self-reported motivation, offering valuable insights into student engagement.

A key source of data on student motivation and academic performance in secondary education comes from the OECD PISA studies. Kriegbaum, Jansen, and Spinath (2015) found that self-efficacy and self-concept significantly predict mathematics performance, even in low-stakes tests with minimal direct consequences. However, achievement motivation emerged as only a moderating variable, influencing test success through intelligence and prior achievements. Their analysis identified self-efficacy as the strongest motivational predictor. Wang et al. (2023) conducted a meta-analysis of PISA-related studies, revealing mixed and sometimes contradictory findings on the relationship between intrinsic motivation and mathematics achievement. The authors failed to establish a clear causal pattern of intrinsic motivation affecting academic outcomes, emphasizing the role of analytical methods and other influential factors such as test anxiety and attitudes toward school. Similarly, Mews and Pöge (2019) reported that intrinsic motivation has only limited explanatory power compared to self-efficacy.

A review of existing literature suggests that the role of intrinsic motivation in academic performance may often be overestimated. Its impact is intertwined with various individual factors, such as self-efficacy, attitudes toward subjects and school, and extrinsic motivators. While intrinsic motivation is often highlighted as a driver of academic success, studies indicate that self-efficacy plays a more influential role in determining student achievement. This study aims to clarify the role of intrinsic motivation in low-stakes academic tests by analyzing a large sample of secondary school students. Through rigorous statistical methods and comprehensive datasets, the research seeks to overcome previous study limitations and offer new insights into how motivation influences test performance. A central question explored is whether post-test motivation serves as a more reliable predictor of academic performance than pre-test motivation, thereby challenging traditional methods of assessing motivational factors in educational research.

To analyze the relationship between self-reported intrinsic motivation, restedness, and academic performance, this study employs a robust methodological approach to examine how these psychological factors interact with test outcomes. The research is structured around three key questions:

RQ1: Does self-reported motivation prior to the test affect test scores and test duration?

RQ2: Does self-reported restedness prior to the test affect test scores and test duration?

RQ3: Is self-reported motivation after the test correlated with test scores and test duration?

These questions are crucial for understanding students' psychological states in low-stakes testing environments, where external pressures are minimized, and intrinsic motivation becomes a key driver of performance. By examining the interplay between motivation, restedness, and test results, the study aims to provide insights that inform educational strategies and assessment practices. By analyzing large-scale data, this research seeks to refine current approaches to motivation assessment and provide reliable insights into how intrinsic motivation functions in low-stakes academic tests. The findings will contribute to discussions on student engagement, mental well-being, and effective assessment techniques, ensuring that academic evaluations account for psychological factors that influence learning outcomes.

2. Method

2.1 Participants

This study analyzed two datasets from a standardization study on German and Math test items. The Math dataset included results from 26,226 students across St. Gallen (11,683) and Zürich (14,543); 48% were male and 52% female. Students were distributed across grades: 35% in 7th, 40% in 8th, and 29% in 9th. Typically, 7th graders are around 12.5 years old. The academic distribution was as follows: 29% attended “Realschule” (lowest achievement level), 49% “Sekundarschule” (intermediate level), and 22% “Gymnasium” (highest level). The German dataset consisted of 27,307 students from St. Gallen (11,820) and Zürich (15,487); 49% male and 51% female. Grade distribution was 34% in 7th, 38% in 8th, and 28% in 9th, with 7th graders averaging 12.5 years of age. The academic distribution mirrored the Math dataset: 29% in “Realschule,” 49% in “Sekundarschule,” and 22% in “Gymnasium.”

