2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #53322

Advancing Growth Mindset Research in Education: A Psychometric Validation of the Growth Mindset Scale

Dr. Jing Li, University of Georgia

Jing Li is currently a 5th-year PhD student in the Quantitative Methodology program at the University of Georgia. Her research interests center on educational measurement, Rasch measurement theory, explanatory item response modeling (EIRM), and applied psychometrics, with a particular focus on evaluating surveys and educational assessments. Jing's research aims to promote the fairness of measurement and to apply appropriate methodologies to analyze data. Her objective is to ensure that data interpretations are both accurate and meaningful and allow the data to tell a story.

Chan Aristella Lu, University of Georgia

Chan Aristella Lu is a senior doctoral researcher at the University of Georgia working at the intersection of education policy, artificial intelligence, and quantitative methods.

Cheng Tang, The University of Georgia
Georgia W. Hodges, University of Georgia

Advancing Growth Mindset Research in Education: A Psychometric Validation of the Growth Mindset Scale

Abstract

Empirical and Full Promoting a growth mindset enables students to recognize that learning is possible for everyone and engaging with challenging tasks can help strengthen their cognitive abilities, particularly in engineering education [1], [2]. Therefore, reliable and valid measurement of growth mindset is essential for evaluating interventions, understanding student motivation, and identifying factors that influence academic achievement. This study aims to validate a 23-item growth mindset scale by conducting a psychometric analysis using Rasch measurement theory (RMT) with a sample of elementary students in the Southeastern United States. Specifically, we will evaluate evidence of construct validity by examining item fit, reliability, dimensionality, and differential item functioning (DIF) across gender and grade levels to assess the measurement properties of the scale and ensure its fair use across diverse student groups. Findings of this study provide empirical evidence that the use and interpretation of scores from the Growth Mindset Scale are fair and valid and further contribute to advancing research and practice in fostering growth mindsets among students.

Keywords: Growth mindset, Validation, Rasch measurement theory, Differential item functioning (DIF)

Introduction

Growth Mindset and Learning in STEM and Engineering Contexts

A growth mindset is the belief that abilities and intelligence can be developed through effort. It has been widely acknowledged as a critical factor influencing learning and achievement in STEM and engineering contexts [3]. Research has shown that students with growth mindset beliefs are more likely to persist when facing challenges, engage in problem solving, and view errors as opportunities for learning. All these characteristics are essential components of engineering education [4], [5]. For example, one quasi-experimental study in elementary classrooms found that serious educational gameplay produced learning gains comparable to hands-on science activities, highlighting the value of engaging, challenge-based learning experiences early in STEM [6]. In engineering and STEM classrooms where tasks often involve complex reasoning and learners struggle while solving complex problems, a growth mindset has been associated with greater engagement, resilience, and long-term participation [7], [8]. As a result, the growth mindset has become an important concept in initiatives to increase participation and support equitable learning experiences in engineering education.

Developmental Importance of Measuring Growth Mindset in Upper Elementary Students

During the critical developmental stage of elementary school, students' academic self-beliefs, motivations, and attitudes toward STEM begin to consolidate [9], [10]. Students acquire early perceptions about their abilities in science, math, and problem-solving at this phase, which have an impact on their later course selection, identity development, and persistence in STEM fields. According to research on engineering identity, students' motivation to participate in difficult STEM-related activities is shaped by their perceptions about learning and competence that developed before middle school [11]. Therefore, it is important to measure growth mindset accurately during upper elementary years in order to uncover early belief patterns and inform interventions aimed at supporting positive learning trajectories before students engage with formal engineering courses.

Challenges in Measuring Growth Mindset in Young Learners

Even though growth mindset scales are widely used, there are a number of difficulties when assessing this concept in younger students. The cognitive development, reading comprehension, and response styles of elementary students differ from those of older adolescents, which further impacts how they understand and react to survey items [12]. This makes it especially important to use a validated, developmentally appropriate measurement tool for elementary learners rather than assuming that instruments designed for older students will function equivalently in this age group.

