RESEP
Research on Socio-Economic Policy

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

forward together
sonke siya phambili
saam vorentoe

RESEP POLICY BRIEF

Department of Economics, Stellenbosch University

Policy Brief No. 05/25

NOVEMBER 2025

This brief was produced as part of the MILAPS project, funded by Optima and utilises Data Driven Districts data

Socio-economic status predicts response patterns in surveys of learner experiences and beliefs

Authors

Glen Takalani, Debra Shepherd & Heleen Hofmeyr

OPTIMA

Policy brief: Socio-economic status predicts response patterns in surveys of learner experiences and beliefs

Date: November 2025

Authors: Glen Takalani, Debra Shepherd and Heleen Hofmeyr (Research on Socio-Economic Policy, Stellenbosch University)

Large-scale international education assessments such as the Trends in Mathematics and Science Study (TIMSS) and the Progress in Reading Literacy Study (PIRLS) contain rich data about learners' experiences of school and beliefs about themselves. These data are often used to compare the experiences and beliefs of learners who attend different types of schools, and inferences are made about the relationship between learner experiences and beliefs and achievement. However, such inter-group comparisons may be misleading. Of chief concern is that learners' schooling experiences may shape how they interpret and respond to questionnaire items about school life and self-beliefs – a phenomenon often described as "reference bias" (Lira et al., 2022). This implies that variation in reported experiences and beliefs may partly reflect differences in interpretation rather than true differences in the underlying constructs. This brief shares findings about the accuracy and reliability with which learner experiences and beliefs are measured in TIMSS and PIRLS data for South Africa.[1] We present evidence that learners attending different types of schools (proxied for by socio-economic status) interpret the TIMSS and PIRLS survey items that ask about their experiences and beliefs in systematically different ways. Based on this evidence, we caution against making inter-group comparisons of learner-reported experiences and beliefs in highly unequal societies such as South Africa.

1 CONTEXT

Across South Africa, policymakers and researchers increasingly rely on large-scale assessments such as TIMSS and PIRLS to understand education system performance to guide interventions in teaching and learning. These assessments include questionnaires that ask learners about their confidence, motivation, and sense of belonging, constructs that are often used to explain differences in achievement (e.g. Hofmeyr (2022)). Yet South Africa's education system remains

[1] The findings summarized here are based on results presented in our working paper, Takalani, G., Shepherd, D. & Hofmeyr, H. 2025. "Do Survey Items Function Equivalently in Diverse Contexts? Evidence from the TIMSS and PIRLS in South Africa". *Resep Working Papers*.

deeply unequal: learners' day-to-day experiences differ sharply by socio-economic status (Böhmer & Wills, 2025). In such a context, assuming that all learners interpret and respond to survey questions in the same way is problematic. This raises a fundamental question: how can we measure learners' experiences and beliefs meaningfully when their reference points, and the realities of schooling they reflect, are so vastly different?

