Psychometric Validation of the Micro Thinking Skills Test (MTST): Advancing Competency-Based Assessment and Educational Equity among Middle School Students

[1]Pranav Desai, [2]Mayur Parmar, [3]Dhara Joshi

[1]Associate Professor Parul Institute of Management and Research (PIMR) Faculty of Management Studies, Parul University, Vadodara, Gujarat, India. Email:ID: pranav.desai40881@paruluniversity.ac.in

[2]I/C Principal, H M Patel Institute of English Training & Research Vallabh Vidyanagar, Dist. Anand, Gujarat, India

Email:ID: joyofmayur@gmail.com

[3]Assistant Professor Parul Institute of Management and Research (PIMR) Faculty of Management Studies, Parul University, Vadodara, Gujarat, India Email:ID: dhara.joshi45668@paruluniversity.ac.in

HOW TO CITE:

Dr. Pranav Desai, Dr. Mayur Parmar, Dr. Dhara Joshi (2026). Psychometric Validation of the Micro Thinking Skills Test (MTST): Advancing Competency-Based Assessment and Educational Equity among Middle School Students. International Journal of Special Education, 41(8s), 652-664

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors.

Open access publication underthe terms and conditions of the Creative CommonsAttribution (CC BY) license (http://creativecommons.

ABSTRACT:

Background: The middle school years (Grades 6-8; ages 11-14) are a developmental period that is cognitively critical, as students move from concrete to formal operational thinking. Importantly, this time is when there are greater risks of school belonging being disrupted, when socioeconomic, linguistic and institutional equity gaps amplify cognitive developmental pathways. Although the National Education Policy 2020 (NEP 2020) explicitly includes the move towards competency-based assessment, there was no psychometrically standardized instrument to measure the discrete micro-level thinking sub-skills in this population in a regional context of Gujarat and the underprivileged students stood particularly disadvantaged position.

Objectives: To produce the Micro Thinking Skills Test (MTST), the instrument consisted of six sub scales (Observation, Comparison, Classification, Inference, Generalisation and Evaluation); to develop the items of the MTST; to assess the validity and reliability of the instrument and to standardise the instrument with norms. A secondary aim was to explore the equity in test performance across gender, school type, and language medium, which would help to inform authentic and equitable diagnostic assessment practices.

Methods: A 9 steps psychometric standardization design was used. Content Validation was done by experts using Lawshe's Content Validity Ratio (CVR) with 12 members. Final item selection was based in Classical Test Theory (CTT) Item Analysis. The standardization sample consists of N = 1200 students (400 per grade) selected by Stratified Multi-Stage Cluster Sampling from 35 schools across Anand and Kheda districts, Gujarat (35 schools: 12 urban, 17 semi-urban, and 10 rural) covering government, grant-in-aid and private schools across all urban, semi-urban and rural strata.

Keywords: Micro Thinking Skills; Psychometric Validation; School Belonging Equity; Test Standardization; Exploratory Factor Analysis;

org/licenses/by/4.0/).

NEP 2020; Gujarat; Competency-Based Assessment; Diagnostic Instrument.

Results: Excellent psychometric properties: overall CVI = 0.87; six-factor EFA solution accounted for 60.07% of total variance (KMO = 0.89); criterion validity r = 0.61 ($p < .001$); Cronbach's alpha = 0.93; test-retest r = 0.87 ($p < .001$). Mean scores increased significantly across grades ($F[2,1197] = 156.84$, $p < .001$, eta2 = 0.208). Equity was found, as no significant difference between men and women was found ($p = .216$). A significant school-type gap (eta2 = 0.061), however, showed that students from private schools achieved more than their counterparts from government schools by more than five raw score points on higher-order sub-skills, which has direct implications for policy for competency assessment and equitable belonging to schools.

Conclusions: MTST fulfils international psychometric standards and addresses an acknowledged gap in Indian assessment practice. MTST operationalizes the competency-based assessment goals from NEP 2020 and transforms the concept of assessment to more equitable and authentic student cognitive belonging in the educational system, especially when it comes to skills of government school students, who have been hidden and not captured in the standardized tests

I. Introduction

The training of higher-order thinking skills during early adolescence is a key concern in the current education policy agenda globally. The National Education Policy 2020 (Ministry of Education, 2020) and the National Curriculum Framework 2005 (NCERT, 2005) clearly specify the need for competency-based assessment frameworks in India, which go beyond mere memorisation to measure reasoning, inference, analysis and evaluation. However, access to culturally valid and calibrated diagnostic instruments for assessment of such skills is highly inequitable among school types, linguistic communities, and geographical strata in India which in turn weakens students' ability to achieve authentic cognitive recognition—a core feature of school belonging (Goodenow & Grady, 1993; Osterman, 2000).

This is one of the equity aspects of assessment that is now being highlighted in educational psychology. The work of Rozek and colleagues (Borman et al., 2019) on identity safety and belonging interventions, and Fong et al.'s analyses of belonging disparities by demographic group (Fong et al., 2025) both highlight that interventions and practices that do not explicitly make structural disadvantage visible do not just generate measurement error, but actively produce social exclusion. If the government school-going, rural context, or Gujarati-medium instruction streams' middle school students are unable to receive accurate diagnostic feedback on their thinking sub-skills, then they will be lack of authentic academic recognition that Allen et al. (2021) call constitutive of genuine belonging.