Furthermore, attitudinal measures given to younger individuals may contain items that function differently across developmental stages or demographic groups, which often show lower sensitivity at the extremes of the latent trait continuum. Without careful psychometric evaluation, scale scores could fail to detect meaningful differences in interpretation, and further lead to potential threats to validity and fairness when comparing groups or assessing interventions.

Advantages of Rasch Measurement Theory for Evaluating the Growth Mindset Scales

Rasch measurement theory [13] offers a rigorous framework for addressing these challenges by evaluating the growth mindset scale at the item level rather than relying solely on total scores. Rasch models estimate person ability and item difficulty on a common scale, allowing researchers to examine item functioning, measurement precision, and targeting across the latent continuum. Importantly, the assessment of differential item functioning (DIF) under the Rasch framework provides direct evidence of whether items function equivalently across gender or grade level. Rasch measurement theory is particularly well-suited for evaluating growth mindset instruments intended for use in diverse and developmental student populations since it emphasizes invariant measurement and construct validity.

Purpose of the Study

The purpose of the present study was to evaluate the psychometric properties of a 23-item growth mindset scale among upper elementary students using Rasch measurement theory. In addition, differential item functioning (DIF) analyses were conducted across gender and grade level to evaluate the fairness and measurement invariance of the scale for group comparisons. The following questions were addressed in this study:

1. To what extent does the 23-item growth mindset scale demonstrate sound psychometric properties when evaluated using Rasch measurement theory among upper elementary students?
2. Do items on the growth mindset scale function equivalently across gender and grade levels and support fair group comparisons?

Methodology

Participants

This study includes a sample of elementary students from the Southeastern United States. Table I summarizes the demographic characteristics of the sample (N=179). The sample includes 93 male students (55%) and 86 female students (45%). In terms of grade level, 98 students (52%) are in 4th-grade, and 81 students (48%) are in 5th-grade.

TABLE I
Sample descriptive statistics (N = 179)

Demographic Characteristics	**N**	Proportion
Gender		
Male	93	0.55
Female	86	0.45
Grade		
4th-grade	98	0.52
5th-grade	81	0.48

23-item Growth Mindset Scale

The growth mindset scale used in this study includes 23 items, which were drawn from prior research on implicit theories of intelligence, motivational processes, and classroom-based assessments of beliefs like effort, feedback, and learning [14], [15]. Throughout the remainder of this study, this instrument is referred to as the *23-item growth mindset scale*. Rather than assuming a fixed a priori scale structure, the items were treated as a candidate item pool and evaluated empirically using Rasch measurement theory.

Model

Rasch measurement theory was developed by Georg Rasch as a scientific approach to measurement [13]. The key idea is that the probability of endorsing an item depends on the difference between a person's underlying level on the trait and how difficult the item is. The traditional Rasch model for dichotomous data can be written as:

$$\log(\frac{P_{ni}}{1 - P_{ni}}) = \theta_n - \delta_i \tag{1}$$

In which P_{ni} is the probability of person n endorsing item i, and the left side of the equation is the log-odds of endorsing an item. On the right side is the difference between a person's ability (θ_n) and item difficulty (δ_i).

Analytical Procedure

Data were analyzed using the Rasch model with the *Winsteps* program [16]. A dichotomous Rasch model was fitted to responses from 179 participants on the 23-item growth mindset scale. Overall model fit and measurement quality were evaluated using person and item fit statistics (infit and outfit mean square errors), separation reliability indices, and variance explained by the Rasch measures to assess evidence of unidimensionality. Differential item functioning (DIF) across gender (male vs. female) and grade level (4th vs. 5th grade) was assessed using the Rasch-Welch procedure, with DIF contrasts interpreted as differences in item difficulty between groups.