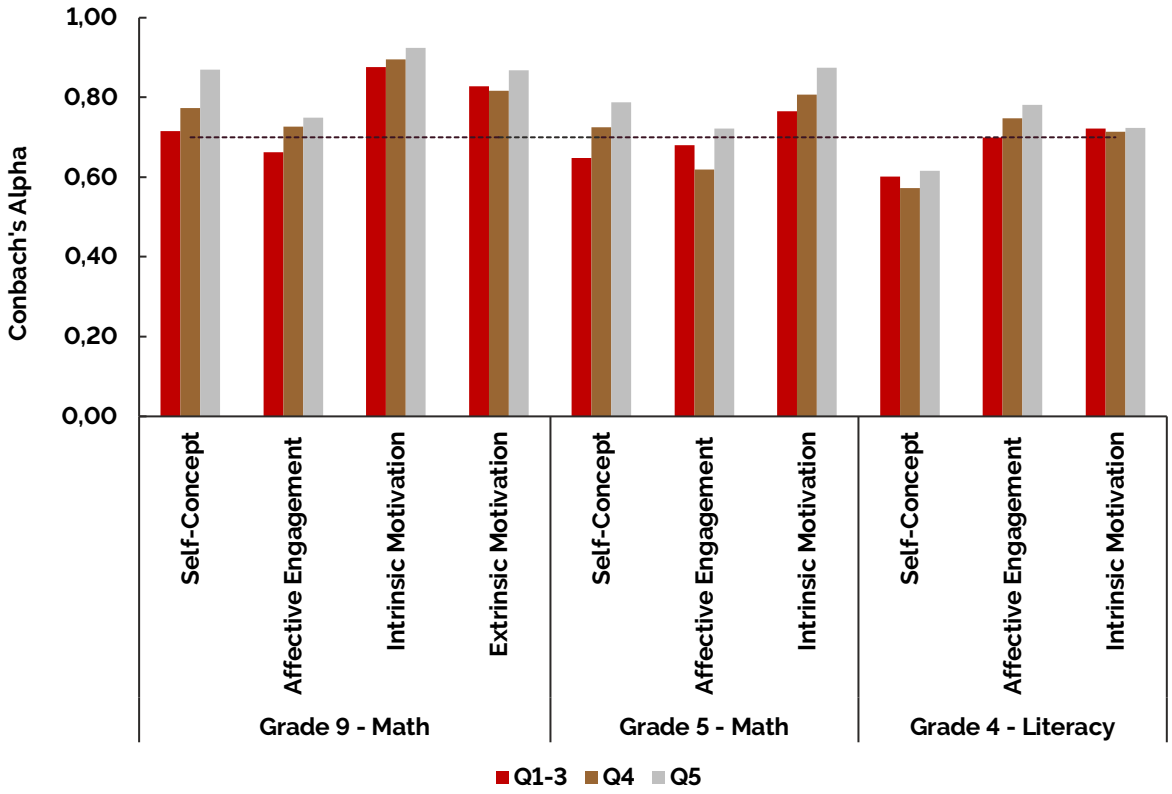
2 KEY FINDINGS

To understand whether self-reported measures from international assessments truly reflect what learners think and feel, we examined both the reliability and the comparability of these measures across different schooling contexts in South Africa. Using data from TIMSS and PIRLS, the analysis first tested whether the sets of questions that make up each self-reported construct (such as those measuring learner confidence or motivation) are internally consistent within groups of schools serving learners from different socioeconomic backgrounds. We then examined whether individual survey items (statements like "I am good at mathematics" or "I like being in school") carry the same meaning across groups. In other words, the study asked not only whether these scales work reliably within each group of learners, but also whether they can be compared fairly between learners in Quintile 1 (the poorest) and Quintile 5 (the wealthiest) schools. Together, these analyses probe the question: can we meaningfully measure and compare learners' reported experiences and beliefs in a system as unequal as South Africa's?

2.1 Scale reliability is not achieved in the poorest 60% of schools

Across most self-reported constructs (such as learners' confidence in mathematics, their enjoyment of reading, and their sense of belonging at school), the internal consistency of the survey scales differs sharply by school wealth. As shown in Figure 1, in wealthier schools (particularly Quintile 5 schools), Cronbach's alpha values exceed the accepted reliability threshold of 0.7. This indicates that the items within each scale work together to measure a coherent concept. In contrast, these values fall below 0.7 in the poorest 60% of schools, suggesting that learners' responses to the items that make up a given scale are not aligned in a consistent way. In other words, when learners in under-resourced schools answer questions intended to measure a construct like "maths confidence," their responses to related items vary less predictably. This implies that the instrument does not capture a stable or well-defined construct in these contexts. This pattern shows that the measurement tools themselves are least reliable for the learners whose experiences policymakers most need to understand.

Figure 1: Reliability scores by Grade and school quintile



Notes: the horizontal dashed line at 0.7 indicates minimum acceptable reliability.

2.2 Individual survey items do not carry the same meaning across school contexts

Beyond problems of overall scale reliability, the analysis shows that individual survey items often function differently for learners in different school contexts. In statistical terms, these items fail tests of measurement invariance, meaning that the same question does not measure the same underlying idea across groups. For example, when asked to agree or disagree with statements such as "I like being in school" or "I am good at solving difficult mathematics problems," learners in poorer schools appear to interpret these statements differently from those in wealthier schools (see Table 1). In under-resourced schools, where curriculum coverage tends to be lower (Hoadley, 2016), "difficult problems" may refer to simpler material than it does in wealthier schools. Similarly, "being proud of one's school" carries different connotations when school facilities and resources diverge starkly. These differences mean that even when learners share the same true level of confidence or engagement, their responses on the survey can look systematically different, undermining the validity of cross-school comparisons.

Table 1: Examples of survey items that function differently across school contexts

Dataset	Construct	Non-Invariant Items	Interpretation
TIMSS Grade 5	Affective engagement	"I like being in school"; "I am proud to go to this school"	Learners interpret school pride and belonging differently depending on school context.
TIMSS Grade 5	Intrinsic value of mathematics	"I learn many interesting things in mathematics"	Differences likely reflect unequal exposure to rich mathematics content and learning experiences.
TIMSS Grade 9	Mathematics self-concept	"I am good at solving difficult mathematics problems"	What counts as a 'difficult problem' varies with curriculum coverage and instructional complexity.
TIMSS Grade 9	Utility value of mathematics	"I would like a job that involves using mathematics"; "My parents think it is important that I do well in mathematics"	Socioeconomic and cultural factors shape how learners view mathematics as a career pathway.