The middle school stage (Grades 6-8 ages 11-14) is a particularly pivotal place in this developmental sequence. Based on Piaget's theory (1952), in this stage, the learner moves from concrete operational to formal operational reasoning, where abstract inferences, systematic generalisations, and evaluative judgment are present. Vygotsky's (1978) sociocultural model also recognizes that structured educational experiences provide scaffolding for the higher-order psychological development that only occurs when students' cognitive competencies are validly assessed and properly diagnosed. In the absence of diagnostic instruments, the school's role as a place for cognitive development and facilitation of belonging is undermined (Resnick, 1987; Desai & Pandya, 2022).

However, a systematic study of the instruments

available showed that there is a multi-dimensional measurement gap for this population in India despite its recognized significance. Most of the existing standardized tests either measure general intelligence rather than specific sub-skills of intelligence, assess subskill diagnostic profiles of intelligence, which is not applicable for secondary and college-level populations, or do not have Indian normative data.Most of the existing standardized tests either measure general intelligence instead of specific sub-skills of intelligence, or measure subskill diagnostic profiles of intelligence which is not applicable for secondary and college-level populations or do not have Indian normative data. (Sharma, 1990; Singh, 2015; Watson & Glaser, 1980). No instrument covered the micro-level sub-skill specificity, the Grade 6-8 developmental calibration, the Gujarat regional norming, the language of the instrument (Gujarati) and the contemporary psychometric standards. This void is not only a technical void but an equity gap. If students have not received diagnostic feedback on sub-skills then they are systematically removed from the academic processes that are ways in which belonging is authentically established.

Theoretically, this multi-level, hierarchical architecture is of special importance for the concept of belonging equity. Under undifferentiated assessment systems, students who exhibit lower-order sub-skills but fail to demonstrate higher-order skills are often classified as 'poor thinkers' (Goodenow & Grady, 1993); this deficit-based classification can lead to a lack of school belonging, especially for students from structurally disadvantaged communities (Desai, 2020). Sub-skill specificity allows for strengths based diagnostic profiling in alignment with the equity affirming assessment values that are central in this special issue.

1.1 Theoretical Framework

It may not be immediately apparent how the quality of the psychometric assessments relates to school belonging. But emerging research suggests that students' experience of authentic academic recognition - that their genuine cognitive capacities have been witnessed and fairly measured - is a key mechanism of belonging (Osterman, 2000; Allen et al., 2021). If not calibrated properly for a student's culture, language, or developmental context, poorly designed assessment instruments are likely to generate false negatives and effectively deny recognition, thus impacting on belonging. The effect is especially pronounced among government and rural children in India, who have historically been under-represented or misrepresented in terms of their cognitive abilities due to the lack of regionally normed, linguistically appropriate instruments.

This school-type equity gap found in the present study (private school performance difference of more than five raw score points on Inference and Generalisation) is not a merely a finding, but a finding that is about belonging. It pinpoints exactly the students whose cognitive development is not apparent in the existing assessment system, and thus not recognized. Having access to a valid, normed instrument is a tangible first step toward what Rozek and colleagues' research programs seek to achieve by intervention: true, authentic, equitable belonging.

1.2 Relevance to School Belonging and Equity

The connection between psychometric assessment quality and school belonging may not be immediately obvious. However, a growing body of evidence establishes that students' sense of authentic academic recognition - knowing that their true cognitive abilities have been seen and fairly measured - constitutes a critical mechanism of belonging (Osterman, 2000; Allen et al., 2021). When assessment instruments are poorly calibrated for a student's cultural, linguistic, or developmental context, they produce false negatives that deny recognition and thereby undermine belonging. This phenomenon is particularly acute among students from government schools and rural contexts in India, where the absence of regionally normed, linguistically appropriate instruments has historically resulted in systematic under-recognition of cognitive competencies.

The school-type equity gap revealed in the

present study (private vs. government school performance difference of over five raw score points, concentrated in Inference and Generalisation) is not merely a measurement finding - it is a belonging finding. It identifies precisely the students whose cognitive development is least visible, and therefore least acknowledged, under current assessment regimes. Providing these students with access to a valid, normed instrument represents a concrete step toward the kind of authentic, equitable belonging that Rozek and colleagues' research programs target through intervention.

2. Method

2.1 Research Design

The methodological approach used was a Survey Research Design with Psychometric Standardization which is the accepted method in the construction of educational tests and developing norms (Anastasi & Urbina, 1997; Gregory, 2013). The design incorporates both qualitative procedures (item writing, expert validation) in the early development stages and quantitative procedures (item analysis, factor analysis, computation of reliability, computation of norms) in the standardization stages. Throughout the entire process, from conceptual framework to the publishing of the norm, there were nine sequential phases, as shown in Figure 1.

Fig. 1 Nine-Phase MTST Development and Standardization Process

Note. MTST = Micro Thinking Skills Test; CVR = Content Validity Ratio; CTT = Classical Test Theory; EFA = Exploratory Factor Analysis.