2.2 Materials

The data were collected in the 2017 standardization of the Stellwerk test (https://stellwerk.ch), a high-stakes computer adaptive test (CAT; Bock & Gibbons, 2021) using Rasch-scaled (Fischer & Molenaar, 1995) dichotomous items. Widely used in German-speaking Swiss cantons, the test is mandatory in some regions (typically in 8th grade) and optional in others. It assesses Mathematics, German (mother tongue), English, French, and Science (Biology, Physics, Chemistry), with results utilized for educational monitoring and student placement. Employers often require test scores for apprenticeships and job applications. The Stellwerk test selects items dynamically from a large pool, necessitating calibration through large-scale standardization studies to determine item difficulty. Results are reported on a standardized 200-800 scale. The Math test assesses algebra, geometry, and data/probability (Figure 1). Item formats include multiple-choice, sorting, selection tasks, and free-text entries. The standardization process employed test booklets, each containing 28 difficulty-balanced items tailored to grade levels. The test duration averaged 50 minutes, with valid completion times ranging from 12 minutes 30 seconds to 2 hours. Outliers were omitted. Motivation was assessed using brief four-item Likert-scale surveys before and after testing. The pre-test survey included statements like “I will make an effort in doing the test,” while the post-test survey adapted them to past tense. A restedness scale gauged student energy levels, ensuring minimal disruption to testing.

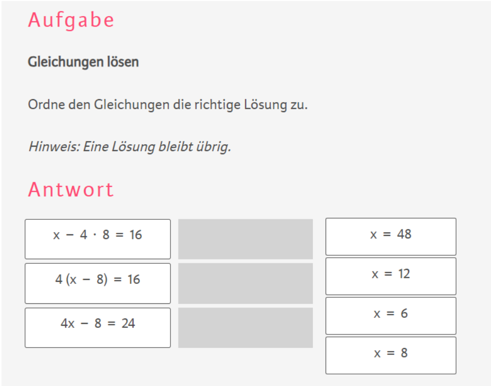


Figure 1: Example test item for Mathematics.

3. Results

For the present paper, we analyzed the data of the Stellwerk test in Mathematics. Prior to analyses, we carried out a data cleaning. Implausible data were omitted from the dataset, that is, to short test times (< 750 sec.) and too long test times (> 7000 sec.). In addition, all incomplete tests were deleted; these are test where students did not work on all tasks (e.g., by clicking "next question" without giving an answer).

Overall, the dependent variables were all normally distributed. The analysis included 26226 cases. The average test score (i.e., correctly solved items) was 9.67 (SD = 5.00, Min = 0, Max = 28), the average test time was 2227.73 seconds (SD = 854.84, Min = 750, Max = 8487), the motivation prior the test (MOT_PRE) was 12.45 (SD = 2.89, Min = 4, Max = 16), the average restedness (ENERGY_PRE) was 9.45 (SD = 1.78, Min = 4, Max = 16), and the average motivation after the test (MOT_POST) was 10.83 (SD = 3.27, Min = 4, Max = 16). The participants, therefore, had an average individual loss in motivation (MOT_DIFF) from prior to after the test of -1.62 (SD = 2.77).

For class types (A, B, C – where A is the type with the highest academic performance) we found the expected differences. As shown in Table 1, the higher the class type the higher were the scores and the time spent on the test. For gender, we did not find substantial differences (Table 1). In terms of motivation (ENERGY_PRE, MOT_PRE, MOT_POST) we did not find differences across class types and gender (Table 2).

Table 1: Descriptive results for test variables for class type and gender (N = 26226).

	Class Type			
	Score		Time	
	M	SD	M	SD
A	11.62	4.54	2346.23	883.09
B	10.22	5.06	2212.09	815.80
C	6.38	3.41	2133.05	869.41
	Gender			
	Score		Time	
	M	SD	M	SD
Male	10.01	5.11	2191.22	845.51
Female	9.49	4.80	2277.99	855.30

Table 2: Descriptive results for motivation variables for class type and gender (N = 26226).

	Class Type					
	ENERGY_PRE		MOT_PRE		MOT_POST	
	M	SD	M	SD	M	SD
A	9.68	1.61	12.24	2.67	10.66	3.00
B	9.63	1.48	12.67	2.65	11.13	3.08
C	9.66	1.39	12.82	2.50	11.04	3.00
	Gender					
	ENERGY_PRE		MOT_PRE		MOT_POST	
	M	SD	M	SD	M	SD
Male	9.68	1.59	12.47	2.75	10.87	3.14
Female	9.62	1.37	12.70	2.50	11.09	2.94

The Pearson correlations between performance and motivation variables (Table 3), however, are rather low. The highest correlation exists between score and motivation after the test (r = .274). It is noteworthy, that MOT_POST has a clearly stronger relationship to scores than MOT_PRE.