2.3 Observed non-invariance reflects well-documented social comparison influences on self-perception

Research on social comparison processes shows that learners form self-perceptions primarily by comparing themselves to immediate peers rather than absolute standards (Marsh & Parker, 1984; Hattie, 2009). Over and above the curriculum coverage and instructional complexity of mathematics content, 'being good at difficult mathematics problems' is interpreted relative to what classmates can do, which would make the same response reflect different underlying competence levels across schools with different achievement distributions. Additionally, Bourdieu and Passeron's (1990) concept of cultural capital could explain why items about career aspirations and parental expectations function differently. Affluent students often possess the cultural and social capital that makes certain careers visible and achievable (Archer et al., 2014), while lower-SES students possibly interpret the same items through the awareness of structural barriers (Branson et al., 2024). Stereotype threat research (Steele. 1997) further shows that lower-SES students are aware of negative societal expectations about their group's academic capabilities, potentially leading them to underrate themselves even at equivalent ability levels (Croizet & Claire, 1998). Together, these mechanisms demonstrate that survey items tap into different psychological and social realities across school contexts, not merely different levels of the same construct.

3 RECOMMENDATIONS

The findings highlight a central challenge for evidence-based education policy in South Africa: when learners' contexts differ so profoundly, the tools used to measure their experiences and beliefs must be designed and interpreted with care.

1. Treat learner self-reports as diagnostic tools, not comparative metrics. Until the underlying measurement issues are addressed, policymakers and researchers should avoid using TIMSS, PIRLS, or similar survey data to compare learner experiences or beliefs across socio-economic groups. Instead, these data can be used within comparable groups of schools to identify local patterns of engagement, confidence, or belonging that may inform targeted interventions.

2. Strengthen the validity and fairness of survey instruments. Education authorities should invest in the local adaptation and validation of learner questionnaires to ensure that items function *equivalently across socio-economic contexts*. Our analysis shows clear evidence of non-measurement invariance across SES groups, meaning that learners with different backgrounds may interpret or respond to items differently – even when they have similar underlying abilities or attitudes. Validation processes should therefore explicitly compare item functioning across SES groups, including through interviews with learners, careful translation and back-translation where needed, and routine psychometric testing for reliability and measurement invariance prior to reporting or interpreting results. Building internal psychometric capacity within research institutions, the DBE, and provincial assessment teams is essential to achieving this.

3. Build transparency into reporting and policy use. Research that draws on learner survey data should clearly communicate any evidence of non-measurement invariance, particularly across socio-economic groups, so that users understand when comparisons between groups may not be valid. Transparent reporting of these limitations – including how invariance was tested and what the results imply for interpretation – will help prevent overconfident or misleading conclusions and ensure that policy decisions are grounded in evidence that is appropriately qualified.

4 REFERENCES

Archer, L., DeWitt, J. & Willis, B., 2014. Adolescent boys' science aspirations: Masculinity, capital, and power. *Journal of Research in Science Teaching*, 51(1), pp. 1-30. https://doi.org/10.1002/tea.21122

Böhmer, B. & Wills, G. 2025. COVID-19 and inequality in reading outcomes in South Africa. *Large-scale Assessments in Education*. https://doi.org/10.1186/s40536-025-00256-9

Bourdieu, P. & Passeron, J. -C., 1990. Reproduction in Education, Society and Culture. New Delhi: *Sage Publications*.

Croizet, J-C., & Claire, T., 1998. Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and social Psychology Bulletin*, 24(6), pp. 588-594. https://doi.org/10.1177/0146167298246003

Hattie, J., 2009. Visible learning: A synthesis of over 800 meta-analyses relating to achievement. London: *Routledge*.

Hoadley, U. (2016). *A review of the research literature on teaching and learning in the Foundation Phase in South Africa.* Stellenbosch University.

Hofmeyr, H. 2022. Why do girls do better? Unpacking South Africa's gender gap in PIRLS and TIMSS. *International Journal of Educational Development*. 94: 102648. https://doi.org/10.1016/j.ijedudev.2022.102648

Lira, B., O'Brien, J.M., Peña, P.A., Galla, B.M., D'Mello, S., Yeager, D.S., Defnet, A., Kautz, T., Munkacsy, K., Duckworth, A.L. (2022). Large studies reveal how reference bias limits policy applications of self-report measures. *Scientific Reports*. 12: 19189. https://doi.org/10.1038/s41598-022-23373-9

Marsh, H. W. & Parker, J. W., 1984. Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well*?. Journal of Personality and Social Psychology*, 47(1), pp. 213-231. https://doi.org/10.1037/0022-3514.47.1.213

Steele, C. M., 1997. A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), pp. 613-629. https://doi.org/10.1037/0003-066X.52.6.613