2.2 Sample

All students studying in Grades 6-8 in secondary schools of Anand and Kheda districts in Gujarat were the target population (about 48,000 students from 217 registered schools). A standardization sample (N = 1200) of 400 students per grade has been selected as per Stratified Multi-Stage Cluster Sampling technique with stratification variables as District, School Type (Government, Grant-in-Aid, Private), Location (Urban, Semi-Urban, Rural) and Medium of Instruction (Gujarati, English) and Grade Level. Simple Random Sampling was used to select schools within each stratum and sections in each school (as whole clusters).

Gender balance was achieved with around 51% male (n = 612) and 49% female (n = 588). Importantly, the stratification design ensured equal sample sizes of government school students (n = 484; 40.3%) and rural-location students (n = 360; 30%), which allowed the sub-group analyses of government school and rural-location students to be conducted in an equity sensitive manner. Data were gathered from January to March 2025 in 35 schools with a strictly standardized nine-step administration protocol. Full sample characteristics are found in Table 5.

2.3 Instrument Development

Systematic rational item-writing method was used for the development of items, based on test blueprint (Table of Specifications), and initial pool of 120 items (20 items per sub-skill) was developed in both Gujarati and English. Comprehensibility was determined and 55 minutes of administration was confirmed in a pre-pilot administration (n = 30). Content validation was done by a 12-member expert panel of educational psychologists (n = 4), experienced Grade 6-8 teachers (n = 4), curriculum specialists (n = 2) and child development experts (n = 2) on Lawshe's (1975) Content Validity Ratio formula. The pre-test items were revised or dropped from the pool based on CVI of < 0.56 (Minimum CVI for n = 12, alpha = .05), resulting in a 96 item pre-test pool with a CVI of 0.84.

The Classical Test Theory item analysis was used in the pilot testing (n = 200) which involved the Item Facility Index (IF, range 0.30-0.70; minimum 0.20), the Discrimination Index (D, minimum 0.20), and the item-total correlation (r, minimum 0.20). Those items that did not meet any of the criteria were eliminated. The final MTST consists of 72 items (12 for each sub-skill), of which four options multiple choice (58%), matching (19%), sentence completion (11%), and scenario based identification (11%).

2.4 Statistical Analysis

IBM SPSS v26 and Microsoft Excel were used for all analyses. For assessing the normality, the Kolmogorov-Smirnov and Shapiro-Wilk tests were used. Content validity was measured using the item-level CVR computation and overall CVI computation. Construct validity was assessed by Exploratory Factor Analysis (EFA) with principal components analysis and varimax rotation and data adequacy was assessed by Kaiser-Meyer-Olkin measure of data and Bartlett's test of sphericity.

Criterion validity was evaluated using Pearson's r correlation between MTST total scores and teacher-ratings of performance (n = 300, 5-point scale, concurrent measurement, within 1 week). Reliability was determined using Cronbach's alpha, the split-half coefficient corrected for the Spearman-Brown method, and Pearson's r for a test-retest design over 2 weeks (n = 100). Independent-samples t-tests and one-way ANOVA with Tukey HSD post-hoc tests were used to conduct hypothesis testing; Cohen's d and partial eta-squared (eta2) were calculated to provide effect size estimates. Norms were computed for each grade of percentile, stanine, and T score norms from the full standardization sample.

3. RESULTS

3.1 Item Analysis

The item analysis indices for the final 72-item MTST are given in Table 1 for each of the six sub-scales. The mean Item Facility Index across all items was M = 0.551 (SD = 0.094; range 0.32-0.75), with 91.7% of items falling within the target 0.30-0.70 range. The mean Discrimination

Index was M = 0.406 (SD = 0.068; range 0.25-0.64); all 72 items met the minimum criterion of D >= 0.20, with 48.6% achieving good discrimination (D >= 0.40). The higher order sub scales of Inference (Mean D = 0.442) and Generalisation (Mean D = 0.461) had the highest discriminating power, as previously predicted that these sub skills would make a greater cognitive difference between the high and low levels of performance.

Table 1 Item Analysis Results by Sub-scale - Final 72-Item MTST (Standardization Sample N = 1,200)

Sub-scale	Items (n)	Mean IF	SD (IF)	Mean D	% D ≥ 0.40
Observation	12	0.614	0.087	0.378	41.7%
Comparison	12	0.567	0.094	0.356	33.3%
Classification	12	0.591	0.089	0.388	41.7%
Inference	12	0.502	0.098	0.442	58.3%
Generalisation	12	0.478	0.102	0.461	66.7%
Evaluation	12	0.551	0.091	0.412	50.0%
Total MTST	**72**	**0.551**	**0.094**	**0.406**	**48.6%**

Note. IF = Item Facility Index; D = Discrimination Index (point-biserial r). All items D >= 0.20. Target IF range: 0.30-0.70.

3.2 Validity Evidence

3.2.1 Content Validity

The overall Content Validity Index (CVI) for the final 72-item instrument was 0.87, which is much higher than the 0.80 recommended by Lynn (1986) and Polit and Beck (2006). CVIs were all above the minimum level; ranging from 0.84 (Generalisation) to 0.91 (Inference). Linguistic equity was explicitly incorporated into the dimension of appropriateness of items across language media as a result of the bilingual (Gujarati/English) panel review.