Table 3: Correlation table of performance and motivation variables.

	SCORE	TIME	ENERGY_PRE	MOT_PRE	MOT_POST
SCORE	1.000	0.337	0.007	0.194	0.274
TIME	0.337	1.000	0.006	0.207	0.196

For a statistical comparison of variables, we grouped the participants in high and low motivated individuals. For the variables we used the median as the cut-off score (medians: ENERGY_PRE = 10, MOT_PRE = 13, MOT_POST = 11). The average scores for ENERGY_PRE_LOW were 9.53 (SD = 4.94) for ENERGY_PRE_HIGH 9.77 (SD = 5.04), the average scores for MOT_PRE_LOW were 8.77 (SD = 4. 74) for MOT_PRE_HIGH 10.49 (SD = 5.08), and the average scores for MOT_POST_LOW were 8.38 (SD = 4.53) for MOT_POST_HIGH 10.64 (SD = 5.11). This means that the LOW groups had lower scores than the HIGH groups; the effect is highest for MOT_POST. Figure 2 shows the 2-way interaction diagrams.

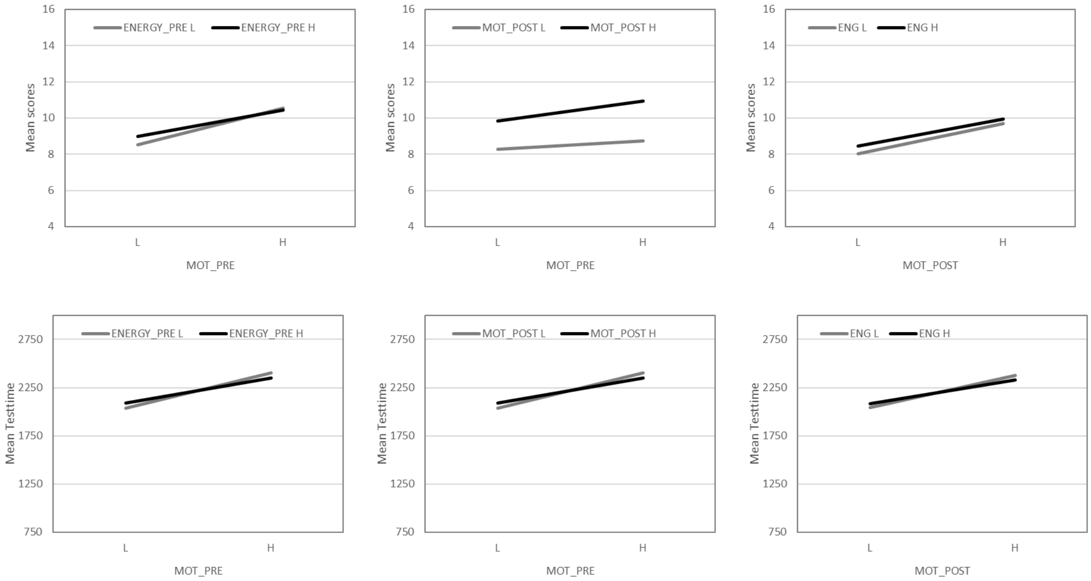


Figure 2: Interactions diagrams of restedness (energy), motivation prior and after the test for scores (top row) and test times (bottom row).

The average test times for ENERGY_PRE_LOW were 2223.34 (SD = 852.21) for ENERGY_PRE_HIGH 2231.05 (SD = 856.82), the average scores for MOT_PRE_LOW were 2068.78 (SD = 825.80) for MOT_PRE_HIGH 2371.65 (SD

= 855.21), and the average scores for MOT_POST_LOW were 2065.36 (SD = 829.40) for MOT_POST_HIGH 2349.35 (SD = 853.43). This means that the LOW groups had lower scores than the HIGH groups; the effect is highest for MOT_POST. Figure 2 shows the 2-way interaction diagrams.