Results

Rasch Analysis

We summarized the dichotomous Rasch analysis results for 179 participants and 23 items. The mean person measure was 0.05 logits ($SD = 0.34$), and the mean item difficulty was centered at 0 logits ($SD = 0.86$). Overall model fit was acceptable: average infit MSE was 1.01 for both person ($SD = 0.52$) and items ($SD = 0.14$), and average outfit MSE was 1.01 for person ($SD = 0.58$) and items ($SD = 0.13$), indicating close correspondence between observed responses and model expectation.

Measurement precision differed for persons versus items. Person separation reliability was 0.44, indicating limited differentiation among respondents along the latent trait continuum. Internal consistency was assessed using Cronbach's alpha ($\alpha = .69$), suggesting moderate and generally acceptable reliability. In contrast, item separation reliability was very high (0.99), demonstrating a well-defined item hierarchy. The Rasch measures accounted for 38.1% of total variance, exceeding the commonly used 20% guideline for supporting unidimensionality [16].

Fig. 1 presents the Wright map showing the joint distribution of person ability estimates and item difficulty parameters on the same logit scale. Person distribution is centered near the mean of the item difficulty, indicating reasonable targeting of the scale to the sample. Most respondents are clustered around the mid-range of the latent trait (around 0 logits), suggesting that the scale is most informative for individuals with a moderate level of growth mindset. Item difficulty range indicates adequate coverage of the latent growth mindset continuum, though minor gaps at the extremes suggest reduced precision for individuals with very low or very high growth mindsets. Several items are located at the higher end of the scale (e.g., items 18 and 20), indicating that endorsement of these items requires a higher level of growth mindset. Items such as item 9 appear at the lower end of the continuum, reflecting lower endorsement thresholds.

We found item difficulty estimates ranged from -1.9 to 1.76 logits. Item 9 (-1.9 logits) and item 13 (-0.92 logits) are the simplest, whereas item 8 (1.76 logits), 20 (1.57 logits), and 7 (1.11 logits) are the hardest to endorse. Standard errors of the item estimates range from 0.13 to 0.29, suggesting adequate precision in the estimation of item locations. Item fit statistics indicated acceptable model-data fit across all items, with infit MSE ranging from 0.83 to 1.31, and outfit MSE ranging from 0.82 to 1.31. Using Wright and Linacre's [17] guidelines, all items fell within the recommended range of 0.5 to 1.5, supporting the adequacy of the item set for Rasch measurement and providing further evidence of the scale's internal consistency and construct representation.

Fig. 1. Wright map.