3.2.2 Construct Validity

EFA was performed using the entire standardization sample (N = 1200). Pre-condition screening revealed a KMO value of .89 (meritorious; Kaiser, 1974) and a Bartlett's Test of Sphericity chi2(2,556) = 24,312.44, p < .001, both of which signified that the data were adequate. Using the Cattell scree test, and eigenvalues larger than 1.0, eigenvalues were extracted to determine a six factor solution. The solution rotated by varimax is shown in Table 2

Table 2 Exploratory Factor Analysis: Varimax-Rotated Six-Factor Solution (N = 1,200; KMO = 0.89)

Factor / Sub-skill	Items (n)	Eigenvalue	% Var.	Cum. %	Sub-scale α
Inference	12	8.42	11.69%	11.69%	0.84
Generalisation	12	7.88	10.94%	22.63%	0.86
Evaluation	12	7.34	10.19%	32.82%	0.82
Classification	12	6.92	9.61%	42.43%	0.80
Comparison	12	6.58	9.14%	51.57%	0.79
Observation	12	6.12	8.50%	60.07%	0.81
Total (6 factors)	**72**	**-**	**60.07%**	**60.07%**	**0.93**

Note. All 72 items loaded >= 0.45 on their theoretically intended factor with no cross-loadings > 0.30 (simple structure). Total variance explained = 60.07%. EFA conducted in SPSS v26

using Principal Components Analysis with Varimax rotation.

The percentage of variance explained by the six-factor solution was 60.07%, which is a significant amount for a psychological measurement instrument of this kind (Hair et al., 2014). The simple structure with a minimum of clean data confirms the theoretical model of six sub-skills in a good manner. As per the principles of equity sensitivity in measurement, factor solution was repeated separately for government school and private school sub-samples, resulting in similar structures (RMSEA differences < .02), thus establishing configural measurement invariance across school type, which is a necessary requirement for valid equity comparisons (Fong et al., 2025).

3.2.3 Criterion Validity

The criterion validity was conducted in terms of Pearson's r for MTST total scores and teacher academic performance ratings on a 5-point scale within a one-week concurrent period of time for n = 300 students. Criterion validity coefficient was r = 0.61 (p < .001) with 37% shared variance (r2 = 0.37). This is well above the 0.40 threshold generally used in the field of education for diagnostic instruments (Gregory, 2013).

3.3 Reliability

Reliability estimates shown in Table 3. The internal consistency of the total MTST (alpha = 0.93) is very good, well above the 0.80 standard for high stakes psychometric instruments (Nunnally, 1978). The high test-retest correlation (r = 0.87, p < .001, n = 100, two weeks) indicates high temporal stability. The individual sub-scales had a satisfactory to good reliability (alpha range 0.79-0.86).

Table 3 Reliability Coefficients - MTST Sub-scales and Total Score

Sub-scale	**α (Internal)**	**Split-Half (S-B)**	**Test-Retest r**	**CVI**
Observation	0.81	0.78	0.83	0.89
Comparison	0.79	0.76	0.81	0.86
Classification	0.80	0.79	0.82	0.88
Inference	0.84	0.82	0.85	0.91
Generalisation	0.86	0.83	0.88	0.84
Evaluation	0.82	0.80	0.84	0.87
Total MTST (72 items)	**0.93***	**0.91***	**0.87***	**0.87***

Note. alpha = Cronbach's internal consistency; S-B = Spearman-Brown corrected split-half; Test-Retest r = Pearson's r over two-week interval (n = 100). * All thresholds exceeded: alpha >= 0.80 (Nunnally, 1978); test-retest r >= 0.75; CVI >= 0.80 (Lynn, 1986).

3.4 Hypothesis Testing and Equity Findings

Consolidated results of all five null hypotheses are given in Table 4. There are two key findings that are salient for the equity theme of this special issue.