A multivariate analysis of variance (ANOVA) reported a series of statistically significant group differences (Table 4). Given the large sample size, we computed the Cohen's d for the main effects to indicate the effect sizes between the HIGH and LOW groups. For ENERGY_PRE d = .048 for Score and .009 for Time. For MOT_PRE d = .350 for Score and d = .360 for Time. For MOT_POST d = .464 for Score and d = .337 for Time. According to Cohen (1992) these a rather small effect sizes, whereas the strongest effect was found for MOT_POST, meaning the self-reported motivation after the test.

Table 4: Results of a multivariate analysis of variance (ANOVA).

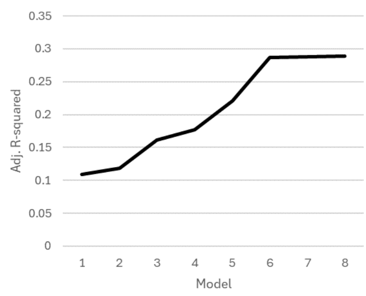
	Variable	SS	df	MS	F	Sig.
EN_PRE_HL	SCORE	3094.35	1	3094.35	131.17	<.001
	TIME	245198198	1	245198198	349.637	<.001
M_PRE_HL	SCORE	61.145	1	61.145	2.592	0.107
	TIME	859670.107	1	859670.107	1.226	0.268
M_POST_HL	SCORE	16767.332	1	16767.332	710.772	<.001
	TIME	147637921	1	147637921	210.522	<.001
EN_PRE_HL * M_PRE_HL	SCORE	268.004	1	268.004	11.361	<.001
	TIME	8474728.37	1	8474728.37	12.084	<.001
EN_PRE_HL * M_POST_HL	SCORE	429.554	1	429.554	18.209	<.001
	TIME	3584592.59	1	3584592.59	5.111	0.024
M_PRE_HL * M_POST_HL	SCORE	11.29	1	11.29	0.479	0.489
	TIME	1778162.44	1	1778162.44	2.536	0.111
EN_PRE_HL * M_PRE_HL * M_POST_HL	SCORE	5.475	1	5.475	0.232	0.63
	TIME	862265.267	1	862265.267	1.23	0.268

For a deeper analyses of the manifest variables, we specified two multiple linear regression models, one without motivation variables and one including them in the model. The results are shown in Table 5. The model without motivation variables resulted in an adjusted R^2 of 0.2212 ($F(5, 13823) = 786.08$, $p < .001$); the model including motivation variables in an adjusted R^2 of 0.2888 ($F(8, 13823) = 703.1$, $p < .001$). There is an increase in explained variance with the motivation variables, however, overall, the variance explained by these predictor variables is unexpectedly low. An ANOVA between the two models in R yielded that the model including motivation variables is a significantly better predictor for test scores ($F(-3) = 438.8$, $p = <.001$). When we investigate the adjusted R^2s upon stepwise inclusion of variables, we see that the variables school grades in mathematics (GRADE_M) and MOT_POST have the strongest explanatory power to predict test scores, followed by the parents' educational level (EDU_IX) and the achievement level of the class (CLASS_TYPE).

4. Discussion

This study examined the impact of self-reported motivation on academic test performance, focusing on standardized test data in Mathematics from the Swiss Stellwerk test. The research explored whether self-reported restedness and motivation—before and after the test—affected test duration and scores. While the hypothesis suggested a clear influence of motivation, statistical significance played only a minor role due to the large sample size. Unexpectedly, correlations between motivation and performance variables were weak. Self-reported restedness showed the smallest effect, followed by motivation before the test, while motivation after the test had the strongest impact. Despite examining several key predictors, the latent and manifest variables accounted for no more than 30% of test score variability, likely due to high variance in performance measures.