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    2                +
                     |
                     |T I8
                     |  I20
                     |
                     |
                     |
                     |  I7
    1                +  I14
                 .#  |S I21
                 .# T|  I11
                 .#  |
             .#####  |  I16
                .## S|
            .######  |  I1     I22
           ########  |
    0      .####### M+M I12    I15    I4
       .###########  |
             .##### S|  I17
               .###  |  I5     I6
                ###  |  I10    I2     I3
                  # T|  I19
                  .  |  I23
                     |S I13
   -1                +  I18
                     |
                     |
                     |
                     |
                     |
                     |T
                     |  I9
   -2                +
               <less>|<freq>
 EACH "#" IS 3: EACH "." IS 1 TO 2
```

Differential Item Functioning

Gender Table II presents the results of the DIF analysis across gender using the Rasch-Welch procedure (Rasch-Welch t-test) [18]. DIF contrast represents male minus female difficulty: positive values indicate items that were more difficult for females, and negative values indicate items that were more difficult for males. Most items demonstrate negligible DIF, with non-significant Rasch-Welch statistics, suggesting that the majority of items functioned equivalently across gender. Five items were found with significant DIF across gender. Items 7 and 8 showed significant negative DIF contrasts (item 7: $DIF = -0.46, p = .019$; item 8: $DIF = -0.64, p = .013$), suggesting that male students found them more challenging than female students. However, item 13 and item 19 demonstrated significant positive DIF (item 13: $DIF = 0.43, p = .038$; item 19: $DIF = 0.42, p = .026$), indicating that female students found them more difficult than male students. Furthermore, male students are less likely to agree that effort is required beyond intelligence, according to item 20's negative DIF ($DIF = -0.48, p = .038$). Appendix A displays the content of the DIF items across gender.

Grade Table III presents the results of the DIF analysis across grade level (4th-grade minus 5th-grade) using the Rasch-Welch procedure. Most of the items functioned similarly across grade levels. Six items showed significant DIF across grades. Items 4, 7, 8, and 16 demonstrated significant negative DIF contrasts (item 4: $DIF = -0.34, p = .045$; item 7: $DIF = -0.41, p = .038$; item 8: $DIF = -0.61, p = .023$; item 16: $DIF = -0.54, p = .002$), suggesting 4th-grade students found these items more difficult than 5th-grade students. These items reflect beliefs that personal characteristics are fixed (item 4), sports ability is innate (item 7), science is easier for boys than girls (item 8), and intelligence cannot be changed (item 16). In contrast, items 18 and 22 have significant positive DIF (item 18: $DIF = 0.49, p = .020;$ item 22: $DIF = 0.35, p = .037$). It appears that 5th-grade students have more difficulty endorsing these two items. This could reflect growing social awareness and classroom experience across grade levels. Appendix B displays the content of the DIF items across grades.

Although statistically significant, the magnitude of DIF contrasts ranged from approximately 0.34 to 0.61 logits, indicating small to moderate DIF [16]. Overall, results support essential measurement invariance across grades and gender, while suggesting that a small set of items, those related to fixed beliefs and social collaborative experiences, may need additional review in further work.

TABLE II
Differential item functioning across gender

					Rasch-Welch		
Item	**Female**	**Male**	**DIF Contrast**	**Joint SE**	***t***	***d.f.***	***Prob.***
1	0.22	0.22	0	0.17	0	174	1.000
2	-0.49	-0.56	-0.06	0.18	-0.35	174	0.726
3	-0.64	-0.48	0.17	0.18	0.92	172	0.359
4	0.02	0.02	0	0.17	0	173	1.000
5	-0.46	-0.38	0.08	0.18	0.44	174	0.664
6	-0.51	-0.31	0.2	0.18	1.14	173	0.258
7	**1.37**	**0.91**	**-0.46**	**0.19**	**-2.38**	**164**	**0.019**
8	**2.13**	**1.49**	**-0.64**	**0.25**	**-2.51**	**156**	**0.013**
9	-2.13	-1.7	0.43	0.28	1.51	168	0.132
10	-0.38	-0.57	-0.19	0.18	-1.05	173	0.293
11	0.83	0.72	-0.11	0.17	-0.62	171	0.534