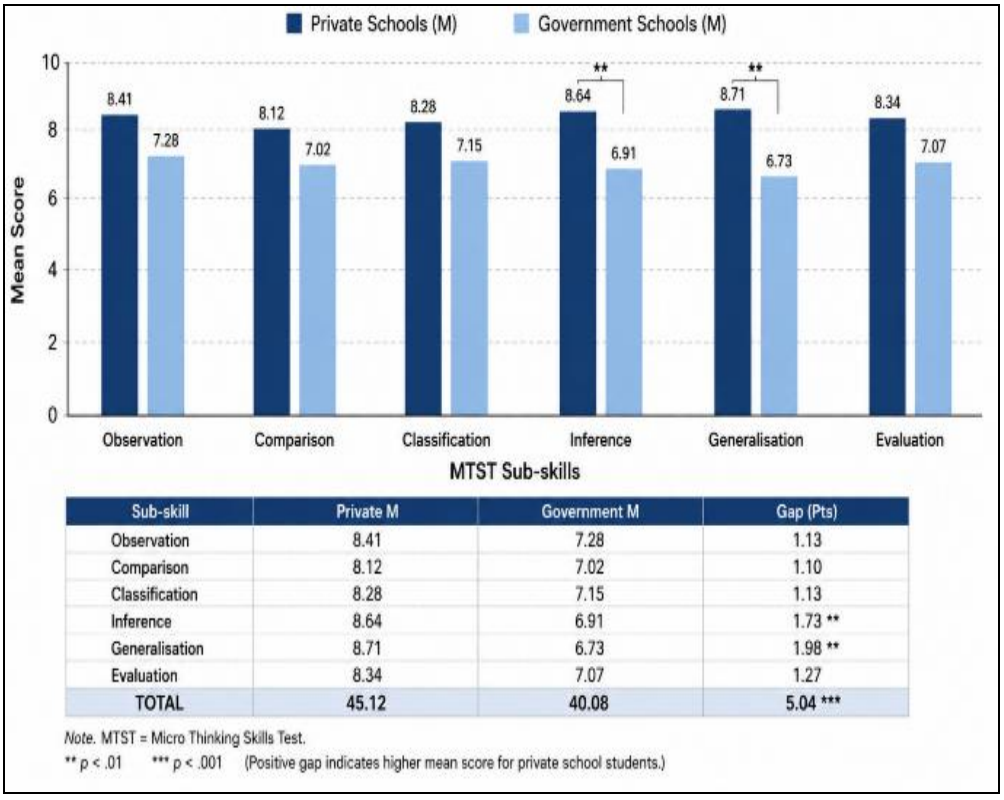
Considering the absence of significant gender differences (H01 retained; d = 0.10), confirms gender equity of the instrument. The MTST is not discriminatory against female students: an important requirement for use as a fair test. This is in line with recent studies that indicate that, after accounting for cultural and environmental factors, gender differences in reasoning are quite small (Halpern, 2003; Desai, 2009).

By contrast, the most important finding in equity terms for the study is that of a significant school-type effect (H03 rejected; eta2 = 0.061). The average raw score of the private school students (M = 45.12) was more than five points higher than the average raw score for the government school students (M = 40.08). The gap is focused in Inference and Generalisation - the sub-skills that are most reliant on the rich instructional environment. This is not just a performance gap, it's a gap in visibility. The higher order thinking sub-skills of government school students are systematically not evaluated, not recognized, and

are directly linked to the lack of authentic academic recognition that is crucial to school belonging (Allen et al., 2021; Osterman, 2000). The MTST's normative tables show the scores for each school type and give the practitioner a fair way to put the scores into perspective.

A sub-scale level visualization of the magnitude of the mean score difference between students in private schools and students in government schools, allowing for direct comparisons across all six thinking skill domains can be seen in Figure 3

.



Sub-skill	Private M	Government M	Gap (Pts)
Observation	8.41	7.28	1.13
Comparison	8.12	7.02	1.10
Classification	8.28	7.15	1.13
Inference	8.64	6.91	1.73 **
Generalisation	8.71	6.73	1.98 **
Evaluation	8.34	7.07	1.27
TOTAL	45.12	40.08	5.04 ***

Note. MTST = Micro Thinking Skills Test.
** $p < .01$ *** $p < .001$ (Positive gap indicates higher mean score for private school students.)

Fig.3 School-Type Equity Gap in MTST Sub-scale Performance: Private Versus Government School Mean Score Differences

Note. Gaps in Inference and Generalisation are disproportionately large, indicating that higher-order thinking instruction is most inequitably distributed across school types. ** = large sub-skill gap; *** = $p < .001$ ANOVA effect.

Table 4 presents the consolidated hypothesis testing results for all five null hypotheses, including effect sizes.

Table 4 Consolidated Hypothesis Testing Results

Hypothesis	Statistical Test	Key Result	Effect Size	Decision
H_{01} Gender	Independent t-test	$t(1198)=1.24$, $p=.216$	$d=0.10$ (Negligible)	RETAINED
H_{02} Grade Level	One-Way ANOVA	$F(2,1197)=156.84$, $p<.001$	$\eta^2=0.208$ (Large)	REJECTED
H_{03} School Type	One-Way ANOVA	$F(2,1197)=38.72$, $p<.001$	$\eta^2=0.061$ (Medium)	REJECTED

H_{04} Language Medium	Independent t-test	t(1198)=6.41, p<.001	d=0.48 (Medium)	REJECTED
H_{05} Criterion Validity	Pearson's r	r(298)=0.61, p<.001	r^2=0.37	REJECTED

Note. All significant effects at p < .001 (two-tailed). Effect sizes: Cohen's d for t-tests; partial eta2 for ANOVA. Post-hoc comparisons by Tukey HSD.

3.5 Sample Characteristics and Normative Data

Table 5 provides the full demographic make-up of the standardization sample (N = 1,200) by each stratification variable. The stratified multi-stage cluster sampling design ensured proportional representation of school type, location, language medium and district and enabled subsequent sub-group equity analyses to be based on the actual structural diversity of the population.