Table 5: Results of stepwise regression.

Model*	Adj. R^2	Summary
GRADE_M	**0.1087**	
GRADE_D + GRADE_M	0.1188	
GRADE_D + GRADE_M + EDU_IX	**0.1617**	
GRADE_D + GRADE_M + EDU_IX + LANG	0.1771	
GRADE_D + GRADE_M + EDU_IX + LANG + CLASS_TYPE	**0.2212**	
GRADE_D + GRADE_M + EDU_IX + LANG + CLASS_TYPE + MOT_POST	**0.2868**	
GRADE_D + GRADE_M + EDU_IX + LANG + CLASS_TYPE + MOT_POST + MOT_PRE	0.2875	
GRADE_D + GRADE_M + EDU_IX + LANG + CLASS_TYPE + MOT_POST + MOT_PRE + ENERGY_PRE	0.2888	

* GRADE_M = Mathematics grade in the last report card; GRADE_D = German grade in the last report card; EDU_IX = educational level of the parents; LANG = langue (mother tonge, primary langue, secondary langue, novice); CLASS_TYPE = achievement level of the school class; MOT_PRE = motivation prior to the test, MOT_POST = motivation after the test; ENERGY_PRE = restedness prior to the test.

A potential explanation for the limited predictive power of motivation lies in the context of standardization studies. These studies, conducted periodically in Swiss schools, are seen as burdensome by teachers and students, offering little direct benefit. Furthermore, schools are selected through representative sampling rather than voluntary participation. Unlike OECD PISA studies, Stellwerk standardization tests are administered by regular class teachers rather than external proctors, potentially influencing student commitment. For instance, PISA 2018 showed a correlation of r = .57 between math grades and PISA math scores (Pulkkinen & Rautopuro, 2022), while this study found a lower correlation of r = .25. Additionally, self-report surveys face challenges such as straight-lining—when respondents select uniform responses across Likert-scale items without differentiation.

Straightlining may indicate differences in participant motivation. Positive straightliners (n = 24) did not differ significantly from other respondents in their scores (M = 10.21, SD = 2.52 vs. M = 10.27, SD = 2.39; t(74) = -0.11, p = .912), suggesting consistent attitudes or response patterns. In contrast, negative straightliners (n = 7) scored notably lower (M = 7.86, SD = 2.27; t(57) = -2.38, p = .021), which may reflect low motivation, frustration, or protest behavior. However, interpreting straightlining as a motivation indicator is not straightforward. Without additional evidence—such as response time, quality of open-ended answers, or performance on attention checks—it remains unclear whether the behavior reflects genuine attitudes or systematic bias. Straightlining alone is insufficient for valid motivation assessment. Researchers should therefore combine multiple indicators to better evaluate data quality and participant engagement. While straightlining can serve as a potential signal, it only becomes meaningful when considered alongside other metrics. A nuanced interpretation is essential to avoid misjudging participant intent or data reliability.

This study highlights that academic performance in Mathematics is shaped by multiple factors. The strongest predictors include: (i) school type (e.g., Realschule, Sekundarstufe, Gymnasium, or achievement levels A, B, C), (ii) main language spoken by students, (iii) parental educational background, (iv) self-reported math grades from prior school reports, and (v) intrinsic motivation reported before and after the test. Factors such as test booklet versions, student preferences for Mathematics, and gender showed no substantial effects on performance. Structural equation modeling suggests that performance is primarily influenced by personal background, academic achievements, and motivation. Future research should focus on better ways to evaluate intrinsic motivation and its role in academic effort.