12	-0.08	0.03	0.11	0.17	0.64	171	0.526
13	**-1.15**	**-0.72**	**0.43**	**0.21**	**2.09**	**166**	**0.038**
14	1.04	0.92	-0.12	0.18	-0.64	172	0.526
15	-0.01	-0.05	-0.05	0.17	-0.28	174	0.782
16	0.48	0.43	-0.05	0.17	-0.28	174	0.776
17	-0.28	-0.28	0	0.17	0	173	1.000
18	-1.12	-0.87	0.25	0.21	1.22	172	0.224
19	**-0.83**	**-0.41**	**0.42**	**0.19**	**2.24**	**169**	**0.026**
20	**1.84**	**1.36**	**-0.48**	**0.23**	**-2.09**	**163**	**0.038**
21	0.95	0.69	-0.26	0.18	-1.5	172	0.135
22	0.18	0.32	0.14	0.17	0.87	174	0.386
23	-0.83	-0.74	0.09	0.19	0.45	174	0.654

Note. DIF contrast is male minus female.

TABLE III
Differential item functioning across grades

Item	4th-grade	5th-grade	DIF Contrast	Joint SE	Rasch-Welch		
					t	*d.f.*	*Prob.*
1	0.08	0.38	-0.29	0.17	-1.76	170	0.081
2	-0.4	-0.67	0.27	0.18	1.49	169	0.138
3	-0.42	-0.71	0.28	0.18	1.53	168	0.128
4	**-0.14**	**0.20**	**-0.34**	**0.17**	**-2.02**	**171**	**0.045**
5	-0.3	-0.56	0.27	0.18	1.5	169	0.136
6	-0.46	-0.35	-0.11	0.18	-0.63	172	0.528
7	**0.94**	**1.36**	**-0.41**	**0.2**	**-2.1**	**154**	**0.038**
8	**1.54**	**2.15**	**-0.61**	**0.27**	**-2.3**	**138**	**0.023**
9	-1.86	-1.95	0.09	0.28	0.33	170	0.745
10	-0.48	-0.48	0	0.18	0	171	1.000
11	0.77	0.75	0.03	0.17	0.15	165	0.881
12	-0.07	0.04	-0.11	0.17	-0.68	168	0.500
13	-0.92	-0.92	0	0.2	0	167	1.000
14	1.12	0.8	0.32	0.18	1.74	170	0.084
15	0.03	-0.1	0.13	0.17	0.77	170	0.440
16	**0.21**	**0.75**	**-0.54**	**0.17**	**-3.16**	**166**	**0.002**
17	-0.25	-0.31	0.06	0.17	0.35	170	0.727
18	**-0.77**	**-1.26**	**0.49**	**0.21**	**2.36**	**164**	**0.020**
19	-0.57	-0.65	0.08	0.19	0.45	169	0.655
20	1.57	1.57	0	0.22	0	163	1.000
21	0.76	0.87	-0.11	0.18	-0.64	166	0.526
22	**0.41**	**0.07**	**0.35**	**0.17**	**2.1**	**170**	**0.037**
23	-0.68	-0.9	0.22	0.19	1.11	169	0.267

Note. DIF contrast is 4th-grade minus 5th-grade.

Discussion

Using Rasch measurement theory, this study investigated the psychometric quality of a 23-item growth mindset scale among elementary students, focusing on item functioning, dimensionality, reliability, and differential item functioning (DIF) across gender and grade levels. Overall, the findings confirm the validity and fairness of using the scale for upper elementary schoolers and suggest that several items need further examination.

The Rasch analysis indicated acceptable overall model fit, strong item separation reliability, and evidence consistent with unidimensionality. The Wright map showed acceptable alignment between item difficulty and person ability, and the scale is most informative for students with moderate growth mindset beliefs. There are minor gaps at the extremes of the continuum that indicate reduced sensitivity for students with very low or very high growth mindsets, a limitation commonly observed in attitudinal measures administered to younger learners [12]. Item-level analyses further supported the adequacy of the item set, with all items demonstrating acceptable infit and outfit statistics. All these findings support the internal structure and construct representation of the scale within this age group.

Differential Item Functioning across Gender and Grade

Differential item functioning analyses indicated that most items functioned equivalently across gender and grade level, supporting largely invariant measurement of growth mindset for 4th- and 5th-grade students. A small number of items showed statistically significant DIF related to domain-specific ability beliefs, stereotypes, feedback orientation, and learning motivation, but DIF magnitudes (0.34-0.64 logits) were small to moderate, and did not suggest meaningful scale-level bias.

Implications for Engineering Education and Early STEM Learning

The results have important implications for engineering education and early STEM learning because a growth mindset is associated with persistence, problem solving, and engagement with challenging tasks [15], [19], [3], [20]. The 23-item growth mindset scale provides a psychometrically sound instrument for evaluating mindset beliefs during a developmental period when STEM attitudes begin to form [21]. By demonstrating measurement invariance for the majority of items in the scale across gender and grade, this study supports the use of the scale for identifying potential disparities and evaluating early interventions aimed at fostering growth-oriented beliefs before students encounter more formal engineering curricula.