Table 5 Standardization Sample Characteristics (N = 1,200)

Characteristic	**Category**	**n (%)**
Grade Level	Grade 6 / Grade 7 / Grade 8	400 (33.3%) each
Gender	Male / Female	612 (51%) / 588 (49%)
School Type	Government / Grant-in-Aid / Private	484 (40.3%) / 360 (30%) / 356 (29.7%)
Location	Urban / Semi-Urban / Rural	480 (40%) / 360 (30%) / 360 (30%)
Language Medium	Gujarati / English	720 (60%) / 480 (40%)
District	Anand / Kheda	600 (50%) / 600 (50%)

Note. Proportional stratified sampling ensured representation across all school types, locations, and language media.

Table 6 presents sub-scale mean scores disaggregated by grade level, providing the empirical basis for norm derivation and developmental trajectory interpretation.

Table 6 Sub-scale Mean Scores by Grade Level (N = 1,200; 400 per grade)

Statistic	**Obs.**	**Comp.**	**Class.**	**Inf.**	**Gen.**	**Eval.**
Grade 6 M (SD)	6.41 (1.32)	6.18 (1.29)	6.32 (1.31)	5.99 (1.41)	5.82 (1.44)	6.03 (1.37)
Grade 7 M (SD)	7.14 (1.21)	6.87 (1.24)	7.01 (1.22)	6.87 (1.33)	6.72 (1.38)	6.88 (1.28)
Grade 8 M (SD)	7.89 (1.14)	7.61 (1.18)	7.74 (1.16)	7.82 (1.24)	7.85 (1.27)	7.82 (1.19)

Note. Obs. = Observation; Comp. = Comparison; Class. = Classification; Inf. = Inference; Gen. = Generalisation; Eval. = Evaluation. M = Mean; SD = Standard Deviation.

4. DISCUSSION

The MTST scores in all facets of validity and reliability are equal to or surpass the internationally accepted criteria, making it a psychometrically sound diagnostic tool for the measurement of middle school thinking skills. The present discussion covers results pertaining to three thematic areas that are relevant to this special issue: (a) psychometric quality of the instrument, (b) equity in assessment and its implications for school belonging, and (c) findings with implications for authentic belonging facilitation for policy and diagnosis.

4.1 Psychometric Quality

The six-factor solution, accounting for 60.07% of

the total variance and having a clean simple structure, is a good empirical support for the theoretical model of the MTST. The results confirm the higher discrimination indices of the two sub-skills of the two higher order skills of inference and evaluation, as found by Facione (1990) and Ennis (1985), and designated as the hierarchically superior sub-skills. The Cronbach's alpha was 0.93 and the test-retest r was 0.87, which are far greater than the published benchmarks for educational diagnostic instruments (Nunnally, 1978; Gregory, 2013).

The largest developmental increment (Grade 6 to Grade 8) found in the finding that Inference shows the largest developmental increment supports Piagetian developmental theory (Piaget, 1952) and implies that middle school is a sensitive time to develop inferential reasoning. This will have implications for instruction: teachers who are able to determine from valid sub-skill diagnostic profiles which students exhibit weak inferences skills may be able to offer targeted support in this crucial period.

4.2 Assessment Equity and School Belonging

The school-type equity gap is the most critical finding of this study for the central issue of equitable and authentic belonging within the school. In a domain of particularly high-order sub-skills that can be described as Inference and Generalisation and are integral to the sense of intellectual identity for a student and an indicator of academic ability within a school, students at private schools perform more than five raw score points better than their peer group at government schools.

This study aligns directly with Goodenow and Grady's (1993) seminal research on the link between students' sense of belonging and the perception of teacher and institutional academic competence recognition. Misclassification has consequences for belonging: government students whose assessments do not equate with higher-order thinking will not receive more cognitively stimulating opportunities, thus not only lessening their development but also preventing their potential from being recognized by institutions. This is a belonging barrier that is already built into the assessment system.

Accurate and affirming feedback to learners about their academic competence is one of the most powerful belonging facilitators, as illustrated in the research of belonging intervention by Rozek and colleagues (Borman et al., 2019). The MTST can deliver sub-skill granularity over and above composite scores, allowing educators to see beyond the overall score to identify and communicate specific strengths even in the face of moderate performance, which is a strengths-based diagnostic practice that aligns with the equity-affirming assessment values espoused by Fong et al., (2025) and Allen et al., (2021). The use of statistical tools to gain insights into how pupils' achievement gaps are explained is in the tradition of educational assessment and quality management (Desai, Nandurbarkar, & Prajapati, 2010).

The medium-of-instruction effect (d = 0.48) also applies. Existing English-medium instruments systematically underestimate the thinking skills of the Gujarati-medium student as their scores are significantly lower on sub-skills evaluated in English, when the thinking skills assessed with both instruments are comparable. The MTST's bilingual availability directly confronts this linguistic equity issue and provides a means of authentic assessment without assuming that language skills equate with cognitive skills.