In conclusion, in educational research—particularly within technology-enhanced learning—intrinsic motivation is often regarded as a crucial factor for successful knowledge acquisition, retention, and recall. Game-based learning, gamification, and multimedia learning are frequently cited as methods that enhance intrinsic motivation, thereby improving performance, engagement, and diligence. However, research on the actual effects of intrinsic motivation remains inconclusive, often limited by methodological weaknesses, selective sampling, or small sample sizes. This study analyzed a representative dataset in Mathematics and found surprisingly weak correlations between self-reported motivation and test performance. One clear limitation of the study is that motivation and restedness were assessed using only a few questionnaire items rather than a comprehensive, theory-based scale. Nevertheless, these findings suggest that the role of intrinsic motivation and restedness in education—particularly in technology-enhanced learning—warrants further scrutiny. Additional analyses, including structural equation modeling and an evaluation of the same dataset in German as a first language, yielded similar results, though they are beyond the scope of this paper. A plausible explanation for the lack of correlation is that students may reach a substantial level of effort with only moderate motivation, meaning that higher motivation does not necessarily translate into improved academic performance. Future experimental studies should explore this hypothesis in greater depth to determine under what conditions intrinsic motivation significantly influences learning outcomes.

Ethics Declaration

This study is a re-analysis of existing data sets. The data were originally collected under the supervision and with approval of the cantonal departments of education.

AI Declaration

AI was used to edit, optimize the text body and adjust its length. AI was not used to generate any content and not for surveying the existing literature.

References

Afzal, H., Ali, I., Khan, M. and Hamid, K. (2010) ‘A study of university students’ motivation and its relationship with their academic performance’, *International Journal of Business and Management*, 5(4). Available at: https://doi.org/10.5539/ijbm.v5n4p80.

Akhtar, H. and Firdiyanti, R. (2023) ‘Test-taking motivation and performance: Do self-report and time-based measures of effort reflect the same aspects of test-taking motivation?’, *Learning and Individual Differences*, 102, p. 323. Available at: https://doi.org/10.1016/j.lindif.2023.102323.

Bandhu, D., Mohan, M. M., Nittala, N. A. P., Jadhav, P., Bhadauria, A. and Saxena, K. K. (2024) ‘Theories of motivation: A comprehensive analysis of human behavior drivers’, *Acta Psychologica*, 244, p. 104177.

Bock, R. D. and Gibbons, R. D. (2021) ‘Computerized adaptive testing’, in Bock, R. D. and Gibbons, R. D. *Item response theory* (Chapter 8). Wiley. Available at: https://doi.org/10.1002/9781119716723.ch8.

Botnaru, D., Orvis, J., Langdon, J., Niemiec, C. P. and Landge, S. M. (2021) ‘Predicting final grades in STEM courses: A path analysis of academic motivation and course-related behavior using self-determination theory’, *Learning and Motivation*, 74, p. 101723.

Chan, D. et al. (2015) ‘In-lecture learning motivation predicts students’ motivation, intention, and behaviour for after-lecture learning: examining the trans-contextual model across universities from UK, China, and Pakistan’, *Motivation and Emotion*, 39(6), pp. 908–925. Available at: https://doi.org/10.1007/s11031-015-9506-x.

Cerasoli, C. P., Nicklin, J. M. and Ford, M. T. (2014) ‘Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis’, *Psychological Bulletin*, 140(4), pp. 980–1008. Available at: https://doi.org/10.1037/a0035661.

Cohen, J. (1992) ‘Statistical power analysis’, *Current Directions in Psychological Science*, 1(3), pp. 98–101. Available at: https://doi.org/10.1111/1467-8721.ep10768783.

Deci, E. L. and Ryan, R. M. (2008) ‘Self-determination theory: A macro theory of human motivation, development, and health’, *Canadian Psychology*, 49(3), pp. 182–185.

Fischer, G. H. and Molenaar, I. W. (1995) *Rasch models: Foundations, recent developments, and applications*. Berlin: Springer.

Keller, J. M. (1987) ‘Development and use of the ARCS model of instructional design’, *Journal of Instructional Development*, 10, pp. 2–10.

Kriegbaum, K., Jansen, M. and Spinath, B. (2015) ‘Motivation: A predictor of PISA's mathematical competence beyond intelligence and prior test achievement’, *Learning and Individual Differences*, 43, pp. 140–148. Available at: https://doi.org/10.1016/j.lindif.2015.08.026.