At the same time, the identification of items related to domain-specific ability beliefs and gender stereotypes underscores the importance of examining how students' beliefs about science, effort, and talent may influence their engagement with engineering-related activities. These findings suggest that educators and researchers should attend not only to overall growth mindset scores, but also to specific belief patterns that may affect students' willingness to persist in engineering-relevant tasks.

More broadly, this study also shows the value of applying Rasch measurement theory in engineering education research. Through a detailed examination of item functioning and fairness, Rasch analysis provides a rigorous framework for validating instruments used to study motivation and mindset in STEM contexts.

Building on these findings, the validated scale can be used in early engineering learning settings to support evidence-based instruction. For example, teachers may administer it before and after engineering design challenges to examine whether revising designs over multiple rounds, reflecting on what went wrong, and using feedback are associated with shifts toward growth-oriented beliefs. Results can then inform targeted instructional support for students or groups who show lower growth-oriented beliefs. For gender and grade comparisons, DIF-flagged items should be handled cautiously by reporting sensitivity analyses (e.g., estimating results with and without those items) and avoiding conclusions that rely heavily on a small subset of DIF items.

Limitation and Future Suggestions

Several limitations should be acknowledged. First, the findings are limited to 4th- and 5th-grade students, so the psychometric properties observed in this study may not generalize to older populations. Future research should examine whether the scale demonstrates similar measurement properties and invariance in older student populations. In addition, the Wright map indicated reduced measurement precision at the extreme ends of the latent continuum, indicating the need to refine existing items or expand item coverage to better capture very low or very high growth mindset beliefs. Although DIF was small to moderate for a few items, content related to domain-specific ability beliefs, gender stereotypes, and feedback orientations may be interpreted differently across groups. A qualitative approach can be applied to better understand students' interpretations. Finally, this study employed a cross-sectional design and did not examine relationships between growth mindset scores and external outcomes. Future research should extend this work by examining longitudinal measurement invariance and exploring how growth mindset scores relate to engineering-specific outcomes such as problem-solving performance, design thinking, and persistence in STEM learning environments.

Conclusion

This study provides empirical evidence supporting the psychometric quality of a 23-item growth mindset scale for use with upper elementary students. Using Rasch measurement theory, the results demonstrated acceptable model fit, strong item separation reliability, evidence consistent with unidimensionality, and invariant measurement across gender and grade levels. The Wright map indicated reasonable targeting of the scale, with strong measurement precision concentrated in the mid-range of the growth mindset continuum, while item-level analyses supported the adequacy and construct representation of the item set.

Even though we found a small number of items with statistically significant DIF, the magnitude of these effects was small to moderate and didn't indicate systematic bias at the scale level. Overall, the findings support the fair and meaningful use of the scale for assessing growth mindset among 4th- and 5th-grade students and across genders. By validating a growth mindset measure at an early developmental stage, this study contributes to engineering and STEM education research by providing a psychometrically sound tool for examining foundational beliefs that support persistence, engagement, and learning in challenging contexts, while demonstrating the value of Rasch measurement theory for strengthening validity arguments and informing educational practice.