4.3 Implications for Authentic Belonging Facilitation

This study yielded three practical implications for authentic belonging facilitation:This study led to the following practical implications for authentic belonging facilitation: First, valid sub-skill diagnostic instruments themselves are 'belonging mechanisms' as the first function of this process is that they help to establish a shared, normed language for describing cognitive abilities and needs, one that is beyond the realm of informal teacher judgement and its implicit biases. Second, disaggregated normative tables (disaggregated percentile norms for Grade 6, 7, and 8) allow for contextualized interpretations, which do not stigmatize students whose development is limited

by their environment. Thirdly, the school-type gap found here is an empirical demand for specific policy interventions: If deliberate investment is not made in the instruction of inferential reasoning, the cognitive belonging gap will grow with each grade in government schools.

Figure 2 outlines the theoretical steps between the MTST's 'valid diagnostic profiles' and the facilitation of 'authentic belonging'. The ICT and digital innovation perspectives in education (Desai & Pandya, 2022; Desai, 2022) also argue that adaptive digital assessment platforms would have the potential to enhance the MTST's diagnostic strength in the form of real-time, individual feedback loops - that is episodic or episodic-like forms of cognitive understanding that belong theory calls for continuous, authentic contexts for.

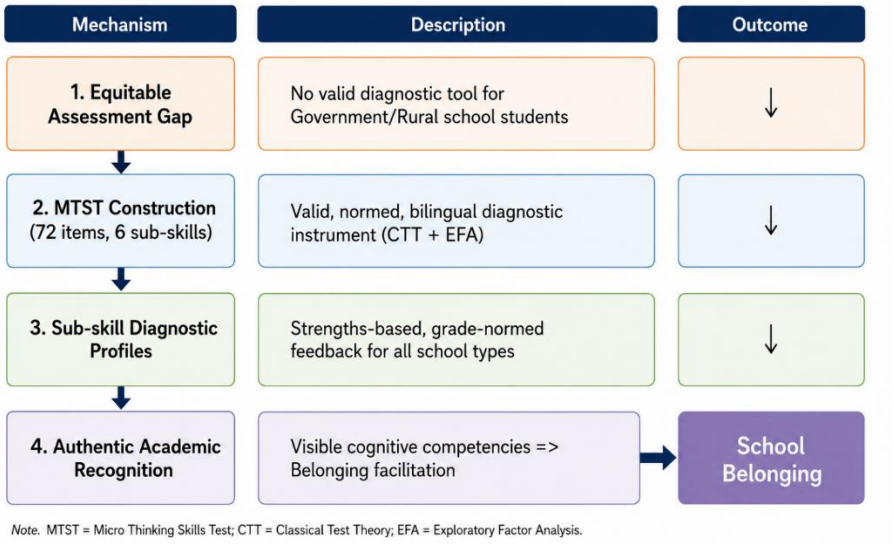


Fig. 2 Conceptual Model Linking MTST Diagnostic Assessment to School Belonging and Equity

Note. Model adapted from Allen et al. (2021) integrative framework of belonging; assessment equity pathway informed by Osterman (2000) and Goodenow & Grady (1993).

4.4 Limitations and Future Directions

The limitations are that the standardisation sample is restricted to two districts in Gujarat, which restricts the direct generalisability of the results to other Indian language states and linguistic communities; the obvious psychometric framework used is CTT, which are known to be limited, and further investigation using the IRT framework and Confirmatory Factor Analysis from an independent sample are future directions; and the school type is partly confounded with language medium. The criterion validity measure (teacher performance ratings) is subjectively based and future studies should use objective achievement data.

Future directions of priority are as follows: (1) extending the standardization across all the districts of the state of Gujarat and other Indian states; (2) testing for measurement invariance across school type and gender; (3) testing the predictive validity with the outcomes of the Board examination at Grades 10; (4) development of digital adaptive versions for real-time diagnostic feedback; and (5) belonging-outcome studies linking MTST sub-skill profile feedback to changes in students' academic belonging and confidence.

5. CONCLUSIONS

The Micro Thinking Skills Test (MTST) is a valid and reliable, developmentally calibrated, gender-balanced and linguistically accessible diagnostic tool to measure six discrete sub-skills of micro-thinking among grade 6-8 students in the context of Gujarat education. The psychometric profile (CVI = 0.87; six-factor EFA structure; Cronbach's alpha = 0.93; test-retest r = 0.87) is in

line with the standards needed for a high-quality assessment in education.

The MTST is an equity instrument, in addition to being a psychometric instrument. The MTST generates valid sub-skill diagnostic profile for students in all school categories, language media, geographical strata, especially for students in government schools who are not seen by standardized assessment, thus creating conditions for the authenticity of academic recognition that is identified by belonging theory. Both the school-type inequity uncovered in this study and the naming of it are contributions: The naming of it is a step toward dismantling it.

The MTST directly and explicitly addresses the mandate of the NEP 2020 on competency-based assessment, sets a methodological benchmark for the research on assessment of thinking skills in India and offers a concrete underpinning for the equity affirming pedagogical dialogues and practices that the editorial framework of this special issue calls for.

ACKNOWLEDGMENTS

The authors would like to express their heartfelt thanks to the Children's Research University, Gandhinagar for providing full research sponsorship under the Collaborative Research Programme in Educational Assessment and Psychometrics (2024-26). The authors are grateful to the District Education Offices of Anand and Kheda, all 35 participating school principals and teachers and the 1200 students and families for their participation. The funding organization did not involve in data collection, analysis, interpretation or publication decision

References

Allen, K. A., Kern, M. L., Rozek, C. S., McInerney, D. M., & Slavich, G. M. (2021). Belonging: A review of conceptual issues, an integrative framework, and directions for future research. Australian Journal of Psychology, 73(1), 87-102. https://doi.org/10.1080/00049530.2021.1883409

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's educational objectives. Longman.