Lo, K., Ngai, G., Chan, S. and Kwan, K. (2022) ‘How students’ motivation and learning experience affect their service-learning outcomes: a structural equation modeling analysis’, *Frontiers in Psychology*, 13. Available at: https://doi.org/10.3389/fpsyg.2022.825902.

Malone, T. and Lepper, M. (1987) ‘Making Learning Fun: A Taxonomy of Intrinsic Motivations for Learning’, in Snow, R. and Farr, M. J. (eds.) *Aptitude, Learning, and Instruction Volume 3: Conative and Affective Process Analyses*. Hillsdale, NJ.

Mews, S. and Pöge, A. (2019) ‘Das Zusammenspiel von Selbstbildern, motivationalen und emotionalen Orientierungen sowie deren Einfluss auf die Mathematikleistung in der PISA-Studie 2012’, *Zeitschrift für Erziehungswissenschaften*, 22, pp. 899–924.

Pulkkinen, J. and Rautopuro, J. (2022) ‘The correspondence between PISA performance and school achievement in Finland’, *International Journal of Educational Research*, 114, p. 102000.

Ramírez-Arellano, A. et al. (2018) 'Factors affecting student learning performance: a causal model in higher blended education', *Journal of Computer Assisted Learning*, 34(6), pp. 807–815. Available at: https://doi.org/10.1111/jcal.12289.

Reisenzein, R. (2017) 'Varieties of cognition-arousal theory', *Emotion Review*, 9, pp. 17–26.

Reuning, K. and Plutzer, E. (2020) 'Valid vs. Invalid Straightlining: The complex relationship between straightlining and data quality', *Survey Research Methods*, 14(5), pp. 439–459.

Silm, G., Pedaste, M. and Täht, K. (2020) 'The relationship between performance and test-taking effort when measured with self-report or time-based instruments: A meta-analytic review', *Educational Research Review*, 31, p. 100335.

Svinicki, M. D. and Vogler, J. S. (2012) 'Motivation and learning: Modern theories', *Encyclopedia of Science Learning*, pp. 2336–2339.

Urhahne, D. and Wijnia, L. (2023) 'Theories of Motivation in Education: an Integrative Framework', *Educational Psychology Review*, 35, p. 45.

Ulum, H. and Küçükaydın, M. A. (2024) 'The interaction of mathematics achievement, academic motivation and social context among Turkish adolescents', *Learning and Motivation*, 88, p. 102050.

Vu, T. et al. (2022) 'Motivation-Achievement Cycles in Learning: a Literature Review and Research Agenda', *Educational Psychology Review*, 34, pp. 39–71. Available at: https://doi.org/10.1007/s10648-021-09616-7.

Wang, L. (2022) 'Student intrinsic motivation for online creative idea generation: mediating effects of student online learning engagement and moderating effects of teacher emotional support', *Frontiers in Psychology*, 13. Available at: https://doi.org/10.3389/fpsyg.2022.954216.

Wang, M. D. and Hau, K.-T. (2024) 'In Likert scale, is ticking options consecutively at two ends equally problematic as ticking in the middle?', *Journal of Official Statistics*, 40(1), pp. 123–145.

Wang, X. S. et al. (2023) 'Factors predicting mathematics achievement in PISA: a systematic review', *Large-scale Assessments in Education*, 11(24).

Weiner, B. (1972) 'Attribution theory, achievement motivation, and the educational process', *Review of Educational Research*, 42, pp. 203–215.

Xu, U. et al. (2020) 'Intrinsic motivation, favorability, time management, and achievement: A cross-lagged panel analysis', *Learning and Motivation*, 72, p. 101677.

Yu, Z., Ming-le, G. and Wang, L. (2020) 'The effect of educational games on learning outcomes, student motivation, engagement and satisfaction', *Journal of Educational Computing Research*, 59(3), pp. 522–546. Available at: https://doi.org/10.1177/0735633120969214.

Zhang, C. and Conrad, F. G. (2014) 'Speeding and straightlining: Survey satisficing and its implications', *Public Opinion Quarterly*, 78(4), pp. 955–968.