References

[1] D. S. Yeager and C. S. Dweck, “What can be learned from growth mindset controversies?” *American Psychologist*, vol. 75, no. 9, pp. 1269–1284, 2020, doi: 10.1037/amp0000794.

[2] A. L. Campbell, I. Direito, and M. Mokhithi, “Developing growth mindsets in engineering students: A systematic literature review of interventions,” *European Journal of Engineering Education*, vol. 46, no. 4, pp. 503–527, 2021, doi: 10.1080/03043797.2021.1903835.

[3] A. Godwin, G. Potvin, Z. Hazari, and R. Lock, “Identity, critical agency, and engineering: An affective model for predicting engineering as a career choice,” *Journal of Engineering Education*, vol. 105, no. 2, pp. 312–340, 2016, doi: 10.1002/jee.20118.

[4] R. Ackerman and L. Levontin, “Mindset effects on the regulation of thinking time in problem-solving,” *Thinking & Reasoning*, vol. 30, no. 3, pp. 479–508, 2024, doi: 10.1080/13546783.2023.2259550.

[5] J. Suh, S. Graham, T. Ferranone, G. Kopeinig, and B. Bertholet, “Developing persistent and flexible problem solvers with a growth mindset,” in *Motivation and Disposition: Pathways to Learning Mathematics*, D. J. Brahier and W. R. Speer, Eds. Reston, VA, USA: National Council of Teachers of Mathematics, 2011, pp. 169–184.

[6] G. W. Hodges, K. Flanagan, J. Lee, A. Cohen, S. Krishnan, and C. Ward, “A quasi-experimental study comparing learning gains associated with serious educational gameplay and hands-on science in elementary classrooms,” *Journal of Research in Science Teaching*, vol. 57, pp. 1460–1489, 2020, doi: 10.1002/tea.21661.

[7] M. Boonma, “The power of resilience in 21st-century education: How growth mindset and mental well-being support teachers and students,” *Nakhon Ratchasima Journal of Humanities and Social Sciences*, vol. 1, no. 3, pp. 1–26, 2025. [Online]. Available: https://so11.tci-thaijo.org/index.php/NJHSS/article/view/1941. [Accessed February 11, 2026].

[8] S. Murphy, A. MacDonald, C. A. Wang, and A. Cardoso, “Towards an understanding of STEM engagement: A review of the literature on motivation and academic emotions,” *Canadian Journal of Science, Mathematics and Technology Education*, vol. 19, no. 3, pp. 304–320, 2019, doi: 10.1007/s42330-019-00054-w.

[9] K. Olive, X. Tang, A. Loukomies, K. Juuti, and K. Salmela-Aro, “Gendered differences in motivational profiles, achievement, and STEM aspirations of elementary school students,” *Frontiers in Psychology*, vol. 13, Art. no. 954325, 2022, doi: 10.3389/fpsyg.2022.954325.

[10] T. Jansen, J. Meyer, A. Wigfield, and J. Möller, “Which student and instructional variables are most strongly related to academic motivation in K–12 education? A systematic review of meta-analyses,” *Psychological Bulletin*, vol. 148, no. 1–2, pp. 1–26, 2022, doi: 10.1037/bul0000354.

[11] A. Y. Kim, G. M. Sinatra, and V. Seyranian, “Developing a STEM identity among young women: A social identity perspective,” *Review of Educational Research*, vol. 88, no. 4, pp. 589–625, 2018, doi: 10.3102/0034654318779957.

[12] N. Borgers, E. de Leeuw, and J. Hox, “Children as respondents in survey research: Cognitive development and response quality,” *Bulletin of Sociological Methodology / Bulletin de Méthodologie Sociologique*, vol. 66, no. 1, pp. 60–75, 2000, doi: 10.1177/075910630006600106.

[13] G. Rasch, *Probabilistic Models for Some Intelligence and Attainment Tests*, Expanded ed. Chicago, IL, USA: The University of Chicago Press, 1980.

[14] C. S. Dweck, C.-Y. Chiu, and Y.-Y. Hong, “Implicit theories and their role in judgments and reactions: A world from two perspectives,” *Psychological Inquiry*, vol. 6, no. 4, pp. 267–285, 1995, doi: 10.1207/s15327965pli0604_1.

[15] C. S. Dweck, *Mindset: The New Psychology of Success*. New York, NY, USA: Random House, 2006.

[16] J. M. Linacre, *Winsteps® Rasch Measurement Computer Program User’s Guide*, ver. 5.10.4, 2025. [Online]. Available: https://www.winsteps.com. [Accessed February 11, 2026].

[17] B. D. Wright and J. M. Linacre, “Reasonable mean-square fit values,” *Rasch Measurement Transactions*, vol. 8, no. 3, p. 370, 1994.

[18] B. L. Welch, “The generalization of ‘Student’s’ problem when several different population variances are involved,” *Biometrika*, vol. 34, no. 1–2, pp. 28–35, 1947, doi: 10.2307/2332510.