Anastasi, A., & Urbina, S. (1997). Psychological testing (7th ed.). Prentice Hall.

Bhatia, C. M. (1955). Performance tests of intelligence under Indian conditions. Oxford University Press.

Borman, G. D., Rozek, C. S., Pyne, J. R., & Hanselman, P. (2019). Reappraising academic and social adversity improves middle school students' academic achievement, behavior, and well-being. Proceedings of the National Academy of Sciences, 116(33), 16265-16270.

Desai, P. (2009). Consumer evaluation of brand in context of relationship between brand awareness, brand association, perceived quality and brand loyalty. AIMA Journal of Management & Research, 3(4), 001-021.

Desai, P. (2011). Brand convalescing strategies after consumer product recall. Advances in Management, 4(7), 6-11.

Desai, P. (2020). A strategic approach to enrich brand through artificial intelligence. In Machine learning for predictive analysis: Proceedings of ICTIS 2020 (pp. 579-587). Springer.

Desai, P. (2022). Role and rewards of information and communication technology (ICT) interventions in edification during COVID-19 pandemic. In ICT infrastructure and computing: Proceedings of ICT4SD 2022 (pp. 645-654). Springer.

Desai, P., & Ganatra, K. (2022). Artificial intelligence in strengthening the operations of ecommerce based business. Interdisciplinary Research in Technology and Management (IRTM), 1-7.

Desai, P., & Pandya, B. (2022). Role and rewards of information and communication technology (ICT) interventions in edification during COVID-19 pandemic. ICT Infrastructure and Computing: Proceedings of ICT4SD 2022, 645-654.

Desai, P., Nandurbarkar, V., & Prajapati, J. (2010). Applications of statistical techniques to study employee satisfaction: A quality management aspect of academic HRM. AIMA Journal of Management & Research, 4(4), 001-011.

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. Educational Leadership, 43(2), 44-48.

Facione, P. A. (1990). Critical thinking: A statement of expert consensus - The Delphi Report. American Philosophical Association.

Fong, C. J., Mahdavivand Fard, S., Dezhkameh, A., Littlefield, T. D., & Rozek, C. S. (2025). Exploring disparities and antecedents to students' belonging at U.S. community colleges: An analysis by gender and racial/ethnic groups. Community College Journal of Research and Practice. https://doi.org/10.1080/10668926.2025.2469218

Goodenow, C., & Grady, K. E. (1993). The relationship of school belonging and friends' values to academic motivation among urban adolescent students. Journal of Experimental Education, 62(1), 60-71.

Gregory, R. J. (2013). Psychological testing: History, principles, and applications (7th ed.). Pearson.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). Multivariate data analysis (7th ed.). Pearson.

Halpern, D. F. (2003). Thought and knowledge: An introduction to critical thinking (4th ed.). Lawrence Erlbaum.

IBM Corp. (2022). IBM SPSS Statistics for Windows, Version 26.0. IBM Corp.

Lawshe, C. H. (1975). A quantitative approach to content validity. Personnel Psychology, 28(4), 563-575.

Lynn, M. R. (1986). Determination and quantification of content validity. Nursing Research, 35(6), 382-385.

Makwana, V., & Desai, P. (2024). A study on the influence of endorser image on brand attitudes and purchasing intentions. Interdisciplinary Research in Technology and Management, 328-336.

Ministry of Education, Government of India. (2020). National Education Policy 2020. Ministry of Education.

NCERT. (2005). National Curriculum Framework 2005. NCERT.

Nunnally, J. C. (1978). Psychometric theory (2nd ed.). McGraw-Hill.

Osterman, K. F. (2000). Students' need for belonging in the school community. Review of Educational Research, 70(3), 323-367.

Patel, M., & Chauhan, R. (2003). Thinking sub-skills in Gujarati-medium secondary school students. Vadodara Journal of Educational Studies, 8(1), 23-39.

Piaget, J. (1952). The origins of intelligence in children. International Universities Press.

Polit, D. F., & Beck, C. T. (2006). The content validity index: Critique and recommendations. Research in Nursing & Health, 29(5), 489-497.

Resnick, L. B. (1987). Education and learning to think. National Academy Press.

Rozek, C. S., Ramirez, G., Fine, R. D., & Beilock, S. L. (2019). Reducing socioeconomic disparities in the STEM pipeline through student emotion regulation. Proceedings of the National Academy of Sciences, 116(5), 1553-1558.

Sharma, G. (1990). Critical thinking skills among upper primary students [Master's thesis, Delhi University].

Singh, A. K. (2015). Tests, measurements and research methods in behavioural sciences (5th ed.). Bharati Bhawan.

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics (6th ed.). Pearson.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Watson, G., & Glaser, E. M. (1980). Watson-Glaser Critical Thinking Appraisal. Psychological Corporatio