[19] L. S. Blackwell, K. H. Trzesniewski, and C. S. Dweck, “Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention,” *Child Development*, vol. 78, no. 1, pp. 246–263, 2007, doi: 10.1111/j.1467-8624.2007.00995.x.

[20] D. S. Yeager and C. S. Dweck, “Mindsets that promote resilience: When students believe that personal characteristics can be developed,” *Educational Psychologist*, vol. 47, no. 4, pp. 302–314, 2012, doi: 10.1080/00461520.2012.722805.

[21] B. M. Capobianco, B. F. French, and H. A. Diefes-Dux, “Engineering identity development among pre-adolescent learners,” *Journal of Engineering Education*, vol. 101, no. 4, pp. 698–716, 2012, doi: 10.1002/j.2168-9830.2012.tb01125.x.

Appendix A
DIF items across gender

Item	Item Content
1	Your intelligence is something very basic about you that you can’t change very much.
2	No matter how much intelligence you have, you can always change it quite a bit.
3	You can always substantially change how intelligent you are.
4	You are a certain kind of person, and there is not much that can be done to really change that.
5	You can always change basic things about the kind of person you are.
6	Music talent can be learned by anyone.
7	**Only a few people will be truly good at sports – you have to be “born with it.”**
8	**Science is much easier for boys than girls.**
9	The harder you work at something, the better you will be at it.
10	No matter what kind of person you are, you can always change substantially.
11	Trying new things is stressful for me and I avoid it.
12	Some people are good and kind, and some are not – it’s not often that people change.
13	**I appreciate when people, parents, coaches, teachers give me feedback about my performance.**
14	I often get angry when I get feedback about my performance.
15	All human beings without a brain injury or birth defect are capable of the same amount of learning.
16	You can learn new things, but you can’t really change how intelligent you are.
17	You can do things differently, but the important parts of who you are can’t really be changed.
18	Human beings are basically good, but sometimes make terrible decisions.
19	**An important reason why I do my school work is that I like to learn new things.**
20	**Truly smart people do not need to try hard.**
21	In school, if I get the wrong answer more than once, then I am not very good at that subject.
22	When I work in a group, I find it frustrating when we don't use my ideas.
23	When I work in a group on a project, we create something better than I could do by myself.

Appendix B
DIF items across grade

Item	Item content
1	Your intelligence is something very basic about you that you can't change very much.
2	No matter how much intelligence you have, you can always change it quite a bit.
3	You can always substantially change how intelligent you are.
4	**You are a certain kind of person, and there is not much that can be done to really change that.**
5	You can always change basic things about the kind of person you are.
6	Music talent can be learned by anyone.
7	**Only a few people will be truly good at sports – you have to be "born with it."**
8	**Science is much easier for boys than girls.**
9	The harder you work at something, the better you will be at it.
10	No matter what kind of person you are, you can always change substantially.
11	Trying new things is stressful for me and I avoid it.
12	Some people are good and kind, and some are not – it's not often that people change.
13	I appreciate when people, parents, coaches, teachers give me feedback about my performance.
14	I often get angry when I get feedback about my performance.
15	All human beings without a brain injury or birth defect are capable of the same amount of learning.
16	**You can learn new things, but you can't really change how intelligent you are.**
17	You can do things differently, but the important parts of who you are can't really be changed.
18	**Human beings are basically good, but sometimes make terrible decisions.**
19	An important reason why I do my school work is that I like to learn new things.
20	Truly smart people do not need to try hard.
21	In school, if I get the wrong answer more than once, then I am not very good at that subject.
22	**When I work in a group, I find it frustrating when we don't use my ideas.**
23	When I work in a group on a project, we create something better than I could do by myself.